ISCB International Society for Clinical Biostatistics

ISCB45 THESSALONIKI 2024

45th Annual Conference of the International Society for Clinical Biostatistics

FINAL PROGRAMME & BOOK OF ABSTRACTS

21-25 July 2024
Thessaloniki Concert Hall

www.iscb2024.info

CONTENTS

CONTENTS

WELCOME MESSAGE

Dear colleagues,

On behalf of the LOC and the SPC, it is our pleasure to welcome you to the **45th Annual Conference** of the **International Society for Clinical Biostatistics (ISCB45)**, taking place in **Thessaloniki Greece**, at the Thessaloniki Concert Hall on **21 to 25 July 2024**.

ISCB Conferences are open to clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics. We are pleased to see various disciplinary fields, professional ranks, and career stages represented.

The SPC has put a lot of effort in bringing to you a lively and interactive Main Conference Programme (Monday to Wednesday), consisting of abstract contributions of delegates, as well as focused topics delivered by world renowned experts in the invited sessions. The first day of the Conference (Sunday) is dedicated to Pre-Conference Courses, were limited attendance combined with focused topics ensures efficient didactic results. Finally, on the last day of the Conference (Thursday), the ISCB Statistics in Regulatory Affairs and STRATOS initiative are organising their Mini-Symposia, in parallel with the Early Career Biostatisticians programme.

ISCB45 also offers many networking opportunities, in the Students Networking Event (Sunday), the Welcome Reception (Monday), the Excursions offered at a small fee (Tuesday), and of course the Conference Dinner (Wednesday).

We have tried to keep as many virtual features as possible available during the ISCB45 Conference, to enable participants to follow parallel tracks of scientific content offered for post-viewing through our dedicated platform after completion of the event.

We are very excited to have you with us in Thessaloniki and wish you an enjoyable and rewarding ISCB45 Conference.

Vana Sypsa
ISCB45 LOC Chair

Giota Touloumi
ISCB45 SPC Chair

ORGANISATION - COMMITTEES - AUSPICES

Organised by

The International Society for Clinical Biostatistics (ISCB)

The International Society for Clinical Biostatistics (ISCB) was founded in 1978 to stimulate research into the principles and methodology used in the design and analysis of clinical research and to increase the relevance of statistical theory to the real world of clinical medicine.

Membership is open to all interested individuals who share the Aims of the Society. ISCB's membership includes clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics.

ISCB has an Executive Committee and 6 Subcommittees: Conference Organising, Early Career Biostatisticians, Education, National Groups, Statistics in Regulatory Affairs (SiRA), Student Conference Awards.

Executive Committee

President
Tomasz Burzykowski (BE)

Vice-President
Thomas Jaki (UK)

Secretary
Elaine Pascoe (AU)

Treasurer
Chris Metcalfe (UK)

Member and Media & Communications Coordinator
Charlotte Bolch (US)

Members
Kinga Salapa (PL)
Laure Wynants (NL)
Milada Cvancarova Småstuen (NO)
Nan van Geloven (NL)
Nikos Pantazis (GR)
Paola Rebora (IT)
Philip S Boonstra (US)

Honorary Members

Michal Abrahamowicz (CA)
Harbajan Chadha-Boreham (FR)
Lutz Edler (DE)
Emmanuel Lesaffre (BE)
Michael Schemper (AT)
Martin Schumacher (DE)
Jørgen Seldrup (FR)
Stephen J. Senn (UK)
Zdenek Valenta (CZ)
Maria Grazia Valsecchi (IT)
Hans C. van Houwelingen (NL)
David W. Warne (CH)
John R. Whitehead (UK)

ISCB Permanent Office

convin

29 K. Varnali Street, 152 33 Chalandri,
Athens, Greece
Call center: +30 210 6833600
Mobile contact: +30 6956 665669
Website: www.iscb.international
Email: office@iscb.international

ORGANISATION - COMMITTEES - AUSPICES

Local Organising Committee (LOC)

Chair:
Vana Sypsa

Treasurer:
Nikos Pantazis

Members:
Pantelis Bagos
Grigorios Chlouverakis
Nikos Demiris
Evangelos Evangelou†
Anna-Bettina Haidich
Evangelos Kritsotakis
Loukia Meligkotsidou
Christos Nakas
Dimosthenis Panagiotakos
Sotirios Roussos
Evi Samoli
Fotios Siannis
Giota Touloumi
Grigorios Trypsiannis
Kostas Tsilidis

Professional Congress Organiser

convin

29 K. Varnali Street,
152 33 Chalandri, Athens, Greece
Corporate W: www.convin.gr
Corporate E: info@convin.gr
ISCB45 W: www.iscb2024.info
ISCB45 E: info@iscb2024.info

Scientific Programme Committee (SPC)

Chair:
Giota Touloumi (GR)

Members:
Michal Abrahamowicz (CA)
Mouna Akacha (CH)
Abdel Babiker (UK)
Mark Baguelin (UK)
Giorgos Bakoyannis (US)
Nadine Binder (DE)
Apostolos Bournetas (GR)
Urania Dafni (GR)
Issa Dahabreh (US)
Spiros Denaxas (UK)
Els Goetghebeur (BE)
Guadalupe Gomes (ES)
Miguel Hernan (US)
Dimitris Karlis (GR)
Paola Rebora (IT)
Dimitris Rizopoulos (NL)
Marcia Rueckbeil (NL)
Georgia Salanti (CH)
Rodolphe Thiébaut (FR)
Ian White (UK)
Marvin N. Wright (DE)
Donglin Zeng (US)

UNDER THE AUSPICES OF:

ARISTOTLE UNIVERSITY OF THESSALONIKI
HELLENIC REPUBLIC National and Kapodistrian University of Athens EST. 1837
REGION OF CENTRAL MACEDONIA
CITY OF THESSALONIKI

AWARDS & SUPPORT SCHEME

Student Conference Award (StCA) Winners	Title of presentation	Session code	University / Organisation	Country
Awan Afiaz	Optimal sandwich variance estimator in penalised GEE for nearly separated longitudinal binary data with small samples	OC13-5	University of Washington, Seattle WA; ISRT, University of Dhaka	United States; Bangladesh
Haya Elayan	Correcting for case-mix shift when developing clinical prediction models	OC12-5	University of Manchester	United Kingdom
Léa Orsini	Bayesian analysis of restricted mean survival time adjusted on covariates using pseudo-observations	OC30-5	Paris-Saclay University	France
Achilleas Stamoulopoulos	Joint modelling of longitudinal data and informative visiting process to predict subject-specific probabilities of future gaps in care	OC21-1	National and Kapodistrian University of Athens	Greece
Conference Award for Scientists (CASc) Winners	**Title of presentation**	**Session code**	**University / Organisation**	**Country**
Mahnaz Ibrahim	On estimation of relative risk regression model for binary data	OC32-3	University of Dhaka	Bangladesh
Shafayet Khan Shafee	Interval estimation of median odds ratio for measuring contextual effects in multilevel data using binary logistic model	P-B06-06	Pathao Limited, Dhaka	Bangladesh
Nabil Ahmed Uthso	Machine learning methods in dynamic survival prediction under standard and mixed-model landmarking framework	P-B01-07	University of Dhaka	Bangladesh
Conference Fund for Developing Countries (CFDC) Recipients	**Title of presentation**	**Session code**	**University / Organisation**	**Country**
Behnaz Alafchi	Joint modelling of ordinal repeated measurements and multi-state survival data: A simulation study	P-A07-01	Hamadan University of Medical Sciences	Iran
	Simultaneous modelling of ordinal repeated measurements and multi-state data in the presence of cure fraction: A joint model	P-A07-02		

AWARDS & SUPPORT SCHEME

Beryl Ang'iro	Multivariate survival analysis for high dimensional data: A sample splitting approach	**P-A08-01**	Karatina University	Kenya
Ömer Faruk Dadas	A comparison between the joint model and landmark approach for dynamic prediction of competing risks in survival analysis with time-dependent covariates	**P-B08-03**	Ege University, Izmir	Turkey
Anand Hari	Impact of baseline hazards in jointly modelling longitudinal and terminal outcomes in cancer survival	**OC09-4**	Regional Cancer Centre, Thiruvananthapuram	India
Md Mynul Islam	Child malnutrition at Upazila Level in Bangladesh: A small-area estimation approach	**P-B09-02**	University of Dhaka	Bangladesh
Jagathnath Krishna Kumarapillai Mohanan Nair	Joint frailty modelling of multiple time-to-events and longitudinal measures	**OC27-2**	Regional Cancer Centre, Thiruvananthapuram	India
Lara Lewis	Impact of clinic community ART referral on service delivery for HIV patients newly initiated on antiretroviral therapy	**P-B02-17**	CAPRISA, Durban	South Africa
Shivani Saxena	Clustering high-dimensional spaces using a modified EM algorithm with fractional order assignment	**P-A08-08**	Institute of Advanced Research, Ahmedabad	India
Haoning Shen	Assessing treatment effects with adjusted restricted mean time lost in observational competing risks data	**OC27-5**	Southern Medical University, Guangzhou	China
Elaheh Talebi-Ghane	Simultaneous modelling of longitudinal measurements and multi-state data with a cured fraction using a joint model	**OC21-3**	Hamadan University of Medical Sciences	Iran
	A correlated joint frailty model for recurrent and terminal events: Application to breast cancer patients	**P-A07-11**		

KEYNOTE SPEAKERS

Keynote Speaker

Professor Sir David Spiegelhalter

Emeritus Professor of Statistics, University of Cambridge, United Kingdom

Monday 22 July 2024 | 09:30-10:30

Lecture Title: *Trustworthy communication of statistical evidence: What is it, and how we can get more of it*

President's Invited Speaker

Professor Chris Holmes

Professor of Biostatistics, Department of Statistics, University of Oxford, United Kingdom

Wednesday 24 July 2024 | 11:00-12:00

Lecture Title: *Biostatistics in the AI Era*

GENERAL INFORMATION A-Z

ABSTRACT BOOK

Abstracts presented during the ISCB45 Annual Conference are included in the Final Programme & Book of Abstracts published in electronic format and made available through the website and Conference App, as well as the ISCB website.

APP

Android devices

Search for The Event App by EventsAIR on PlayStore. When you are prompted for a code upon launching the App, insert **iscb45**

iOS devices

Search for The Event App by EventsAIR on AppStore. When you are prompted for a code upon launching the App, insert **iscb45**

CASH POINTS

Several banks with ATM machines (such as Eurobank, Alpha, Piraeus) are located in Vassilis Olgas Avenue, in close vicinity to the venue. Banks in Greece operate on weekdays from 08:00 to 14:00.

More specifically,

- ATM Piraeus Bank 650m
- ATM Alpha Bank 850m
- ATM National Bank of Greece 650m

However, all stores and restaurants accept credit/debit card payments.

CONFERENCE DATES

Sunday 21 July 2024

Pre-Conference Courses

Monday 22 July to Wednesday 24 July 2024

Main Conference

Thursday 25 July 2024

Mini-Symposia & ECB Day

CONGRESS VENUE

Thessaloniki Concert Hall

25 Martiou Street, 546 46 Thessaloniki, Greece

T. +30 231 089 5800

CORPORATE SOCIAL RESPONSIBILITY & SUSTAINABILITY

The International Society for Clinical Biostatistics (ISCB) is dedicated to fostering social responsibility by undertaking and supporting actions and initiatives that align with our identity and commitment to society and its people.

The **ISCB45** Conference advocates and champions the values of sustainability for society and its people. It implements a multitude of measures aimed at minimising the environmental impact of the event. The objective is to ensure that the positive impacts significantly outweigh the negative ones.

Key sustainability measures include:

- Selecting local suppliers to minimise long-distance transportation of goods and personnel
- Minimising printing and striving for a nearly paperless office environment
- Providing contactless check-in for delegates and visitors
- Reducing energy consumption through the use of LED lighting and screens
- Choosing catering partners with compost food and waste programmes
- Choosing Conference material that is produced by recycled materials

CURRENCY / EXCHANGE

The Greek currency is euro (EUR). Exchange offices are located all around the city centre (exchange offices and banks).

ELECTRICITY

The electrical power supply voltage in Greece is 220-240 Volts (US/Canada: 110-120 Volts).

GENERAL INFORMATION A-Z

EMERGENCY CONTACTS

Police: **100**
Fire department: **199**
Medical emergency (ambulance): **166**
European emergency contact number (all above): **112**

ENVIRONMENTAL POLICY

The ISCB45 Conference is observing an environmentally friendly policy. In this context, every effort has been made from preparation to realisation of the Conference, to minimise impact on the environment. The ISCB45 PCO observes an in-house recycling policy, and offices are hosted in energy saving premises.
During the ISCB45 Conference, no plastic badges will be used. All delegate material is environmentally friendly, where possible made from recycled material.

EXHIBITION OPERATING HOURS

All exhibitors are listed in the Programme book. The exhibition will run during Conference dates at the designated area, during official breaks.

SUNDAY	21 July 09:00-17:00
MONDAY	22 July 09:00-17:00
TUESDAY	23 July 09:00-13:30
WEDNESDAY	24 July 09:00-17:30

FACILITIES FOR PERSONS WITH IMPAIRED MOBILITY

The venue premises have been specifically designed to support the needs of persons with impaired mobility. All Conference areas are equipped with facilities that allow easy access, whilst specially designed restrooms, trolley bars and ramps are available in all levels.

FOOD & BEVERAGES

Coffee and refreshments, as well as Lunch during official breaks are included in the delegate registration fee and will be served in designated catering stations in the Exhibition Area.

HYBRID FORMAT

The ISCB45 is a Conference with physical attendance. However, it also provides delegates with the option to watch the Main Conference programme streamed live via the ISCB45 virtual platform, by clicking on the session's name, where they can find more information and the livestream itself.

Live Support: A Live Support system will be available for all attendees who will participate in the Conference, in order to resolve any technical issues that may arise and/or provide guidance through the platform.

Meeting Hub: You will be able to contact all participants and interact with your colleagues through the Meeting Hub section.

IMPORTANT NOTE: ISCB45 sessions will be recorded and made available for post- viewing for a period of five (5) months following Conference dates, that is until the end of year 2024.

INTERNET

Free WiFi access will be available in all Conference areas throughout the duration of the ISCB45 Conference.
Network: **iscb45**

LANGUAGE

English is the official language of the ISCB45 Conference. No simultaneous interpretation is provided.

LIABILITY AND INSURANCE

Registration fees do not include participants' insurance against personal accidents, sickness and cancellations by any party, theft, loss or damage to personal possessions.
Participants are requested to make their own arrangements with respect to health and travel insurance. Neither the ISCB, nor the Conference Local Organising Committee or the PCO Convin SA will accept any liability with this respect.

GENERAL INFORMATION A-Z

LOST & FOUND
A lost and found service is available, in close vicinity to the Registrations Desk.

PARKING
Thessaloniki Concert Hall has an underground parking, which is located in building M2. The parking is operating during announced events and the ticket is 4 EUR per day.

PHOTOS
Photos of sessions and Conference areas will be taken during the ISCB45 Conference. Attendees are advised to contact the Registrations Desk, for any issue or objection with this respect.

PRESS OFFICE
Members of the Press are welcome to register and receive the delegate material from the Registration Desk.
Conference material will be made available upon request and through the official social media of the ISCB.

SELF-CHECK-IN STATIONS
Conference delegates who do not have outstanding payments may collect their name badges through the Self-check-in stations, located near the Registrations Desk at the Conference Venue, thus avoiding queues for badge pick-up.

TELECOMMUNICATIONS
There are 3 main GSM operators in Greece: Cosmote, Vodafone, and Nova. The protocols for digital mobile telephone transmissions are based on GSM technology, operating at the frequencies of 900 and 1800 MHz. Please contact your provider for further details.

TIME
Thessaloniki follows Eastern European Summer Time (UTC/GMT +3 hours) at the time of the ISCB45 Conference.

TIPPING
In restaurants, hotels, taxis or other services, tipping is optional, and customised per case, depending on whether you are pleased with the service. Approximately 5% of the bill is a good guideline.

WEATHER
The weather in Thessaloniki in July is hot, with an average temperature of 25°C (77°F). July is also the month with the most daily sunshine hours, as well as the warmest seawater, with an average sea temperature of 26.2°C (79.2°F).
The usual daytime temperature is 32°C (89.6 °F) and at night the usual temperature is 21°C (69.8°F).

REGISTRATION

BADGES

Registration badges will be used during the ISCB45 Conference. Participants will receive their badges upon check-in at the Registrations Desk. For identification purposes and admission to scientific sessions, participants are requested to wear their badges at all times. Admission to Conference areas will not be allowed without badge identification.

CERTIFICATES OF ATTENDANCE

A Certificate of Attendance will be sent to all registered participants after the Conference via email. As part of this process, you will be asked to fill in an online Evaluation Form. The ISCB greatly values your input and kindly requests that you take the time to provide your feedback on your ISCB45 Conference experience.

REGISTRATION DESK

The Registration Desk is located near the entrance of the Conference areas.

Operating hours:

SUNDAY	21 July 08:00-17:30
MONDAY	22 July 08:30-17:30
TUESDAY	23 July 08:30-14:00
WEDNESDAY	24 July 08:30-18:00
THURSDAY	25 July 08:30-13:00

On Thursday 25 July, the Registration Desk will operate at the upper floor, near Parallel Halls 1 & 2.

ISCB OFFICE

The ISCB will have a representative at the Registration Desk, to staff their onsite Office. Operating hours will be the same as for the Registration Desk.

MAIN CONFERENCE REGISTRATION FEES

CATEGORY	ONSITE FEE	
	Member[1]	Non-Member[2]
Regular	650 EUR	740 EUR
Student[3]	320 EUR	360 EUR
Retired or Senior (65+ y.o.)[5]	380 EUR	-
Reduced rate[4]	380 EUR	420 EUR

Special Notes

[1] Member fees are applicable only if membership status is active. Membership status will be verified and followed up by the ISCB Office, operating onsite during Conference dates. Please address the ISCB Office at office@iscb.international for membership queries.

[2] By registering as a non-member for the ISCB45 Conference, you are provisionally considered a renewed member of the ISCB. If you do not wish to be considered an ISCB member, kindly alert the ISCB Office at office@iscb.international.

[3] For Students' Registration, a Valid Student's ID/ letter from Supervisor needs to be provided. The PCO CONVIN S.A. reserves the right to allocate your registration to the appropriate category, in case relevant documentation is missing or is not valid.

[4] Low-income and medium-low-income countries considered, according to the World Bank list.

[5] Retired or Senior registration category applies ONLY for ISCB members renewed for year 2024 under this membership category.

REGISTRATION

Registration Entitlements:

- Attendance of main Conference programme of 22-24 July 2024
- Attendance of Mini-Symposia / ECB Day programme of 25 July 2024 (registration at no additional cost)
- Access to Poster Area
- Access to Exhibition Area
- Coffee / Lunch breaks according to programme on 22-25 July 2024
- Conference documentation and attendance certificate
- Access to virtual platform and post-viewing rights for a period of five months
- Welcome reception on Monday 22 July 2024

PRE-CONFERENCE COURSES REGISTRATION FEES

CATEGORY	COURSE ONSITE FEE	
	Full Day	Half Day
Regular	370 EUR	250 EUR
Student	250 EUR	170 EUR
Retired or Senior (65+ y.o.)	250 EUR	170 EUR
Reduced rate	250 EUR	170 EUR

Additional Entitlements:

- Attendance of the Pre-Conference Course(s) of 21 July 2024 as selected and paid for
- Coffee / Lunch break according to programme on 21 July 2024
- Course attendance certificate

MINI SYMPOSIA REGISTRATION FEES

Attendance of Mini Symposia is included in the Main Conference registration fee. Below fees apply for those that will ONLY be attending a Mini Symposium on Thursday 25 July and not the Main Conference.

CATEGORY	ONSITE FEE
MINI-SYMPOSIUM 1: Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development	200 EUR
MINI-SYMPOSIUM 2: STRengthening Analytical Thinking for Observational Studies (STRATOS Initiative) – Recent Progress and Foci for the Future	200 EUR

Entitlements:

- Attendance of a Mini-Symposium of 25 July 2024 as selected and paid for
- Coffee break according to programme on 25 July 2024
- Mini-Symposium attendance certificate

INSTRUCTIONS FOR CHAIRS, SPEAKERS & ABSTRACT PRESENTERS

Instructions for session chairs, speakers & oral contributed presenters

All Conference session halls are equipped with the following:

- Projector & 16:9 screen (also suitable for 4:3 presentations)
- Chairs' table and microphones
- Lectern with laptop and wireless microphone
- Audience microphones for discussion purposes

No personal computers will be allowed.

All presentations should be handed in at the Speakers' Preview Desk (CR-2 Hall) latest two (2) hours prior to the respective session, or the day before for early morning presentations. We strongly advise presenters to check their presentations well in advance, to ensure content is depicted according to their wishes.

The Speakers' Preview Desk will be operating throughout Conference dates, according to the below schedule:

SUNDAY	21 July 08:30-18:00
MONDAY	22 July 08:30-17:30
TUESDAY	23 July 08:30-14:00
WEDNESDAY	24 July 08:30-18:00
THURSDAY	25 July 08:30-12:00

Chairs, Speakers and Oral Contributed presenters should be close to the stage and ready for the session at least 15 minutes before its beginning, to check their presence. Chairs are allocated at the panel desk, where fixed microphones will be provided to facilitate their assignment. Chairs are advised to be alert and observe time allocation, as they are responsible for the smooth running of their session and the overall programme flow. Two cards will be provided for the Chairs' use, in order to prompt speakers and presenters that their time is almost over: a yellow card indicating 5 minutes are left and a red card indicating that 1 minute is left.

Please make note that the SPC has decided each oral contributed presentation is allocated 18 minutes: 15 minutes for presentation and 3 minutes for Q&A. The programme already printed unfortunately does not reflect this unified format decided; please consult the App and e-version of the programme for time-slot and presentation order. All Chairs are to observe the same rule and make sure each presenter is given their 15 minutes presentation time, followed directly by 3 minutes of discussion (and not at the end of each session, as indicated in the printed programme). If there are no questions, Chairs should be prepared to discuss for the 3-minute slot allocated to each presentation.

Important note: Chairs are kindly requested not to change the order of presentations during each session, as delegates will not be able to plan ahead the talks they will attend. In case of a non-show, Chairs should be prepared to fill in the time slot with discussion time and should not move to the next presentation. The Scientific Programme Committee kindly requests that you also inform the Secretariat Desk (after your session is over) of any non-shows, as this item is reported to the ISCB after each Conference and affects any presentation certification requested.

Instructions for Poster Presenters

Posters will be displayed in tracks, being mounted and dismounted in the respective date and time.

Mounting date & time:
Poster Sessions Group A (Sessions A-01 to A-10):
Sun 21 Jul @ 17:00-18:00 & Mon 22 Jul @ 08:00-09:00
Poster Sessions Group B (Sessions B-01 to B-12):
Tue 23 Jul @ 12:30-13:30 & Wed 24 Jul @ 08:00-09:00

Display date & time:
Poster Sessions Group A (Sessions A-01 to A-10):
Mon 22 Jul from 09:00 to 17:00
Poster Sessions Group B (Sessions B-01 to B-12):
Wed 24 Jul from 09:00 to 17:00

Dismounting date & time:
Poster Sessions Group A (Sessions A-01 to A-10):
Tue 23 Jul @ 08:00-09:00
Poster Sessions Group B (Sessions B-01 to B-12):
Wed 24 Jul @ 17:30-18:30

Posters printing specifications: A0 dimension - Portrait (vertical) orientation
(maximum 84 cm horizontal x 120 cm vertical)

Presenters are reminded to bring their printed posters to the Conference venue. No printing service will be provided on site. The conference staff will provide materials for fixing posters to boards. The mounting and dismounting schedule will be strictly observed. The ISCB45 Conference will bear no responsibility for posters not mounted / dismounted according to the aforementioned schedule; any posters not removed as instructed will be discarded without further notice.

Poster presenters are kindly requested to be present at the Poster Area during official Lunch Breaks on the day of their presentation (official poster viewing time slot), in order to present their work to their peers and answer to any questions.

Official poster viewing time slots:

- GROUP A: Monday 22 July at 12:30-13:30
- GROUP B: Wednesday 24 July at 13:00-14:00

e-Poster PDF file

All poster presenters have been asked to provide an electronic version of their posters. The electronic versions of all the posters will be available throughout Conference dates, as well as during the post-viewing period, at the Virtual Platform e-Poster Gallery. There will be no option to present an e-poster through the ISCB45 platform.

SOCIAL PROGRAMME & NETWORKING

STUDENTS GATHERING EVENT

WHEN: Sunday, 21 July 2024 at 20:00-23:00
WHERE: ONOMA Hotel Cloud Bar (roof top)
Address: 24, Monastiriou street, 54629 Thessaloniki, Greece
WHAT: Networking opportunity for our Young Researchers, drinks and snacks offered
WHO: All delegates registered under a student category (included in registration fee)*
DRESS CODE: Smart casual
*Please check your attendance at the Conference Registrations Desk upon arrival, as capacity of the venue is limited.
HOW TO GET THERE: private arrangements (no transfer will be provided)

WELCOME RECEPTION

WHEN: Monday, 22 July 2024 at 20:00-23:00
WHERE: Porto Palace Hotel, URSA MINOR bar (roof top)
Address: 65, 26 October street, 546 28 Thessaloniki, Greece
WHAT: Escape in the city during a summer night and enjoy signature welcome cocktails, drinks and snacks, in an experience to remember
WHO: Included in registration fee
DRESS CODE: Smart casual
HOW TO GET THERE: Busses have been booked for transfer to and from the Welcome Reception venue. Departure at 19:30, pick up at Main Entrance of the Conference Venue. Return transfers to leave Porto Palace Hotel at 23:00, drop off outside the Main Entrance of the Conference Venue with a stop at Aristotelous square (city centre) for disembarking

SOCIAL PROGRAMME & NETWORKING

FREE AFTERNOON - CONFERENCE TOURS

WHEN: Tuesday, 23 July 2024
after completion of the
Scientific Programme, followed by
a lunch break (approximately 13:30 onwards)
WHO: Delegates that have pre-booked and paid
a tour as part of the Pre-registration's process.
Service not available for onsite booking.
TOURS OFFERED:

- "Macedonian Treasures" Vergina Tombs
- "Myth & History" Mt. Olympus, Litochoro & Old Panteleimon
- Swimming in Sithonia
- Thessaloniki City Tour & Archaeological Museum & Museum of Byzantine Culture
- Thessaloniki Walking Food Tour
- Gerovasileiou Wine & Dorodouli Tsipouro Tasting

CONFERENCE DINNER

WHEN: Wednesday, 24 July 2024 at 20:00-00:00
WHERE: Polis Convention Thessaloniki
Address: Thermi, 570 01
Thessaloniki, Greece
WHAT: Seated buffet dinner, followed by music
and dancing at an imposing building
with a great view and lush gardens.
WHO: Dinner ticket required at an additional fee
of 80 EUR. Ticket will be available until Mon
22 July during Lunch break
DRESS CODE: Smart casual
HOW TO GET THERE: Busses have been booked
for transfer to and from the Conference
Dinner venue. Departure at 19:15, pick
up at Main Entrance of the Conference
Venue. Return transfers to leave Polis
Convention Thessaloniki starting at 23:30,
last bus leaving at 00:00; drop off outside
the Main Entrance of the Conference
Venue with a stop at Aristotelous square
(city centre) for disembarking.

PROGRAMME OVERVIEW

SUNDAY 21 JULY 2024 | PRE-CONFERENCE COURSES

	PARALLEL 1 Maurice Saltiel I & II Hall	PARALLEL 2 Maurice Saltiel III Hall	PARALLEL 3 Library Hall	PARALLEL 4 CR-1 Hall
09:00-10:30	**PCC1** **Causal Inference in Clinical Trials** Session 1: Pre-specifying the estimand based on counterfactual outcomes	**PCC2** **Dynamic Predictions for Longitudinal and Time-to-Event Outcomes, with Applications in R** Session 1: Introduction of the framework of joint models for longitudinal and time-to-event data	**PCC3** **ROC Analysis for Classification and Prediction in Practice** Session 1: Fundamental concepts and metrics in the evaluation of diagnostic and predictive accuracy	**PCC4** **Adverse Events with Survival Time Outcome: Clinical Questions and Methods for Statistical Analysis based on Hazard Functions** Session 1: Adverse events as first event, requiring a competing risk/multistate framework, with a risk benefit interpretation
10:30-11:00	BREAK			
11:00-12:30	**PCC1** **Causal Inference in Clinical Trials** Session 2: Causal inference methods for treatment-policy estimand: Covariate adjustment	**PCC2** **Dynamic Predictions for Longitudinal and Time-to-Event Outcomes, with Applications in R** Session 2: Joint models to estimate and evaluate dynamic risk predictions for the settings of one event and competing risks	**PCC3** **ROC Analysis for Classification and Prediction in Practice** Session 2: Empirical and model-based estimation of an ROC curve/ Inferential procedures for a single ROC curve and its summary measures	**PCC4** **Adverse Events with Survival Time Outcome: Clinical Questions and Methods for Statistical Analysis based on Hazard Functions** Session 2: Direct impact of the treatment in the risk of developing the adverse event, requiring a latent time variable framework, with an interpretation on causality
12:30-13:30	LUNCH BREAK			

PROGRAMME OVERVIEW

13:30-15:00	**PCC1** **Causal Inference in Clinical Trials** Session 3: Introduction to DAGs	**PCC2** **Dynamic Predictions for Longitudinal and Time-to-Event Outcomes, with Applications in R** Session 3: Dynamic predictions for one longitudinal and one event outcome + JMbayes2 practical to fit joint models and estimate risk predictions (Lab1)	**PCC3** **ROC Analysis for Classification and Prediction in Practice** Session 3: Generalised linear modelling for ROC curves, optimal prediction with combinations of biomarkers	**PCC5** **Statistical and Practical Aspects of the Design and Analysis of Multi-Arm Multi-Stage (MAMS) Platform Trials** Session 1: Introduction to MAMS platform designs
15:00-15:30	BREAK			
15:30-17:00	**PCC1** **Causal Inference in Clinical Trials** Session 4: Causal inference for the hypothetical estimand: Time-varying confounding	**PCC2** **Dynamic Predictions for Longitudinal and Time-to-Event Outcomes, with Applications in R** Session 4: Dynamic predictions for multiple longitudinal outcomes and competing risks + JMbayes2 practical to fit joint models and estimate risk predictions (Lab2)	**PCC3** **ROC Analysis for Classification and Prediction in Practice** Session 4: Multiple-class ROC analysis	**PCC5** **Statistical and Practical Aspects of the Design and Analysis of Multi-Arm Multi-Stage (MAMS) Platform Trials** Session 2: Implementation of the statistical aspects of MAMs - Guidelines on the design and analysis of MAMs
20:00~	STUDENTS' GATHERING @ ONOMA HOTEL (OUTSIDE VENUE)			

PROGRAMME OVERVIEW

MONDAY 22 JULY 2024 | MAIN CONFERENCE – DAY 1

	PLENARY Emilios Riadis Hall	PARALLEL 1 Maurice Saltiel I & II Hall	PARALLEL 2 Maurice Saltiel III Hall	PARALLEL 3 Library Hall	PARALLEL 4 CR-1 Hall
09:00-10:30	**OPENING SESSION & KEYNOTE SPEAKER** **Sir David Spiegelhalter** Trustworthy communication of statistical evidence: What is it, and how we can get more of it				
10:30-11:00	COFFEE BREAK				
11:00-12:30	IS01 | Causal Inference and Machine Learning	OCS01 Clinical Trials I	OCS02 Joint Models I	OCS03 Methods for High Dimensional Data	OCS04 | Statistical Analysis for Complex Data Structures
12:30-13:30	LUNCH BREAK				
13:30-15:00	IS02 | Recent Advances in Survival Analysis with Complex Data Structures	OCS05 Clinical Trials II	OCS06 Meta-Analysis	OCS07 Statistics in Epidemiology I	OCS08 Diagnostics
15:00-15:30	COFFEE BREAK				
15:30-17:00	IS03 | Innovative Complex Adaptive Designs for Confirmatory Clinical Trials with Multiple Primary Research Questions	OCS09 | Joint Models II	OCS10 | Survival Analysis I	OCS11 | Causal Inference I	OCS12 | Prediction and Prognostic Models I
20:00~	WELCOME RECEPTION @ PORTO PALACE HOTEL (OUTSIDE VENUE)				

TUESDAY 23 JULY 2024 | MAIN CONFERENCE – DAY 2

	PLENARY Emilios Riadis Hall	PARALLEL 1 Maurice Saltiel I & II Hall	PARALLEL 2 Maurice Saltiel III Hall	PARALLEL 3 Library Hall	PARALLEL 4 CR-1 Hall
09:00-10:30	IS04 Optimal Individualised Treatment Rules	OCS13 Longitudinal Data Analysis I	OCS14 Artificial Intelligence and Machine Learning	OCS15 Integrative Data Analysis	OCS16 Multiple Testing /Miscellaneous
10:30-11:00	COFFEE BREAK				
11:00-12:30	IS05 Machine Learning Algorithms for Survival Analysis	OCS17 Causal Inference II	OCS18 Using Statistics to Improve Clinical Trials	OCS19 Imperfect Data	OCS20 Clinical Trials III
12:30-13:30	LUNCH BREAK				
13:30~	FREE AFTERNOON | CONFERENCE EXCURSIONS				

PROGRAMME OVERVIEW

WEDNESDAY 24 JULY 2024 | MAIN CONFERENCE – DAY 3

	PLENARY Emilios Riadis Hall	PARALLEL 1 Maurice Saltiel I & II Hall	PARALLEL 2 Maurice Saltiel III Hall	PARALLEL 3 Library Hall	PARALLEL 4 CR-1 Hall
09:00-10:30	IS06 | Bayesian Methods in Clinical Development	OCS21 | Joint Models III	OCS22 | Clinical Trials IV	OCS23 | Competing Risks I	OCS24 | Survival Analysis II
10:30-11:00	COFFEE BREAK				
11:00-12:00	**PRESIDENT'S INVITED SPEAKER** **Prof. Chris Holmes** Biostatistics in the AI Era				
12:00-13:30	**ISCB Annual General Meeting**				
13:30-14:00	LUNCH BREAK				
14:00-15:30	IS07 | Regulators' View of Randomised and Non-Randomised Evidence in Drug Development	OCS25 Longitudinal Data Analysis II	OCS26 Prediction and Prognostic Models II	OCS27 Competing Risks II	OCS28 Electronic Health Data
15:30-16:00	COFFEE BREAK				
16:00-17:30	IS08 | Combining Real-World Data and Randomised Clinical Trials	OCS29 Prediction and Prognostic Models III	OCS30 Survival Analysis III	OCS31 Clinical Trials V	OCS32 Statistics in Epidemiology II
20:00~	CONFERENCE DINNER @ POLIS CONVENTION CENTRE (OUTSIDE VENUE)				

THURSDAY 25 JULY 2024
MINI-SYMPOSIA & EARLY CAREER BIOSTATISTICIANS' DAY

	PARALLEL 1 Maurice Saltiel I & II Hall	PARALLEL 2 Maurice Saltiel III Hall	PARALLEL 3 Library Hall
09:00-10:30	MS1: Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development *(Part I)* In collaboration with the ISCB SiRA SC	MS2: STRengthening Analytical Thinking for Observational Studies (STRATOS Initiative)-Recent Progress and Foci for the Future *(Part I)* In collaboration with the STRATOS Steering Group	EARLY CAREER BIOSTATISTICIANS' DAY *(Part I)* Organised by the ISCB ECB SC
10:30-11:00	COFFEE BREAK		
11:00-12:30	MS1: Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development *(Part II)* In collaboration with the ISCB SiRA SC	MS2: STRengthening Analytical Thinking for Observational Studies (STRATOS Initiative)-Recent Progress and Foci for the Future *(Part II)* In collaboration with the STRATOS Steering Group	EARLY CAREER BIOSTATISTICIANS' DAY *(Part II)* Organised by the ISCB ECB SC

DETAILED PROGRAMME

SUNDAY, 21 JULY 2024

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

09:00-17:00 **PRE-CONFERENCE COURSE 1:** Causal Inference in Clinical Trials

COURSE INSTRUCTORS: Kelly Van Lancker (Ghent BE), **Alex Ocampo** (Basel CH)

SUMMARY: Causal thinking and inference have gained increasing attention in global drug development in light of the recently published ICH E9(R1) guideline on estimands and sensitivity analysis (2019) and the FDA guideline on covariate adjustment (2023). Even so, causal inference remains somewhat of a mystery to many. The goal of this course is to provide insight into the use of causal thinking and methods to better inform decision making. We introduce causal inference methods based on different drug development settings, including estimands, covariate adjustment, and the use of external control data.

PREREQUISITES: This introductory course is aimed at researchers in the pharmaceutical industry and academia working with clinical trial data; it does not demand prior familiarity with causal inference. We foresee a mix of lectures and hands-on exercises. The participants will strengthen their understanding of the concepts and methods explained during the lectures by analyzing real clinical trial data sets during the practical sessions using R

09:00-10:30	Session 1	**Pre-specifying the estimand based on counterfactual outcomes** **OUTLINE:** Important to enable the discussion of causal inference problems is the counterfactual/potential outcomes model. In this session we introduce this framework and explain how it can be helpful to define (causal) estimands.
10:30-11:00		BREAK
11:00-12:30	Session 2	**Causal inference methods for treatment-policy estimand: Covariate adjustment** **OUTLINE:** Covariate adjustment methods have been claimed to yield more powerful intention-to-treat analyses of randomized trials, at no 'cost'. The aim of this session is to provide insight into covariate adjustment: how it succeeds to gain power, and when and how it can be used. In this session we will also cover targeted maximum likelihood estimation. We will also shortly discuss transportability and the use of historical controls.
12:30-13:30		LUNCH BREAK
13:30-15:00	Session 3	**Introduction to DAGs** **OUTLINE:** CAnother important framework, besides counterfactual outcomes, to enable the discussion of causal inference problems is Directed Acyclic Graphs (DAGs). In this session we explain how to use DAGs to represent the causal relationship that we believe exist between the variables of interest. We also discuss how one can recognize problems of selection or confounding bias. In the hands-on part the participants will discover why standard adjustment for time-varying covariates does not provide a valid adjustment for time-dependent confounding, what instrumental variables are and how to utilize them for detecting causal effects.
15:00-15:30		BREAK
15:30-17:00	Session 4	**Causal inference for the hypothetical estimand: Time-varying confounding** **OUTLINE:** Starting from different hypothetical estimands (e.g., due to treatment switching), we review different methods to adjust for time-varying confounding.

SUNDAY, 21 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

09:00-17:00

PRE-CONFERENCE COURSE 2: Dynamic Predictions for Longitudinal and Time-to-Event Outcomes, with Applications in R

COURSE INSTRUCTORS: Dimitris Rizopoulos (Rotterdam NL), **Christos Thomadakis** (Athens GR)

SUMMARY: This course focuses on data collected in follow-up studies. Outcomes from these studies typically include longitudinally measured responses (e.g., biomarkers) and the time until an event of interest occurs (e.g., death, dropout). The aim is often to utilize longitudinal information to predict the risk of the event. An important attribute of these predictions is their time-dynamic nature, i.e., they are updated each time new longitudinal measurements are recorded. In this course, we will introduce the framework of joint models for longitudinal and time-to-event data and explain how it can be used to estimate and evaluate such dynamic risk predictions for the settings of one event and competing risks. We will use the R package JMbayes2 to showcase the capabilities of these models.

PREREQUISITES: This course assumes knowledge of basic statistical concepts, such as regression models and standard statistical inference using maximum likelihood and Bayesian methods. Also, a basic knowledge of R would be beneficial but is not required. Participants are required to bring their laptops with the battery fully charged. Before the course, instructions will be sent for installing the required software.

09:00-10:30	**Session 1**	Introduction of the framework of joint models for longitudinal and time-to-event data
10:30-11:00		BREAK
11:00-12:30	**Session 2**	Joint models to estimate and evaluate dynamic risk predictions for the settings of one event and competing risks
12:30-13:30		LUNCH BREAK
13:30-15:00	**Session 3**	Dynamic predictions for one longitudinal and one event outcome + JMbayes2 practical to fit joint models and estimate risk predictions (Lab1)
15:00-15:30		BREAK
15:30-17:00	**Session 4**	Dynamic predictions for multiple longitudinal outcomes and competing risks + JMbayes2 practical to fit joint models and estimate risk predictions (Lab2)

SUNDAY, 21 JULY 2024

PARALLEL HALL 3
(Library Hall)

09:00-17:00

PRE-CONFERENCE COURSE 3:
ROC Analysis for Classification and Prediction in Practice

COURSE INSTRUCTORS: Christos Nakas (Volos GR; Bern CH), **Constantine A. Gatsonis** (Providence RI US)

SUMMARY: This course will be based on our recently released book on modern ROC methodology (Nakas et al, CRC Press, May 2023). We will examine the conceptual underpinning and sound interpretation of ROC analysis and will discuss the practical implementation ROC methods in diverse scientific fields, with emphasis on the evaluation of biomarkers, imaging modalities, and machine learning tools. Examples will accompany the methodologic discussion using standard statistical software such as R, Matlab and STATA. The course will draw on developments in the field during the past two decades, which was a period of intensive growth in both the methods and the applications of ROC analysis.The instructors will provide a contemporary, integrated exposition of ROC methodology for both classification and prediction and survey methods for multiple-class ROC. The course material would be of interest to graduate level students and researchers from a wide range of disciplines and different industries such as diagnostic medicine, bioinformatics, medical physics, and perception psychology.

PREREQUISITES: Graduate level course intended for researchers and students with background in the fundamentals of statistical methods, including estimation, hypothesis testing, linear and generalized linear modeling. Some familiarity with basic concepts from the evaluation of diagnosis and prediction would be helpful.

09:00-10:30	**Session 1**	Fundamental concepts and metrics in the evaluation of diagnostic and predictive accuracy
10:30-11:00		BREAK
11:00-12:30	**Session 2**	Empirical and model-based estimation of an ROC curve/ Inferential procedures for a single ROC curve and its summary measures
12:30-13:30		LUNCH BREAK
13:30-15:00	**Session 3**	Generalised linear modelling for ROC curves, optimal prediction with combinations of biomarkers
15:00-15:30		BREAK
15:30-17:00	**Session 4**	Multiple-class ROC analysis

SUNDAY, 21 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

09:00-12:30

PRE-CONFERENCE COURSE 4: Adverse Events with Survival Time Outcome: Clinical Questions and Methods for Statistical Analysis based on Hazard Functions

COURSE INSTRUCTOR: Laura Antolini (Milan IT)

SUMMARY: In the study of novel treatments with a survival time outcome the analysis of the occurrence of adverse events is of crucial importance since it may imply the treatment interruption or, if fatal or even the patient death. This justifies the throughout effort in data collection on adverse events that is surprisingly often followed by statistical analyses only based on descriptive methods, such as simple proportions. In this course the analysis of adverse events will be tackled starting from the clinical question, followed by the identification of the estimand of interest and methods for estimation and Inference. Two different approaches will be discussed: 1. the observed occurrence of AE as first event, requiring a competing risk/multistate framework, with a risk benefit interpretation 2. the direct impact of the treatment in the risk of developing the adverse event, requiring a latent time variable framework, with an interpretation on causality

OUTLINE: The course will start by a critically review the commonly used standard theoretical quantities and estimators with reference to their appropriateness for dealing with approaches 1 or 2 in the presence of non fatal and fatal AE. Then the course will follow with the presentation of the suitable methods to address approaches 1 and 2. The theoretical explanation will be paralleled by motivating applications using the softwares Stata and R with data and code fragments made available to students.

TARGET AUDIENCE: Applied biostatisticians/epidemiologists or graduate students familiar with basic survival analysis.

09:00-10:30	**Session 1**	Adverse events as first event, requiring a competing risk/multistate framework, with a risk benefit interpretation
10:30-11:00		BREAK
11:00-12:30	**Session 2**	Direct impact of the treatment in the risk of developing the adverse event, requiring a latent time variable framework, with an interpretation on causality

SUNDAY, 21 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

13:30-17:00

PRE-CONFERENCE COURSE 5: Statistical and Practical Aspects of the Design and Analysis of Multi-Arm Multi-Stage (MAMS) Platform Trials

COURSE INSTRUCTORS: Babak Choodari-Oskooei (London UK), **Max (Mahesh) Parmar** (London UK)

SUMMARY: This course aims to help participants:

- Understand the motivation behind these designs;
- Learn how to choose the design parameters and stopping boundaries, both for lack-of-benefit and efficacy;
- Learn how to deal with overwhelming efficacy;
- Learn about stopping randomisation to research arms;
- Learn how to add a new research arm, and how to control Type I and II error rates in both pre-planned and unplanned addition of a new arm;
- Learn about MAMS designs in which arms are ranked and selectively chosen to continue.

MAMS PLATFORM TRIALS: Typically, in these protocols, randomisation is stopped to insufficiently active treatment arms at interim stages and new research arms can be added during the course of the trial. The MAMS approach is one of the few adaptive designs being deployed in a number of trials and across a range of disease in the phase III setting, including STAMPEDE (prostate cancer), CompARE (TB), TRUNCATE-TB (TB), RAMPART (renal cancer), and ROSSINI-II (wound surgery).

OUTLINE: This half-day course consists of two main sessions, and four lectures in total. The first session introduces MAMS platform designs. The second session focuses on the implementation of the statistical aspects of such trials and provides guidelines on the design and analysis of such trials. It will also explore further design issues such as adding new research arms, and designs in which research arms are ranked and selectively chosen to continue.

13:30-15:00	**Session 1**	Introduction to MAMS platform designs
15:00-15:30		BREAK
15:30-17:00	**Session 2**	Implementation of the statistical aspects of MAMs - Guidelines on the design and analysis of MAMss

MONDAY, 22 JULY 2024

PLENARY HALL (Emilios Riadis Hall)

09:00-10:30 OPENING SESSION & KEYNOTE SPEAKER: Trustworthy communication of statistical evidence: What is it, and how we can get more of it

CHAIR: Vana Sypsa (ISCB45 LOC Chair)

OPENING SESSION	
09:00-09:10	Welcome Address by ISCB45 LOC Chair, **Prof. Vana Sypsa**
09:10-09:20	Welcome Address by ISCB45 SPC Chair, Prof. Giota Touloumi
09:20-09:30	Addresses by Representatives of • The Aristotle University of Thessaloniki School of Medicine • The National & Kapodistrian University of Athens School of Medicine • The Region of Central Macedonia • The City of Thessaloniki
09:30-09:40	Presentation of ISCB45 Student Conference Awards and Conference Awards for Scientists
KEYNOTE LECTURE	
09:40-09:45	Introduction of Keynote Speaker
09:45-10:30	KEYNOTE LECTURE: Trustworthy communication of statistical evidence: What is it, and how we can get more of it Prof. Sir David Spiegelhalter (Cambridge UK)
10:30-11:00	COFFEE BREAK

PLENARY HALL (Emilios Riadis Hall)

11:00-12:30 INVITED SESSION 1: Causal Inference and Machine Learning

CHAIR: Els Goetghebeur (Ghent BE)

11:00-11:30	IS1-1	**Erin Evelyn Gabriel** Copenhagen DK	Double robust: A great asset not to be constructed lightly
11:30-12:00	IS1-2	**Martin Spindler** Hamburg DE	Causal machine learning with DoubleML: An introduction and applications
12:00-12:30	IS1-3	**Maya L. Petersen** Berkeley CA US	Efficient and robust machine-learning-based approaches for simple, cluster randomised, and sequential multiple assignment randomised trial analysis: Illustrations from HIV trials in East Africa

PARALLEL HALL 1 (Maurice Saltiel I & II Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 01: Clinical Trials I

CHAIR: Nicky Best (London UK)

11:00-11:18	OC01-1	**Linxi Han** Bristol UK	SEMAP-curvature: A model-free Bayesian approach for analysis of dose-finding trials
11:18-11:36	OC01-2	**Chiara Micoli** Stockholm SE	Bayesian predictive probability of success for interim monitoring of clinical trials with competing risks data
11:36-11:54	OC01-3	**Alexander Ooms** Oxford UK	A real-time Bayesian re-design of a prostate cancer detection randomised controlled trial
11:54-12:12	OC01-4	**Christopher Selman** Parkville AU	Evaluating the performance of Bayesian cumulative logistic models in randomised controlled trials: A simulation study
12:12-12:30	OC01-5	**Robert Mahar** Melbourne AU	Approximate Bayesian-frequentist power in platform trial design

MONDAY, 22 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 02: Joint Models I

CHAIR: Loukia Meligkotsidou (Athens GR)

11:00-11:18	OC02-1	**Pasquale Dolce** Naples IT	PLS structural equation modelling and quantile composite-based path modelling for medical and healthcare research
11:18-11:36	OC02-2	**Marco Palma** Cambridge UK	Bayesian joint location-scale model for time-to-event and multivariate longitudinal data with association based on within-individual variability
11:36-11:54	OC02-3	**Andreas Markoulidakis** Bristol/Edinburg UK	Using a Multi-Parameter Estimation of Prevalence (MPEP) model to estimate the prevalence of opioid dependence in Scotland
11:54-12:12	OC02-4	**Christos Thomadakis** Athens GR	Joint modelling of longitudinal and competing-risk data under failure cause and non-informative right censoring misclassification
12:12-12:30	OC02-5	**Rehema Ouko** London UK	Bayesian joint modelling of foetal growth measurements and their associations with adverse neonatal outcomes

PARALLEL HALL 3
(Library Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 03: Methods for High Dimensional Data

CHAIR: Denitsa Grigorova (Sofia BG)

11:00-11:18	OC03-1	**Alessia Mapelli** Milan IT	Personalised protein-protein interaction networks via weighted conditional GGMs
11:18-11:36	OC03-2	**Nuria Senar Villadeamigo** Amsterdam NL	Sparse canonical correlation for multiple measurements with latent variable trajectories
11:36-11:54	OC03-3	**Leiv Rønneberg** Cambridge UK	Biomarker discovery for cancer drug-sensitivity screens via multi-output Gaussian Processes
11:54-12:12	OC03-4	**Dominic Edelmann** Heidelberg DE	Distance correlation methods for genome-wide association studies
12:12-12:30	OC03-5	**Charlotte Castel** Oslo NO	Comparison of the LASSO and Integrative LASSO with Penalty Factors (IPF-LASSO) methods for multi-omics data: Variable selection with error control

MONDAY, 22 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 04: Statistical Analysis for Complex Data Structures

CHAIR: Lehana Thabane (Hamilton CA)

11:00-11:18	OC04-1	**Eleni-Rosalina Andrinopoulou** Rotterdam NL	Unveiling the impact of social and environmental determinants of health on lung function decline in cystic fibrosis through data integration using the US Registry
11:18-11:36	OC04-2	**Nastaran Sharifian** Galway IE	Automating functional data analysis in real time: An application to pressure sensor data in the treatment of venous leg ulcers
11:36-11:54	OC04-3	**Fabiola Del Greco Miglianico** Bolzano/Bozen IT	Machine learning algorithm to predict biomarker levels using metabolomics data
11:54-12:12	OC04-4	**Cristina Ehrmann** Nottwil CH	Generalised partial credit model with covariates: Comparing the efficacy of two implementations in a Bayesian framework
12:12-12:30	OC04-5	**Patric Tippmann** Freiburg DE	Domain adaptation approaches for harmonising multi-site tabular data

12:30-13:30 LUNCH BREAK

PLENARY HALL
(Emilios Riadis Hall)

13:30-15:00 INVITED SESSION 2: Recent Advances in Survival Analysis with Complex Data Structures

CHAIR: Giorgos Bakoyannis (Indianapolis IN US)

13:30-14:00	IS2-1	**Eleni-Rosalina Andrinopoulou** Rotterdam NL	Joint modelling of (un)bounded longitudinal markers, competing risks, and recurrent events in cystic fibrosis data
14:00-14:30	IS2-2	**Ying Zhang** Omaha NE US	Semiparametric estimation of misclassified semi-competing risks data under gamma-frailty conditional Markov model
14:30-15:00	IS2-3	**Dipankar Bandyopadhyay** Richmond VA US	Bayesian semiparametric modelling of spatially-referenced multistate current status data

MONDAY, 22 JULY 2024

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

13:30-15:00 ORAL CONTRIBUTED SESSION 05: Clinical Trials II

CHAIR: Thomas Jaki (Regensburg DE)

13:30-13:48	OC05-1	**Libby Daniells** Lancaster UK	Flexible borrowing of historical information in a basket trial via the exchangeability-nonexchangeability (EXNEX) model
13:48-14:06	OC05-2	**Marco Ratta** Turin IT	Dual criteria approach to early conditional approval in time-to-event group sequential trials via historical borrowing
14:06-14:24	OC05-3	**Juliana Schneider** Potsdam DE	Multimodal outcomes in N-of-1 trials: A simulation study
14:24-14:42	OC05-4	**Eline Anslot** Ghent BE	Improving interim decisions for single arm trials by adjusting for baseline covariates and short-term endpoints
14:42-15:00	OC05-5	**Thomas Gaertner** Potsdam DE	Adjusting for time-dependencies in N-of-1 trials: A comparative analysis of statistical methods

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

13:30-15:00 ORAL CONTRIBUTED SESSION 06: Meta-Analysis

CHAIR: Anna-Bettina Haidich (Thessaloniki GR)

13:30-13:48	OC06-1	**David Phillippo** Bristol UK	multinma: A comprehensive R package for network meta-analysis of survival outcomes with aggregate data, individual patient data, or a mixture of both
13:48-14:06	OC06-2	**Noosheen Rajabzadehtahmasebi** Freiburg DE	Path-based approach for detecting and assessing inconsistency in network meta-analysis: A novel method
14:06-14:24	OC06-3	**Theofanis Ioannidis** Thessaloniki GR	Extensions of threshold analysis in network meta-analysis
14:24-14:42	OC06-4	**Emmanouil Androulakis** Piraeus GR	Meta-analysis of time-to-event data under model misspecification, based on median ratios
14:42-15:00	OC06-5	**Samuel Perren** Bristol UK	Extending multilevel network meta-regression to disconnected networks and single-arm studies: A case study on plaque psoriasis treatments

MONDAY, 22 JULY 2024

PARALLEL HALL 3
(Library Hall)

13:30-15:00 ORAL CONTRIBUTED SESSION 07: Statistics in Epidemiology I

CHAIR: Klea Katsouyanni (Athens GR)

13:30-13:48	OC07-1	**Ellen Bradley** London UK	Modelling of global estimates: An application of Bayesian modelling for low birthweight and preterm births
13:48-14:06	OC07-2	**Vasiliki Engeli** Athens GR	Estimating smooth age-specific contact rates in the population of Greece during the COVID-19 pandemic
14:06-14:24	OC07-3	**Lander Rodriguez** Bilbao ES	Dissecting COVID-19's spread in the Basque Country: A Bayesian spatio-temporal analysis
14:24-14:42	OC07-4	**Christos Chalitsios** Ioannina GR	A data-driven approach for studying the interplay between psoriasis and multiple conditions: A phenome-wide association and Mendelian randomisation study
14:42-15:00	OC07-5	**Alessandro Grosso** Antwerp BE	Heterogeneity in the acquisition of multiple infections: A mathematical and statistical modelling perspective

PARALLEL HALL 4
(CR-1 Hall)

13:30-15:00 ORAL CONTRIBUTED SESSION 08: Diagnostics

CHAIR: Christos Nakas (Volos GR; Bern CH)

13:30-13:48	OC08-1	**Sunil Mathur** Houston TX US	Improving hypotheses testing with external information knowledge for achieving higher efficiency in non-normal population
13:48-14:06	OC08-2	**Ainesh Sewak** Zurich CH	Construction and evaluation of optimal diagnostic tests
14:06-14:24	OC08-3	**Leire Garmendia Berges** Bilbao ES	Development of a time-dependent AUC estimator for competing risks models
14:24-14:42	OC08-4	**Nia Kang** Montreal CA	Community-driven Bayesian inference for sociocultural questionnaire instrument validation
14:42-15:00	OC08-5	**Juan Carlos Pardo-Fernandez** Vigo ES	Methods for comparing ROC curves under the presence of covariates

15:00-15:30 COFFEE BREAK

MONDAY, 22 JULY 2024

PLENARY HALL
(Emilios Riadis Hall)

15:30-17:00 INVITED SESSION 3: Innovative Complex Adaptive Designs for Confirmatory Clinical Trials with Multiple Primary Research Questions

CHAIR: Babak Choodari-Oskooei (London UK), **Ian White** (London UK)

15:30-15:55	IS3-1	**Cyrus Mehta** Cambridge MA US	Graph based adaptive MAMS designs for trials with multiple endpoints
15:55-16:20	IS3-2	**Babak Choodari-Oskooei** London UK	Pairwise and familywise error rate control in platform trials: Impact on sample size, trial timelines and analysis
16:20-16:45	IS3-3	**Annette Kopp-Schneider** Heidelberg DE	Using Bayesian methods to include non-concurrent controls in the analysis of platform trials: Benefits and limitations
16:45-17:00	Discussion	**Max (Mahesh) Parmar** London UK	

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

15:30-17:00 ORAL CONTRIBUTED SESSION 09: Joint Models II

CHAIR: Jan Beyersmann (Ulm DE)

15:30-15:48	OC09-1	**Nina Van Gerwen** Rotterdam NL	Combining and optimising dynamic predictions with Super Learner
15:48-16:06	OC09-2	**Teun Petersen** Rotterdam NL	Personalised monitoring schedules for multiple types of measurements with application in chronic heart failure
16:06-16:24	OC09-3	**Pedro Miranda Afonso** Rotterdam NL	Dynamically predicting survival benefit to prioritise liver transplantation in hepatocellular carcinoma patients
16:24-16:42	OC09-4	**Anand Hari** Thiruvananthapuram IN	Impact of baseline hazards in jointly modelling longitudinal and terminal outcomes in cancer survival
16:42-17:00	OC09-5	**Andrea Mario Vergani** Milan IT	Diagnosis of cardiovascular diseases using interpretable cardiac magnetic resonance-derived latent factors

MONDAY, 22 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

15:30-17:00 ORAL CONTRIBUTED SESSION 10: Survival Analysis I

CHAIR: Kleio Kipourou (London UK)

15:30-15:48	OC10-1	**Damjan Manevski** Ljubljana SI	Regression modelling in an extended multi-state model using relative survival
15:48-16:06	OC10-2	**Molly Wells** Leicester UK	Use of pseudo-observations in the development and validation of prognostic models for competing risks
16:06-16:24	OC10-3	**Theofanis Balanos** Indianapolis IN US	Semiparametric regression analysis of misclassified competing risk data
16:24-16:42	OC10-4	**Oskar Laverny** Marseille FR	Non-parametric estimation of net survival under dependence between death causes
16:42-17:00	OC10-5	**Kathrin Möllenhoff** Cologne DE	Survival analysis under non-proportional hazards: Investigating non-inferiority or equivalence in time-to-event data

PARALLEL HALL 3
(Library Hall)

15:30-17:00 ORAL CONTRIBUTED SESSION 11: Causal Inference I

CHAIR: Geir Egil Eide (Bergen NO)

15:30-15:48	OC11-1	**Pantelis Samartsidis** Cambridge UK	A Bayesian multivariate factor analysis model for policy evaluation from time-series data
15:48-16:06	OC11-2	**Georgios Kazantzidis** Basel CH	Quantifying the proportion of treatment effect based on joint models in tumour kinetics and survival
16:06-16:24	OC11-3	**Lizbeth Burgos-Ochoa** Tilburg NL	Harnessing survey data for causal inference: Generalising treatment effects in observational studies
16:24-16:42	OC11-4	**Claus Ekstrom** Copenhagen DK	Comparison of causal discovery approaches to utilise temporal information for life-course Epidemiology
16:42-17:00	OC11-5	**Andrea Callegaro** Ottignies BE	Covariate-adjusted Robust Mixture Prior approach in clinical trials with historical controls

MONDAY, 22 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

15:30-17:00 ORAL CONTRIBUTED SESSION 12: Prediction and Prognostic Models I

CHAIR: Laure Wynants (Maastricht NL)

15:30-15:48	OC12-1	**Willi Sauerbrei** Freiburg DE	The Linearity Assumption: Unravelling its impact on prediction accuracy in multivariable models
15:48-16:06	OC12-2	**Alex Carriero** Utrecht NL	The harms of class imbalance corrections for calibration in machine learning: A simulation study
16:06-16:24	OC12-3	**Yuxuan Jin** Cleveland OH US	Prediction models at the crossroads of statistical inference and machine learning
16:24-16:42	OC12-4	**Aiden Smith** Leicester UK	Assessing the impact on predictive performance of prognostic models when missing data mechanisms vary across model development phases: A simulation study
16:42-17:00	OC12-5 StCA	**Haya Elayan** Manchester UK	Correcting for case-mix shift when developing clinical prediction models

20:00~ **WELCOME RECEPTION AT PORTO PALACE HOTEL**

TUESDAY, 23 JULY 2024

PLENARY HALL
(Emilios Riadis Hall)

09:00-10:30 INVITED SESSION 4: Optimal Individualised Treatment Rules

CHAIRS: Giorgos Bakoyannis (Indianapolis IN US), **Rodolphe Thiébaut** (Bordeaux FR)

09:00-09:30	IS4-1	**Menggang Yu** Ann Arbor MI US	Robust sample weighting to facilitate individualized treatment rule learning for a target population
09:30-10:00	IS4-2	**Sukjin Han** Bristol UK	Policy learning with distributional welfare
10:00-10:30	IS4-3	**Hyung Park** New York NY US	Functional additive models for interaction effects between a treatment and functional covariates

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 13: Longitudinal Data Analysis I

CHAIR: Jeremy Taylor (Ann Arbor MI US)

09:00-09:18	OC13-1	**Chiara Degan** Leiden NL	Multivariate longitudinal modelling with non-normally distributed outcomes and endogenous covariates
09:18-09:36	OC13-2	**Marion Leila Claudine Kerioui** Cambridge UK	A Bayesian partition model for high-dimensional functional data
09:36-09:54	OC13-3	**Roula Tsonaka** Leiden NL	Inference in longitudinal data analysis with terminating events
09:54-10:12	OC13-4	**Davood Roshan Sangachin** Galway IE	Challenge your limits!
10:12-10:30	OC13-5 StCA	**Awan Afiaz** Seattle WA US; Dhaka BD	Optimal sandwich variance estimator in penalised GEE for nearly separated longitudinal binary data with small samples

TUESDAY, 23 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 14: Artificial Intelligence and Machine

CHAIR: Claus Ekstrom (Copenhagen DK)

09:00-09:18	OC14-1	**Mark Clements** Stockholm SE	Continuous time Markov multistate models for health economics, with an application to artificial intelligence assisted pathology
09:18-09:36	OC14-2	**Konstantina Thaleia Pilali** Padua IT	UBEP Ethical Assessment Tool in clinical studies
09:36-09:54	OC14-3	**Manel Rakez** Bordeaux FR	Enhancing breast cancer prediction with statistical and deep learning approaches for longitudinal imaging data
09:54-10:12	OC14-4	**Mehran Moazeni** Utrecht NL	Real-time detection of Atrial fibrillation using meta-learning in the Amsterdam UMC Database
10:12-10:30	OC14-5	**Fabian Kabus** Freiburg DE	An end-to-end modelling approach for capturing spatiotemporal patterns in two-photon imaging data

PARALLEL HALL 3
(Library Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 15: Integrative Data Analysis

CHAIR: Philip S. Boonstra (Ann Arbor MI US)

09:00-09:18	OC15-1	**Darren Scott** Cambridge UK	Borrowing from historical control data in a Bayesian time-to-event model with flexible baseline hazard function
09:18-09:36	OC15-2	**Nusrat Rabbee** San Francisco CA US	Incorporating machine learning with network analysis for imbalanced multi-class classification
09:36-09:54	OC15-3	**Myra McGuinness** Melbourne AU	Quantifying between-study heterogeneity in one-stage individual participant data meta-analyses of treatment-covariate interactions: A simulation study
09:54-10:12	OC15-4	**Marie Analiz April Limpoco** Hasselt BE	Federated mixed effects logistic regression based on one-time shared summary statistics
10:12-10:30	OC15-5	**Theodoros Evrenoglou** Freiburg DE; Paris FR	Producing treatment hierarchies in network meta-analysis using probabilistic models and treatment choice criteria

TUESDAY, 23 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 16: Multiple Testing / Miscellaneous

CHAIR: Dimitris Karlis (Athens GR)

09:00-09:18	OC16-1	**Robin Ristl** Vienna AT	A comparison of multivariate hypothesis tests for longitudinal outcomes
09:18-09:36	OC16-2	**Christina Yap** London UK	Enhancing excellence in early phase dose-finding trials through SPIRIT and CONSORT Dose-FINding Extensions (DEFINE) guidance
09:36-09:54	OC16-3	**Lan Kelly** Adelaide AU	Cluster summary analysis of cluster randomised crossover trials: Advantages and disadvantages
09:54-10:12	OC16-4	**Matthias Becher** Berlin DE	Multiple contrast tests under variance heteroscedasticity in general factorial designs
10:12-10:30	OC16-5	**Christiana Kartsonaki** Oxford UK	Some approximations to the path formula for some non-linear models

10:30-11:00 COFFEE BREAK

PLENARY HALL
(Emilios Riadis Hall)

11:00-12:30 INVITED SESSION 5: Machine Learning Algorithms for Survival Analysis

CHAIRS: Michal Abrahamowicz (Montreal CA), **Dimitris Rizopoulos** (Rotterdam NL)

11:00-11:30	IS5-1	**Malka Gorfine** Tel Aviv IL	Machine learning procedures in survival analysis
11:30-12:00	IS5-2	**Harald Binder** Freiburg DE	How to leverage the success of ChatGPT for longitudinal and time-to-event data
12:00-12:30	IS5-3	**Qixian Zhong** Xiamen CN	Hypothesis testing for the Deep Cox Model

TUESDAY, 23 JULY 2024

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 17: Causal Inference II

CHAIR: Issa Dahabreh (Boston MA US)

11:00-11:18	OC17-1	**Jean-Baptiste Baitairian** Paris FR	Sensitivity analysis of the average potential outcome with unobserved confounders
11:18-11:36	OC17-2	**Ilaria Prosepe** Leiden NL	Bridging the gap between multi-state modelling and causal inference to estimate the effect of treatment delay
11:36-11:54	OC17-3	**Fenny Ong** Leuven BE	Surrogate endpoint evaluation based on causal inference and information theory in binary-continuous setting
11:54-12:12	OC17-4	**Felix Clouth** Tilburg NL	The parametric G-formula for latent Markov models
12:12-12:30	OC17-5	**Anca Chis Ster** London UK	A new estimator and Stata command for estimating causal mediation effects with non-adherence in the presence of missing data

PARALLEL HALL 2
(Maurice Saltiel III Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 18: Using Statistics to Improve Clinical Trials

CHAIR: Saskia le Cessie (Leiden NL)

11:00-11:18	OC18-1	**Vivienn Weru** Heidelberg DE	Bayesian dynamic borrowing using the mixture prior: Impact of parameter choices
11:18-11:36	OC18-2	**Pierre-Emmanuel Poulet** Paris FR	Prediction powered inference for clinical trials
11:36-11:54	OC18-3	**Aris Perperoglou** London UK	Targeted maximum likelihood estimation for clinical trials with urvival outcome
11:54-12:12	OC18-4	**Sarah Ren** Bristol UK	Sensitivity analysis for unmeasured confounding in indirect treatment comparisons
12:12-12:30	OC18-5	**Marion Leila Claudine Kerioui** New York NY US; Cambridge UK	New clinical trial design borrowing information across patient subgroups based on fusion-penalized regression models

TUESDAY, 23 JULY 2024

PARALLEL HALL 3
(Library Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 19: Imperfect Data

CHAIR: Guadalupe Gómez-Melis (Barcelona ES)

11:00-11:18	OC19-1 Invited	**Rebecca Betensky** New York NY US	Pseudo-observations for censored covariates
11:18-11:36	OC19-2 Invited	**Jesus Vazquez** Chapel Hill NC US	Estimators under informative covariate censoring: An application to Huntington's disease
11:36-11:54	OC19-3 Invited	**Guadalupe Gómez-Melis** Barcelona ES	Regression and goodness-of-fit with interval-censored covariates
11:54-12:12	OC19-4	**Vasiliki (Valia) Baralou** Athens GR	Variable selection for survival data with multiply imputed covariates: A comparative simulation study
12:12-12:30	OC19-5	**Rheanna Mainzer** Parkville AU	Incomplete reporting of incomplete data: Findings and recommendations from a scoping review

PARALLEL HALL 4
(CR-1 Hall)

11:00-12:30 ORAL CONTRIBUTED SESSION 20: Clinical Trials III

CHAIR: Katherine Lee (Melbourne AU)

11:00-11:18	OC20-1	**Enyu Li** Coventry UK	Bootstrap confidence intervals in two-stage adaptive enrichment designs
11:18-11:36	OC20-2	**Babak Choodari-Oskooei** London UK	Designing efficient multi-arm multi-stage randomised clinical trials: With an application to surgical trials
11:36-11:54	OC20-3	**Ziyan Wang** Southampton UK	Modelling time-treatment interactions to increase power in multi-arm multi-stage, and platform trials
11:54-12:12	OC20-4	**Stef Baas** Cambridge UK; Twente NL	A general and computationally tractable approach for exact statistical analysis in response-adaptive designs
12:12-12:30	OC20-5	**Yisheng Li** Houston TX US	An extended Bayesian semi-mechanistic dose-finding design for phase I oncology trials using pharmacokinetic and pharmacodynamic information

12:30-13:30 LUNCH BREAK

13:30~ **FREE AFTERNOON | CONFERENCE EXCURSIONS**

WEDNESDAY, 24 JULY 2024

PLENARY HALL
(Emilios Riadis Hall)

09:00-10:30 INVITED SESSION 6: Bayesian Methods in Clinical Development

CHAIR: Marcia Rueckbeil (Amsterdam NL)

09:00-09:30	IS6-1	**Rebecca Turner** London UK	Applying Bayesian methods in clinical trials: Opportunities and challenges
09:30-10:00	IS6-2	**Nicky Best** London UK	Bayesian approaches in clinical development: Methods and case studies
10:00-10:30	IS6-3	**Andrew Thomson** Amsterdam NL	Regulatory considerations for the acceptability of Bayesian methods

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 21: Joint Models III

CHAIR: KyungMann Kim (Madison WI US)

09:00-09:18	OC21-1 StCA	**Achilleas Stamoulopoulos** Athens GR	Joint modelling of longitudinal data and informative visiting process to predict subject-specific probabilities of future gaps in care
09:18-09:36	OC21-2	**Anaïs Rouanet** Bordeaux FR	Joint modelling of the temporal relationships between multivariate longitudinal markers of Alzheimer's disease progression and clinical endpoints
09:36-09:54	OC21-3	**Elaheh Talebi-Ghane** Hamadan IR	Simultaneous modelling of longitudinal measurements and multi-state data with a cured fraction using a joint model
09:54-10:12	OC21-4	**Clemens Schaechter** Freiburg DE	Analysis of disease trajectories and treatment effects by combining variational autoencoders with random effects models
10:12-10:30	OC21-5	**Felix Boakye Oppong** Rotterdam NL; Brussels BE	Joint modelling of time-dependent biomarker variability and time-to-event outcomes

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 22: Clinical Trials IV

CHAIR: Stefan Michiels (Villejuif FR)

09:00-09:18	OC22-1	**Zhujie Gu** Cambridge UK	Joint modelling of survival and longitudinal data for sequential decision-making in clinical trials
09:18-09:36	OC22-2	**Kelly Van Lancker** Ghent BE	Automated, efficient and model-free inference for randomised clinical trials via data-driven covariate adjustment
09:36-09:54	OC22-3	**Harry Parr** London UK	An overview of prognostic scoring adjustments: Applied to neurodegenerative diseases
09:54-10:12	OC22-4	**Mollie Payne** London UK	Principal treatment effects: A novel approach to non-linear dose-response modelling in complex interventions
10:12-10:30	OC22-5	**Teresa Lee** London UK	Estimating the Complier Average Causal Effect (CACE) for randomised therapy trials – Is a binary compliance definition appropriate?

PARALLEL HALL 3
(Library Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 23: Competing Risks I

CHAIR: Nikos Pantazis (Athens GR)

09:00-09:18	OC23-1	**Judith Vilsmeier** Ulm DE	Non-Markov non-parametric estimation of complex multistate outcomes after hematopoietic stem cell transplantation
09:18-09:36	OC23-2	**Giorgos Bakoyannis** Indianapolis IN US	Semiparametric regression analysis of competing risks data with missing not at random cause of failure
09:36-09:54	OC23-3	**Kaya Miah** Heidelberg DE	Variable selection for penalised multi-state models incorporating molecular data
09:54-10:12	OC23-4	**Caterina Gregorio** Solna SE	Bridging multi-state models and network theory: An application to multimorbidity and aging research
10:12-10:30	OC23-5	**Juliette Ortholand** Paris FR	A joint spatiotemporal multivariate model with competing risks

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

09:00-10:30 ORAL CONTRIBUTED SESSION 24: Survival Analysis II

CHAIR: Paola Rebora (Milan IT)

09:00-09:18	OC24-1	**Jordache Ramjith** Nijmegen NL	Flexible hazard regression for generalised interval-censored time-to-event data and the gicsurv Rshiny app
09:18-09:36	OC24-2	**Jean Marie Boher** Marseille FR	Group sequential methods based on supremum logrank statistics under proportional and non-proportional hazards
09:36-09:54	OC24-3	**Mathieu Fauvernier** Lyon FR	Penalised flexible model for the mortality ratio of patients with a chronic disease compared to the general population
09:54-10:12	OC24-4	**Georgios Karakatsoulis** Thessaloniki GR	A general mixture cure model taking into account uncensored immune individuals
10:12-10:30	OC24-5	**Jan Beyersmann** Ulm DE	Hazards: Key quantities for analysis, interpretation and understanding of time-to-event data

10:30-11:00 COFFEE BREAK

PLENARY HALL
(Emilios Riadis Hall)

11:00-12:00 PRESIDENT'S INVITED SPEAKER: Biostatistics in the AI Era

PRESIDENT'S INVITED SPEAKER: Prof. Chris Holmes (Oxford UK)
CHAIR: Tomasz Burzykowski (Hasselt BE)

Poster Awards Presentation

PLENARY HALL
(Emilios Riadis Hall)

12:00-13:30 ISCB ANNUAL GENERAL MEETING

13:00-14:00 LUNCH BREAK

WEDNESDAY, 24 JULY 2024

PLENARY HALL
(Emilios Riadis Hall)

14:00-15:30 INVITED SESSION 7: Regulators' View of Randomised and Non-Randomised Evidence in Drug Development

CHAIRS: Abdel Babiker (London UK), **Giota Touloumi** (Athens GR)

14:00-14:15	IS7-1	**Sir Richard Peto** Oxford UK	The magic of randomisation versus the myth of Real-World Evidence
14:15-14:35	IS7-2	**John Concato** Silver Spring MD US	FDA and Real-World Evidence
14:35-14:55	IS7-3	**Andrew Thomson** Amsterdam NL	Use of Real-World Evidence in EU regulatory decision making
14:55-15:10	IS7-4	**Adrian Mander** Cardiff UK	The use of Real-World Evidence within the context of Open-Label Extensions
15:10-15:20	Discussion	**James D. Neaton** Minneapolis MN US	
15:20-15:30		**Round Table Discussion**	

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

14:00-15:30 ORAL CONTRIBUTED SESSION 25: Longitudinal Data Analysis II

CHAIR: Roula Tsonaka (Leiden NL)

14:00-14:18	OC25-1	**Chengfeng Zhang** Guangzhou CN	Dynamic prediction of individual survival time based on RMST using longitudinal covariates
14:18-14:36	OC25-2	**Corentin Ségalas** Bordeaux FR	Dynamic random survival forests using functional principal component analysis for the prediction of survival outcomes from time-varying predictors
14:36-15:04	OC25-3	**Grace Tompkins** Waterloo CA	Weight trimming for flexible inverse probability of treatment and intensity weighting
15:04-15:12	OC25-4	**Sara Baart** Rotterdam NL	Analysing longitudinal semi-continuous outcomes with excess zeros
15:12-15:30	OC25-5	**Arthur Hughes** Bordeaux FR	A variance-component score test for the comparison of gene-set transcriptomic profiles of vaccines

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

14:00-15:30 ORAL CONTRIBUTED SESSION 26: Prediction and Prognostic Models II

CHAIR: Toshiro Tango (Tokyo JP)

14:00-14:18	OC26-1	**Alexandra Hunt** Liverpool UK	Externally validating clinical prediction models for recurrent events
14:18-14:36	OC26-2	**Jose Benitez-Aurioles** Manchester UK	Understanding algorithmic fairness for clinical prediction in terms of net benefit and health equity
14:36-15:04	OC26-3	**Nathalie Weidinger** Biberach an **der Riss DE**	Conceptualisation and visualisation of multivariable models for the analysis of biomarkers for patient selection in clinical trials
15:04-15:12	OC26-4	**Pamela Solano** Regensburg DE	Comparing modern machine learning methods for predicted individual treatment effect
15:12-15:30	OC26-5	**Max Behrens** Freiburg DE	Improving prediction models by incorporating external data with weights based on similarity

PARALLEL HALL 3
(Library Hall)

14:00-15:30 ORAL CONTRIBUTED SESSION 27: Competing Risks II

CHAIR: Nadine Binder (Freiburg DE)

14:00-14:18	OC27-1	**Serigne Lo** Sydney AU	Mixture cure cause-specific hazard Cox models for competing risks data with partly interval censoring
14:18-14:36	OC27-2	**Jagathnath Krishna Kumarapillai Mohanan Nair** Thiruvananthapuram IN	Joint frailty modelling of multiple time-to-events and longitudinal measures
14:36-15:04	OC27-3	**Ilse Cuevas Andrade** London UK	Non-parametric frailty model for the natural history of prostate cancer; using data from a screening trial
15:04-15:12	OC27-4	**Liesbeth de Wreede** Leiden NL	Dynamics of excess mortality during the Covid-19 pandemic in the Netherlands – Impact of vaccination and infection analysed through a multi-state model incorporating relative survival
15:12-15:30	OC27-5	**Haoning Shen** Guangzhou CN	Assessing treatment effects with adjusted restricted mean time lost in observational competing risks data

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 4
(CR-1 Hall)

14:00-15:30 ORAL CONTRIBUTED SESSION 28: Electronic Health Data

CHAIR: Audrone Jakaitiene (Vilnius LT)

14:00-14:18	OC28-1	**Andrea Corbetta** Milan IT; Helsinki FI	Predictive modelling of longitudinal statin adherence trajectories with functional data analysis and machine learning in Finnish nationwide cohort
14:18-14:36	OC28-2	**Clemence Leyrat** London UK	Emulation of target cluster trials of complex interventions: Estimands, methods and application
14:36-15:04	OC28-3	**Hester Allen** Oxford UK	Linking real world data to assess post-acute mortality following COVID-19 infection compared to matched test-negative people in England: Findings from the ECHOES national dataset
15:04-15:12	OC28-4	**Faye Baldwin** Liverpool UK	Emulation of an SLCO1B1*5-guided target trial to guide statin therapy
15:12-15:30	OC28-5	**Jessy Hansen** Melbourne AU	Evaluating the performance of common benchmarking and outlier classification methods in clinical registries – A simulation study

15:30-16:00 COFFEE BREAK

PLENARY HALL
(Emilios Riadis Hall)

16:00-17:30 INVITED SESSION 8: Combining Real-World Data and Randomised Clinical Trials

CHAIR: Nikolaos Demiris (Athens GR)

16:00-16:30	IS8-1	**Brieuc Lehmann** London UK	Data fusion for heterogeneous treatment effect estimation with multi-task Gaussian processes
16:30-17:00	IS8-2	**Nicky J. Welton** Bristol UK	Population average estimates for health technology assessment
17:00-17:30	IS8-3	**Pablo Verde** Düsseldorf DE	Bayesian nonparametric meta-analysis, bias correction methods, and real-world evidence synthesis

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

16:00-17:30 ORAL CONTRIBUTED SESSION 29: Prediction and Prognostic Models III

CHAIR: Rodolphe Thiébaut (Bordeaux FR)

16:00-16:18	OC29-1	**Doranne Thomassen** Leiden NL	Effective sample size of individual predictions: Quantifying uncertainty in machine learning models
16:18-16:36	OC29-2	**Ben Van Calster** Leuven BE; Leiden NL	Properness of common performance measures for risk prediction models: A simple illustration
16:36-16:54	OC29-3	**Francois Petit** Paris FR	Optimal treatment regimes for the net benefit of a treatment
16:54-17:12	OC29-4	**Jessica Xu** Melbourne AU	Optimising dynamic treatment regimens using sequential multiple assignment randomised trials data with missing data
17:12-17:30	OC29-5	**Rebecca Whittle** Birmingham UK	Extended sample size calculations for external validation of a machine learning model for individual risk prediction of a binary outcome

PARALLEL HALL 2
(Maurice Saltiel III Hall)

16:00-17:30 ORAL CONTRIBUTED SESSION 30: Survival Analysis III

CHAIR: Evangelos Kritsotakis (Heraklion Crete GR)

16:00-16:18	OC30-1	**Giuliana Cortese** Padua IT	Regression models for recurrent events in presence of terminal events with efficient estimation
16:18-16:36	OC30-2	**Andrea Mario Vergani** Milan IT	Prediction of incident cardiovascular events using omics-derived latent factors
16:36-16:54	OC30-3	**Lubomir Stepanek** Prague CZ	A novel approach to multiple survival curves comparison based on machine learning: Utilising random forests with reducing assumption dependency
16:54-17:12	OC30-4	**Elsa Coz** Lyon FR	Flexible rate models for recurrent event data
17:12-17:30	OC30-5 StCA	**Léa Orsini** Paris FR; Leuven BE	Bayesian analysis of restricted mean survival time adjusted on covariates using pseudo-observations

WEDNESDAY, 24 JULY 2024

PARALLEL HALL 3
(Library Hall)

16:00-17:30 ORAL CONTRIBUTED SESSION 31: Clinical Trials V

CHAIR: Fotios Siannis (Athens GR)

16:00-16:18	OC31-1	**Stephen Schüürhuis** Berlin DE	Group-sequential methods for generalised pairwise comparisons
16:18-16:36	OC31-2	**Antonios Daletzakis** Amsterdam/Nijmegen NL	Response rate estimation in single-stage basket trials: A comparison of estimators that allow for borrowing across cohorts
16:36-16:54	OC31-3	**Emily Alger** London UK	Patient-centred dose-finding trials using safety, efficacy and patient-reported outcomes
16:54-17:12	OC31-4	**Zhulin Yin** London UK	Using multi-state models to design single-arm adaptive trials with progression-free and overall survival endpoints
17:12-17:30	OC31-5	**Tansy Edwards** London UK	A pragmatic approach to confidence interval estimation for a complier average causal effect for cluster randomised trials

PARALLEL HALL 4
(CR-1 Hall)

16:00-17:30 ORAL CONTRIBUTED SESSION 32: Statistics in Epidemiology II

CHAIR: Maria Grazia Valsecchi (Milan IT)

16:00-16:18	OC32-1	**Maria Vittoria Chiaruttini** Padua IT	Win statistics in observational cancer research: Integrating clinical and quality of life outcomes
16:18-16:36	OC32-2	**Edouard Chatignoux** Saint Maurice FR	Joint disease mapping for bivariate count data with residual correlation due to common cases
16:36-16:54	OC32-3 CASc	**Mahnaz Ibrahim** Dhaka BD	On estimation of relative risk regression model for binary data
16:54-17:12	OC32-4	**Marco Ratta** Turin IT	Accounting for the recruitment process into Bayesian modelling of vaccine data
17:12-17:30	OC32-5	**Georgios Seitidis** Ioannina GR	Stochastic search variable selection in network meta-analysis: A novel approach for detecting inconsistencies

20:00~ **CONFERENCE DINNER AT POLIS CONVENTION CENTRE**

THURSDAY, 25 JULY 2024

PARALLEL HALL 1
(Maurice Saltiel I & II Hall)

09:00-12:30

MINI-SYMPOSIUM 1: Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development

ORGANISERS: Marcia Rueckbeil (Amsterdam NL), **Els Goetghebeur** (Ghent BE), **Mouna Akacha** (Basel CH)

In collaboration with the ISCB Statistics in Regulatory Affairs (SiRA) Subcommitte

SUMMARY: This session will explore the changing landscape of study design options and modeling techniques in the field of drug development. While randomized controlled trials (RCTs) are considered the gold standard for establishing causal relationships and estimating treatment effects, alternative approaches have gained traction in recent years. This shift is also evident in the guidance provided by regulatory agencies such as the FDA and EMA, who have issued recommendations for externally controlled trials, single-arm trials and model-informed drug development. With invited speakers representing academia, industry, and regulatory bodies, this session aims to foster a dynamic discussion on the opportunities and challenges encountered in adopting novel study designs and modelling approaches in drug development.

PART I			
09:00-09:30	**MS1-1**	**Saskia le Cessie** Leiden NL	Beyond conventional randomised controlled trials: Patient reported outcomes in single arm studies
09:30-10:00	**MS1-2**	**Marc Vandemeulebroecke** Basel CH	Applications of designs with external controls / hybrid designs in drug development
10:00-10:30	**MS1-3**	**Sarah Zohar** Paris FR	Voilà: The European collaborative project INVENTS
10:30-11:00		BREAK	
PART II			
11:00-11:30	**MS1-4**	**Kit Roes** Nijmegen NL	Regulatory aspects
11:30-11:45	**MS1-5**	**Martin Posch** Vienna AT	Advanced trial designs and analysis methods to improve confirmatory trials in rare diseases
11:45-12:30	**Panel discussion**	**All Speakers & Organisers** | Moderated by **Tim Friede** (Goettingen DE), **Marcia Rückbeil** (EMA)	

THURSDAY, 25 JULY 2024

PARALLEL HALL 2
(Maurice Saltiel III Hall)

09:00-12:30

MINI-SYMPOSIUM 2: STRengthening Analytical Thinking for Observational Studies (STRATOS Initiative) - Recent Progress and Foci for the Future

ORGANISERS: Willi Sauerbrei (Freiburg DE), **Els Goetghebeur** (Ghent BE)

In collaboration with the STRATOS Steering Committee

SUMMARY: The STRATOS (STRengthening Analytical Thinking for Observational Studies) initiative is a large collaboration of experts in many different areas of biostatistical research. It was launched at a halfday Mini-Symposium at ISCB 2013. The objective of STRATOS is to provide accessible and accurate guidance in the design and analysis of observational studies (www.stratos-initiative.org). We will present recent progress from some topic groups and discuss foci for the near future. Soon after the ISCB we will have our 3rd general meeting at the Lorentz Center in Leiden, Netherlands. Some of the talks look forward to cross topic groups research and include material from the preparation phase for this meeting.

PART I			
09:00-09:23	**MS2-1 for STRATOS**	**Willi Sauerbrei** Freiburg DE	On six foci for the future of STRATOS
09:23-09:46	**MS2-2 for TG3**	**Carsten Oliver Schmidt** Greifswald DE	Building blocks of efficient initial data analysis and data quality assessments – Best practice examples
09:46-10:09	**MS2-3 for TG8**	**Malka Gorfine** Tel Aviv IL	An overview and recent developments in the analysis of multistate processes
10:09-10:30	**MS2-4 for TG6**	**Ben van Calster** Leuven BE	Assessing performance when developing or validating clinical risk prediction models in the era of machine learning
10:30-11:00		BREAK	
PART II			
11:00-11:30	**MS2-5 for TG2/TG4**	**Aris Perperoglou** London UK	Adjusting for covariate measurement error on functional form estimation: Design and early results from a blinded, collaborative STRATOS project
11:30-11:53	**MS2-6 for TG7**	**Els Goetghebeur** Ghent BE	Causal inference moving forward – Embracing joint (dis)appearances
10:09-10:30	**MS2-7 for TG1**	**Katherine Lee** Melbourne AU	Current and future initiatives in missing data
12:16-12:30	**Discussion**		

THURSDAY, 25 JULY 2024

PARALLEL HALL 3
(Library Hall)

09:00-12:30 EARLY CAREER BIOSTATISTICIANS' DAY

Organised by the ISCB Early Career Biostatisticians (ECB) Subcommittee

SUMMARY: We welcome all students and researchers starting their journey in Biostatistics to join the Early Career Biostatisticians' (ECB) Day that will take place on Thursday, 25 July. The aim of the day is to share personal experiences relating to biostatistical research and discuss how to deal with the potential pitfalls of the research process to become better researchers in the process. Whether you are just about to graduate or have already experience working as a researcher or biostatistical consultant, you will benefit from meeting your peers, exchanging your thoughts and ideas, and getting to know more about how to shape a career in biostatistics. Thus, this day will complement the main conference which mainly focuses on research results.

PART I			
09:00-09:05	**WELCOME**		
09:05-09:45	**ECB-1**	**Harald Binder** Freiburg DE	AI, small data, and me
09:45-10:00	**ECB-2**	**Abraham Contreras** Coventry UK	Design and analysis of co-enrolment trials
10:00-10:15	**ECB-3**	**Ajsi Kanapari** Padua IT	How can we leverage the power of Machine Learning and Artificial Intelligence when we design clinical trials? Insights and lesson learned
10:15-10:30	**ECB-4**	**Alice-Maria Toader** Liverpool UK	Working with healthcare systems data in clinical trials
10:30-11:00	BREAK		
PART II			
11:00-11:30	**ECB-5**	**Luca Carlini & Mehdi Gholam** Utrecht NL	Analysis of aggregated data via generation of pseudo–individual participant data at Danone Nutricia Research
11:30-11:45	**ECB-6**	**Mae Chester Jones** Oxford UK	Developing clinical prediction models in maternal health; experiences from a systematic review
11:45-12:00	**ECB-7**	**Mynul Islam** Dhaka BD	Child Malnutrition at Upazila Level in Bangladesh: A Small-Area Estimation Approach
12:00-12:30	**ROUND TABLE DISCUSSION & CLOSING**		

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSIONS GROUP A

POSTER SESSION A-01: CLINICAL TRIALS DESIGN AND SIMULATIONS

P-A01-01	**Danai Andreadi** Lausanne CH	Adaptive design implementation as a way to cope with recruitment difficulties
P-A01-02	**Shenghua Yuan** Framingham MA US	A practical seamless Phase II/III design for special populations and rare disease settings
P-A01-03	**Gianmarco Caruso** Cambridge UK	A response-adaptive multi-arm design for normal endpoints using entropy-based allocation rule
P-A01-04	**Michaela Maria Freitag** Berlin DE	Exploring efficient drug discovery for Major Depressive Disorder: A Phase II platform trial design
P-A01-05	**Vanessa Ihl** Aachen DE	Quantification of allocation bias in clinical trials under a response-adaptive randomisation procedure for normal response variables
P-A01-06	**Ryota Ishii** Tsukuba Ibaraki JP	Bias adjustment in an adaptive seamless design with different binary outcomes
P-A01-07	**Ayon Mukherjee** Frankfurt am Main DE	Covariate-adjusted response adaptive designs for semi-parametric survival models
P-A01-08	**Rainer Puhr** Melbourne AU	A fast, flexible simulation framework for Bayesian adaptive designs with time-to-event endpoints
P-A01-09	**Hong Sun** Boudry CH	Speeding up the clinical studies with biomarker-based randomisation or adaptive Bayesian design with biomarker enrichment
P-A01-10	**Silvia Calderazzo** Heidelberg DE	Adaptive quantification of prior impact in terms of effective sample size
P-A01-11	WITHDRAWN	
P-A01-12	**Ajsi Kanapari** Padua IT	Mind the Gap: A scoping review on Machine Learning and Artificial Intelligence in the design of Clinical Trials
P-A01-13	**Ajsi Kanapari** Padua IT	Stratified propensity score randomisation: The effect of integrating machine learning in the clinical trial design
P-A01-14	**J. Jack Lee** Houston TX US	Augmenting control arm of randomised-controlled trials by incorporating information across multiple external sources with stratified propensity score and data-driven mixture prior
P-A01-15	**Maria Vittoria Chiaruttini** Padua IT	Efficient group sequential design: Harnessing dynamic historical borrowing in medical device trial
P-A01-16	**Dimitris Karlis** Athens GR	Single-arm hierarchical testing with historical controls
P-A01-17	**Cristina Marelli** Genoa IT	A 1:1:1 matching design to account for centre-specific prophylaxis/treatment administration in observational studies

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

P-A01-18	**Guido Moreira** Regensburg DE	Random weights for historical control enrichment
P-A01-19	**Giulia Risca** Milan IT	Basket trials in very rare diseases: Are they feasible?
P-A01-20	**Anais Andrillon** Tournus FR	U-DESPA: A utility-based Bayesian approach for dosage optimisation handling PK, PD safety and efficacy in oncology clinical trials
P-A01-21	**Weishi Chen** Cambridge UK	An efficient approach to operational prior specification in Phase I dose-combination escalation trials
P-A01-22	**Linxi Han** Bristol UK	Bayesian hierarchical model for dose-finding trial incorporating historical data
P-A01-23	**Thomas Jaki** Regensburg DE	Joint TITE-CRM: A design for dose finding studies for therapies with late-onset safety and activity outcomes
P-A01-24	**Daniel Bodden** Cologne DE	Allocation bias in group sequential designs
P-A01-25	**Shun Fu Lee** Hamilton CA	Internal pilot sample size re-estimation for the B-free trial - A cluster randomised crossover trial
P-A01-26	WITHDRAWN	
P-A01-27	**Stefanie Schoenen** Aachen DE	Quantifying the impact of allocation bias in randomised clinical trials with multi-component endpoints
P-A01-28	**Takashi Sozu** Tokyo JP	Sample size determination considering the functional relationship between co-primary endpoints in clinical trials
P-A01-29	**Satomi Okamura** Osaka JP	Non-inferiority trials with evidence of assay sensitivity considering effect modification
P-A01-30	**Ulrike Poetschger** Vienna AT	Pseudo-value regression for the design of non-inferiority studies in Paediatric Oncology
P-A01-31	**Valeria Mazzanti** Dallas TX US	Optimising oncology clinical trial strategies: Comparative analysis of single versus dual endpoint designs
P-A01-32	**Chieh Chiang** New Taipei City TW	The use of two-sided tolerance interval testing with considering the variability of batches in the assessment of biosimilarity
P-A01-33	**Ruochen Du** Singapore SG	Estimating treatment effect in randomised controlled trials with continuous outcomes subject to non-compliance via a CACE framework: A logistic regression based multiple imputation approach
P-A01-34	**Mizuna Itagaki** Tokyo JP	Extension the outcome of two stage design to survival time
P-A01-35	WITHDRAWN	

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

P-A01-36	**Dominic Stringer** London UK	Should we use correlation between clinical trial outcomes as we do for a single trial outcome with repeated measures?
P-A01-37	**Foteini Tsotra** Athens GR	Selection of time to measure the RMST: A simulation study
P-A01-38	**Haotian Wang** Coventry UK	A review of methods for optimal utility-based design of oncology clinical development programmes
P-A01-39	**Xinyue (Sylvia) Zhang** Newcastle upon Tyne UK	Enhancing endpoint analysis in inflammatory bowel disease trials: A comparative study of different modelling strategies
P-A01-40	**Boaz Adler** Dallas TX US	Integrating market access considerations in clinical trial design: Leveraging cloud-native software for holistic optimisation
P-A01-41	**Emily Alger** London UK	U-PRO-CRM: Designing patient-centred dose-finding trials with patient-reported outcomes
P-A01-42	**Ying Lu** Stanford CA US	Cultivating patient preferences in ALS clinical trials: Reliability and Prognostic Value of the Patient-Ranked Order of Function (PROOF)
P-A01-43	**Kazue Yamaoka** Tokyo JP	Intervention to improve supporting skills of registered dietitians in lifestyle modification: A feasibility study
	POSTER SESSION A-02: USING STATISTICS TO IMPROVE THE CONDUCT OF CLINICAL TRIALS	
P-A02-01	**Abraham Richi Contreras** Coventry UK	Design and analysis of co-enrolment trials
P-A02-02	**Christina Habermehl** Darmstadt DE	Applying the estimands framework in oncology dose escalation - A practical example of improving the MTD estimate in the first-in-human Phase 1 trial of antibody-drug-conjugate M9140
	POSTER SESSION A-03: INFECTIOUS DISEASES MODELLING	
P-A03-01	**Bethany Heath** Cambridge UK	Evaluating pooled testing policies in community and healthcare settings for emerging epidemics
P-A03-02	**Kazumi Omata** Tokyo JP	Estimation of the effect of Rapid ART on HIV infection
P-A03-03	**Jose Maria Quintana-Lopez** Usansolo ES	Evaluating forecasting models for COVID-19: Insights from the Omicron period
P-A03-04	**Zdeněk Valenta** Prague CZ	An alternative measure of vaccine effect based on Aalen's additive survival model framework
P-A03-05	**Jan Brink Valentin** Gistrup DK	Redefining the SEIR model for direct modulation of hospital load independent of number of susceptible, exposed, infected, and recovered

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSION A-04: STATISTICAL METHODS FOR HEALTH ECONOMIC EVALUATIONS		
P-A04-01	**Samer Kharroubi** Beirut LB	Use of a nonparametric Bayesian method to model health state preferences: An application to Lebanese SF-6D valuations
P-A04-02	**Samer Kharroubi** Beirut LB	Bayesian modelling of health state preferences: Can existing preference data be used to generate better estimates?
P-A04-03	**Kazutaka Nishio** Osaka JP	Validation of EQ-5D in patients with heart failure with preserved ejection fraction
P-A04-04	**Petros Pechlivanoglou** Toronto CA	Impact of limited sample size and follow-up on partitioned survival or multistate modelling-based health economic models: A simulation study
P-A04-05	**Rahma Said** Leicester UK	Comprehensive evidence synthesis: Embracing single-arm and non-randomised studies in network meta-analysis
P-A04-06	**Junfeng Wang** Utrecht NL	A review of evidence level and causal support in prediction models of complications in health economic models of type 2 diabetes
POSTER SESSION A-05: STATISTICS IN EPIDEMIOLOGY		
P-A05-01	**Vasiliki Bountziouka** Lemnos GR	Assessing the impact of dietary fatty acids on leucocyte telomere length: A substitution model approach using the UK Biobank data
P-A05-02	**Apostolos Gkatzionis** Bristol UK	Statistical methods for genetic studies of disease progression
P-A05-03	**Christiana Kartsonaki** Oxford UK	Identifying metabolic biomarkers associated with risk of pancreatic cancer: A case-subcohort study
P-A05-04	**Zewen Lu** Manchester UK	Does time-adjusted admission severity improve stroke prognostication?
P-A05-05	**Ondřej Májek** Prague CZ	Different age-specific time trends of colorectal cancer incidence rates in the Czech Republic and a potential role of colorectal cancer screening: Application of an age-period-cohort model
P-A05-06	**Aleksander Owczarek** Katowice PL	Risk factors for anticitrullinated protein seropositivity in Older Poles
P-A05-08	**Elsa Coz** Lyon FR	Environmental factors and their impact on Emergency Department visits in Lyon, France
P-A05-09	**Jinheum Kim** Hwaseong KR	Association between the oxidative balance score and the quality of life based on data from the Korean National Health and Nutrition Examination Survey
P-A05-10	**Paolo Dalena** Trieste IT	A lesson for understanding parents' perspective: Perception of quality of care and COVID-19 related fears among users of paediatric health services over the COVID-19 pandemic in 11 facilities in Italy
P-A05-11	**Natchalee Srimaneekarn** Bangkok TH	The current use of reliability analysis for questionnaires in dental research
P-A05-12	**Thomas Stojanov** Basel CH	Challenges in the establishment of minimal important changes (MIC) for outcomes by patients undergoing an arthroscopic rotator cuff repair in Switzerland

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

P-A05-13 LB	**Willi Sauerbrei** Freiburg DE	STRengthening Analytical Thinking for Observational Studies (STRATOS) - Accessible and accurate guidance in the design and analysis of observational studies
P-A05-14	**Jan Brink Valentin** Gistrup DK	Why has rank-based non-parametric hypothesis testing become the gold standard for comparison of non-normal distributed data?
	POSTER SESSION A-06: CAUSAL INFERENCE	
P-A06-01	**Rachid Abbas** Basel CH	Target trial emulation to leverage randomised trial data: Investigate alternative questions of interest in late-stage development of monoclonal antibodies in Alzheimer's disease
P-A06-02	**Nicole Fontana** Milan IT	Drug repurposing of metformin via emulating a target trial for preventing chronic kidney disease progression
P-A06-03	**Josef Fritz** Innsbruck	Target trial emulation to overcome immortal time bias in real-world data of urothelial cancer
P-A06-04	**Derek Hazard** Freiburg DE	Target trial emulation with competing risk analysis using hospital observational data from COVID-19 patients
P-A06-05	**Jingyi Xuan** London UK	Using inverse probability of censoring weighting to estimate hypothetical estimands in clinical trials: Should we implement stabilisation, and if so how?
P-A06-06	**Christiana Drake** Davis CA US	Causal inference for integrating external data in randomised trials
P-A06-07	**Susanne Dandl** Munich DE	Heterogeneous treatment effect estimation for observational data using model-based forests
P-A06-08	**Ilaria Prosepe** Leiden NL	Causal updating of prediction models in a pandemic environment
P-A06-09	**Hiroyuki Shiiba** Tachikawa JP	Marginal structural models for estimating causal risk ratio and causal risk difference in longitudinal studies
P-A06-10	**Charlotte Voinot** Le Chesnay Rocquencourt FR	Estimation of the Average Treatment Effect (ATE) in causal survival: Comparison, applications and practical recommendations
P-A06-11	**Junxian Zhu** Singapore SG	Estimating causal effect on count outcome with excess zeros in randomised clinical trials subject to some degree of noncompliance
P-A06-12	**Alessandro Gasparini** Stockholm SE	Mediation analysis with imperfectly defined mediators: A microsimulation experiment with breast cancer survival, socio-economic status, and stage at diagnosis
P-A06-13	**Katherine Holdsworth** London UK	Causal mediation analysis for a survival outcome with longitudinal mediators, time-varying confounders and left truncation
P-A06-14	**Sharon Lutz** Boston MA US	Estimating correlations when inferring the causal direction between two traits in genetic association studies
P-A06-15	**Elisavet Syriopoulou** Stockholm SE	Mediation analysis with multiple mediators using the relative survival framework: An example exploring socioeconomic differences in cancer survival
P-A06-16 LB	**Julia Krzykalla** Biberach an der Riss DE	Biomarkers in clinical drug development: A mediation analysis approach

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

P-A06-17	**Nilufar Akbari** Berlin DE	Comparison of modelling approaches for continuous covariates in explanatory modelling: Does it matter which methods are used?
P-A06-18	**Fabiola Del Greco Miglianico** Bolzano/Bozen IT	Accounting for the between-study heterogeneity when selecting instrumental variables for two-sample Mendelian randomisation studies
P-A06-19	**Tianyuan Gu** Singapore SG	Implications of non-compliance in the experimental arm on treatment effect metrics in randomised controlled trials: A methodological assessment
P-A06-20	**Laura Güdemann** Exeter UK	Assessing patient subgroups specific safety and effectiveness of oral type 2 diabetes treatments using the Local Instrumental Variable method
P-A06-21	**Emanuele Koumantakis** Turin IT	Dealing with highly imbalanced groups from surveys: Life satisfaction among Italian native, migrant and international adopted adolescents
P-A06-22	**Etienne Peyrot** Paris FR	Comparing covariates balancing methods, a simulation study
POSTER SESSION A-07: JOINT AND LATENT VARIABLES MODELS		
P-A07-01	**Behnaz Alafchi** Hamadan IR	Joint modelling of ordinal repeated measurements and multi-state survival data: A simulation study
P-A07-02	**Behnaz Alafchi** Hamadan IR	Simultaneous modelling of ordinal repeated measurements and multi-state data in the presence of cure fraction: A joint model
P-A07-03	**Denis Rustand** Thuwal SA	Swift inference and dynamic risk predictions for multivariate joint models using INLAjoint
P-A07-04	**Jun Ma** Sydney AU	Joint modelling of longitudinal covariates and partly-interval censored survival data - A penalised likelihood approach
P-A07-05	**Farzana Osman** Durban ZA	Investigating the role of longitudinal cytokine concentrations in HIV acquisition among women: A joint modelling approach
P-A07-06	**Rehema Ouko** London UK	Joint modelling of non-time to event longitudinal data: A systematic review of methodology and applications
P-A07-07	**Edouard Bonneville** Leiden NL	Quantifying associations between immune cell kinetics and (competing) allo-immunological events after allogeneic stem cell transplantation using joint models: Opportunities and challenges
P-A07-08	**Maren Hackenberg** Freiburg DE	Investigating a domain adaptation approach for integrating different measurement instruments in a longitudinal clinical registry
P-A07-09	**Eleni Kalamara** Thessaloniki GR	Using latent profile analysis to understand burnout in a sample of Greek teachers
P-A07-10	**Evripidis Kapanidis** Stockholm SE	Utilising screening cohorts to model DCIS and invasive breast cancer
P-A07-11	**Elaheh Talebi-Ghane** Hamadan IR	A correlated joint frailty model for recurrent and terminal events: Application to breast cancer patients
P-A07-12 LB	**Abderrahim Oussama Batouche** Helsinki FI	A joint frailty model to assess the relationship between time to curative treatment and biochemical recurrence in Prostate Cancer patients

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSION A-08: METHODS FOR HIGH DIMENSIONAL DATA (INCLUDING -OMICS)		
P-A08-01	**Beryl Ang'iro** Karatina KE	Multivariate survival analysis for high dimensional data: A sample splitting approach
P-A08-02	**Niklas Brunn** Freiburg DE	Infusing structural assumptions into dimension reduction for single-cell RNA sequencing data to identify small gene sets
P-A08-03	**Svetlana Cherlin** Newcastle Upon Tyne UK	Utilising high-dimensional data in randomised clinical trials
P-A08-04	**Paolo Dalena** Trieste IT	Prioritisation of differential methylation regions for the prediction of Coeliac disease: A Machine Learning approach
P-A08-05	**Arianna Galotta** Milan IT	Statistical methodologies for Olink data analysis: Different approaches for reliable evaluation of Omics Data
P-A08-06	**Franziska Kappenberg** Dortmund DE	How to benefit from high-dimensional gene expression data for dose-response modelling
P-A08-07	**Elisa Merelli** Milan IT	Spatial proteomics analysis for discriminating autoimmune liver diseases: Insights from MALDI-TOF mass spectrometry imaging
P-A08-08 VIRTUAL	**Shivani Saxena** Ahmedabad IN	Clustering High-dimensional spaces using a modified EM algorithm with fractional order assignment
P-A08-09 LB	**Eva Brombacher** Freiburg DE	Characterising the omics landscape based on 10,000+ datasets
POSTER SESSION A-09: META-ANALYSIS		
P-A09-01	**Thunyarat Anothaisintawee** Bangkok TH	Comparative performance of different depression screening tools in primary care setting: A systematic review and network meta-analysis
P-A09-02	**Maria Konstantina Bouka** Knoxville TN US	Sedentary lifestyle in childhood, adolescence, and young adulthood and its impact on health outcomes: An umbrella review
P-A09-03	**Efthymia Derezea** Bristol UK	Network meta-analysis of diagnostic test accuracy reported at multiple thresholds
P-A09-04	**Konstantina Thaleia Pilali** Padua IT	A Bayesian network meta-analysis with mediation for length of stay and early mortality in ICU setting
P-A09-05	**Sasivimol Rattanasiri** Bangkok TH	Efficacy of targeted therapy or immunotherapy as adjuvant treatment for non-small cell lung cancer: A systematic review and network meta-analysis
P-A09-06	**Ellesha Smith** Leicester UK	The effectiveness of delivery modalities of diabetes prevention programs: A systematic review and component network meta-analysis
P-A09-07 LB	**Maria Petropoulou** Freiburg DE	Network meta-analysis with dose-response relationships
P-A09-08	**Nana-Adjoa Kwarteng** Freiburg DE	Implications of diagnostic meta-analysis studies for the design of future diagnostic studies

MONDAY, 22 JULY 2024

POSTER AREA
(Museum Hall)

P-A09-09	**Pawin Numthavaj** Bangkok TH	Diagnostic performance of WatchPAT peripheral arterial tonometry in detecting obstructive sleep apnea: Systematic review and meta-analysis
P-A09-10	**Sarah Ren** Sheffield UK	Diagnostic accuracy of D-dimer for acute aortic syndromes: Systematic review and meta-analysis
P-A09-11	**Luca Carlini** Utrecht NL	Analysis of aggregated data via generation of pseudo–Individual Participant Data at Danone Nutricia Research
P-A09-12	**Stavros Kanakis** Thessaloniki GR	Single and dual hormone artificial pancreas systems in patients with type 1 diabetes: An overview of systematic reviews
P-A09-13	**Benoit Liquet** Sydney AU	Meta-Analysis models with group structure for pleiotropy detection at gene and variant level by using summary statistics from multiple datasets
P-A09-14	**Kanella Panagiotopoulou** Paris FR	Meta-analysis models relaxing the random effects normality assumption: A simulation study
P-A09-15	**Arnaud Serret-Larmande** Paris FR	Population-adjusted indirect comparisons: A simulation study
P-A09-16	**Pantelis Bagos** Lamia GR	A versatile software package for combining dependent or independent p-values
	POSTER SESSION A-10: MISCELLANEOUS	
P-A10-01	**Lucia Ameis** Cologne DE	Rest-and-jump-and then? – Identifying changes in gene expression
P-A10-02	**Alice Bonomi** Milan IT	Stepwise, forward, backward selection and cross-validation: How to choose the best predictors?
P-A10-03	**Hugo Hadjur** Grenoble FR	Simple approaches for portfolio quantitative decision-making
P-A10-04	**Leroy Jide Ovbude** Wavre BE	Quantitative decision making in the vaccines world: A new paradigm for risk assessment
P-A10-05	**Thomas Forstner** Linz AT	Multiple testing in statistical inference: A historical overview
P-A10-06	**Niklas Hagemann** Cologne DE	Overcoming model uncertainty - How equivalence tests can benefit from model averaging
P-A10-07	**Georgios Koliopanos** Klosters CH	A simple-to-use R package for mimicking study data by simulations
P-A10-08	**Yongxi Long** Utrecht NL	Three questions about ordinal outcomes in neurological trials
P-A10-09	**Yongxi Long** Utrecht NL	An extension of the win odds for ordinal repeated measurements
P-A10-10	**Rachael Stannard** Leicester UK	Incorporation of patient and public involvement in statistical methodology research: Developing resources for both researchers and public contributors
P-A10-11	**Anqi Sui** London UK	Robust estimation of overdispersion parameter for outlier detection in clinical audits

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSIONS GROUP B

POSTER SESSION B-01: ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING METHODS

P-B01-01	**Bushra Alrashdi** Liverpool UK	Epilepsy survival analysis: The integration of machine learning with genetic data from the SANADII Dataset
P-B01-02	**Vasiliki Bountziouka** Lemnos GR	Comparative analysis of ChatGPT and human performance in post classification and sentiment analysis
P-B01-03	**Emanuele Koumantakis** Turin IT	Embedding models for multimorbidity detection: A deep learning framework using healthcare administrative databases
P-B01-04	**Oluwole Nuga** Ota NG	A machine learning model for malaria prediction in an endemic area of Nigeria
P-B01-05	**Sotirios Roussos** Athens GR	Machine learning for HIV prediction in people who inject drugs: Addressing challenges of imbalanced data
P-B01-06	**Asuka Suzuki** Tokyo JP	A novel simple method for evaluating the contribution of explanatory variables in deep learning: An application to pathogenicity prediction results for breast cancer-related variants of unknown significance
P-B01-07 CASc	**Nabil Ahmed Uthso** Dhaka BD	Machine learning methods in dynamic survival prediction under standard and mixed-model landmarking framework
P-B01-08	**Mi Hong Yim** Daejeon KR	Comparison of machine learning models for classification of diabetes mellitus using body composition indicators in men
P-B01-09	**Elina Hypponen** Adelaide AU	LLPowershap: Logistic loss-based automated Shapley values feature selection method

POSTER SESSION B-02: ELECTRONIC HEALTH RECORDS AND REAL-WORLD DATA FOR OBSERVATIONAL AND INTERVENTIONAL STUDIES

P-B02-01	**Elena Albu** Leuven BE	Comparison of random forests modelling strategies for EHR data in the presence of competing risks – A case study on central line-associated bloodstream infections
P-B02-02	**Faye Baldwin** Liverpool UK	Methodological considerations for target trial emulation: A systematic review
P-B02-03	**Giuseppe Occhino** Bari IT	Assessment of frailty in the general population based on electronic health data
P-B02-04	**Cinzia Anna Maria Papappicco** Padua IT	A new strategy to automate the high-volume integration of near real-time data streams in the international multicentre CRICKET (Critical Events in Anaesthetised Kids Undergoing Tracheal Intubation) study
P-B02-05	**Myanca Rodrigues** Toronto CA	Comparative analysis of methods for identifying multimorbidity patterns among people with opioid use disorder in Ontario, Canada
P-B02-06	**Lutfiyya Muhammad** Chicago IL US	Jointly estimating treatment dosage and number of mature oocytes for oocyte cryopreservation using a national clinical outcome database
P-B02-07	**Alexandre Civet** Boulogne Billancourt FR	Propensity score matching to compare the healthcare resource utilisation of patients with multiple sclerosis using the French national health data system

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

P-B02-08	**Audrey D'Abrigeon** Montreuil FR	Integration of historical data into the design and analysis of clinical trials for rare diseases
P-B02-09	**Alice-Maria Toader** Liverpool UK	The use of healthcare systems data for RCTs
P-B02-10	**Rossana Di Staso** Bologna IT	Social support as a key determinant of resilience and well-being of mothers: A network analysis of data from a general population survey
P-B02-11	**Joachim Tan** London UK	Using simulations to compare cohort and case-control study designs for evaluating education outcomes of children with congenital or rare diseases
P-B02-12	**Georgy Gomon** Leiden NL	Missing data in routinely collected electronic health records: An approach to characterize different levels of missing data
P-B02-13	**Marc Dibling** Paris FR	Care pathway heterogeneity in Amyotrophic Lateral Sclerosis: Effects of gender, age and onset
P-B02-14	**Yat Yi Fan** Liverpool UK	Investigating the potential anti-viral effects of Proton Pump Inhibitors on influenza and COVID-19: Intention-to-treat trial emulation using electronic health records
P-B02-15	**Lynne Giles** Adelaide AU	Congenital heart defects and educational outcomes: Findings from a South Australian data linkage study
P-B02-16	**Simona Hapca** Stirling UK	The relationship between acute kidney injury and chronic kidney disease in patients with Type 2 Diabetes: An observational cohort study
P-B02-17	**Lara Lewis** Durban ZA	Impact of clinic community ART referral on service delivery for HIV patients newly initiated on antiretroviral therapy
P-B02-18	**Masako Nishikawa** Tokyo JP	Re-examination of the threshold value of HbA1C in therapeutic guideline of type 2 diabetic patients taking circannual rhythms into account using large registry dataset
P-B02-19	**Paulina Von Stackelberg** Amsterdam NL	Predictive monitoring of depression relapse using administrative data
P-B02-20 LB	**Isao Yokota** Sapporo JP	Comparison of analysis methods for nested case-control studies with risk set sampling in the presence of tied data
	POSTER SESSION B-03: SURVIVAL ANALYSIS	
P-B03-01	**Gizem Ayna Duran** Izmir TR	Identification and validation of survival-associated hub genes in the VAV1 gene network in acute myeloid leukemia
P-B03-02	**Aishwarya Bhaskaran** Sydney AU	A maximum penalised likelihood approach for semiparametric accelerated failure time models with time-varying covariates and partly interval censoring
P-B03-03	**Marie Skov Breum** Copenhagen DK	Efficient nonparametric estimators of discrimination measures with censored survival data
P-B03-04	**Roxane Couturier** Paris FR	Evaluation of the Survival-inferred fragility index to assess the robustness of the estimated treatment effect on survival endpoints
P-B03-05	**Maria De Martino** Udine IT	Assessing different matching methods to evaluate the effectiveness of colchicine in treating myopericarditis

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

P-B03-06	**Fatih Kizilaslan** Oslo NO	Comprehensive analysis of parametric and semiparametric mixture cure models from low- to high-dimensional covariates settings
P-B03-07	**Shirin Moghaddam** Limerick IE	Penalised power-generalised Weibull distributional regression
P-B03-08	**Marilena Müller** Heidelberg DE	Cox regression for comparing a mixture of personalized treatment and standard of care to standard of care
P-B03-09	**Parastoo Niloofar** Galway IE	Translating time to event data using the mean residual function
P-B03-10	**Tomomi Nishikawa** Chiba JP	Examination on analyses for progression free survival as interval-censored data
P-B03-11	**Hugo Pedder** Bristol UK	Information sharing in fractional polynomial network meta-analysis where studies report different summaries and subgroup categories: 2nd line treatments for Non-Small Cell Lung Cancer (NSCLC)
P-B03-12	**Marius Robert** Bordeaux FR	Lead time bias correction in breast cancer screening studies
P-B03-13	**Mathieu Roche** Tours FR	Estimation of the Kendall's Tau in cluster randomised trials with right censored time to event outcomes
P-B03-14	**Maral Saadati** Heidelberg DE	Identification of predictive factors using tree-based models for survival outcomes
P-B03-15	**Marije Sluiskes** Leiden NL	Peeling off the hazard layers: A comparison of relative survival techniques to disentangle different components of excess mortality in vulnerable groups during a crisis
P-B03-16	**Rachael Stannard** Leicester UK	How to make fair comparisons of population-based survival before, during, and after the COVID-19 pandemic: The role of stage-standardisation
P-B03-17	**Yuhe Wang** Leicester UK	Season of neonatal discharge and risk of unplanned Paediatric Intensive Care Unit admission: A use of flexible parametric modelling
P-B03-18	**Yuh-Jenn Wu** Taoyuan City TW	Bayesian inference of Case II interval-censored data with non-proportionality
P-B03-19	**Toby Hackmann** Leiden NL	Effective sample size for the Kaplan-Meier estimator: An intuitive measure of uncertainty?
P-B03-20	**Reshma Kassanjee** Cape Town ZA	The estimation of COVID-19 vaccine effectiveness by time since vaccination in a population-level observational cohort study using linkage of electronic medical records and survival models with time varying exposures
POSTER SESSION B-04: PREDICTION AND PROGNOSTIC MODELS		
P-B04-01	**Lasai Barreñada** Leuven BE	Multicentre flexible calibration curves with binary outcomes using random effects meta-analysis
P-B04-02	**Mae Chester-Jones** Oxford UK	Maternal Early Warning Scores for detecting deterioration in women during pregnancy: A systematic review

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

P-B04-03	**Gordon Forbes** London UK	Accounting for between-study heterogeneity when developing prediction models using IPD meta-analysis: A comparison of methods and recommendations for practice
P-B04-04	**Amy Magona** Oxford UK	Prediction performance effect of added-value biomarkers in multivariable heart failure prognostic models: A systematic review
P-B04-05	**Giuseppe Occhino** Bari IT	Prognostic models for heart failure progression: Systematic review and meta-analysis
P-B04-06	**Jyoti Sehjal** Oxford UK	Prediction models for prognostic outcomes in myelodysplastic syndromes and acute myeloid leukaemia: A systematic review
P-B04-07	**Gareth Ambler** London UK	A sample size calculation for developing a risk prediction model for a binary outcome
P-B04-08	**Menelaos Pavlou** London UK	How to improve model stability when calculating the sample size and accuracy when estimating a prediction model
P-B04-09	**Donna Ankerst** Munich DE	Selection and verification bias adjustments for clinical risk model validation
P-B04-10	**Frank Doornkamp** Leiden NL	Integrating genomic markers for enhanced prediction: Uncertainty assessment of clinical usefulness in early breast cancer
P-B04-11	**Evangelos Kritsotakis** Heraklion Crete GR	Statistical pitfalls and errors in external validation studies of clinical prediction models: A case study with the SAPS II and SAPS 3 prognostic models for critically ill patients
P-B04-12	**Yanakan Logeswaran** London UK	Examining model stability across modelling approaches to predicting poor functioning for individuals at-risk of psychosis
P-B04-13	**Ewan Carr** London UK	Development and validation of a novel longitudinal prognostic model for depressive relapse using passively measured biomarkers
P-B04-14	**Ryo Emoto** Nagoya JP	Development and validation of treatment effect functions in observational studies: Application to predictive analysis of amiodarone in out-of-hospital cardiac arrest patients
P-B04-15	**Shan Gao** Leuven BE	A comparison of regression models for static and dynamic prediction in electronic health records in the absence of censoring
P-B04-16	**Tommi Härkänen** Helsinki FI	Diagnostic tool to assess projected occupation probabilities in illness-death model
P-B04-17	**Joyce (Yun-Ting) Huang** Manchester UK	Quantifying the impact of underdiagnosis in outcome data when developing and validating a clinical prediction model: A simulation study
P-B04-18	**Seyed Amirhossein Jalali** Limerick IE	Prognosis value of immunocyte ratios in neoadjuvant chemotherapy-treated breast cancer
P-B04-19	**David Jenkins** Manchester UK	Use of statistical process control to monitor calibration-in-the-large of a clinical prediction model
P-B04-20	**Carolien Maas** Rotterdam NL	Predicting individualised treatment effects: A comparison of different modelling approaches and performance metrics

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

P-B04-21	**Ryunosuke Machida** Tokyo JP	Predicting study duration based on dynamic predictions using joint models in clinical trials with a time-to-event endpoint
P-B04-22	**Roberto Melotti** Bolzano/Bozen IT	Tremor classification by digital spiral analysis in a large population sample
P-B04-23	**Asanao Shimokawa** Tokyo JP	Construction of decision tree using prior information
P-B04-24	**Thomas Stojanov** Basel CH	Development and validation of a model predicting post-operative shoulder stiffness by patients undergoing an arthroscopic rotator cuff repair in Switzerland
P-B04-25	**Cristian Tebe** Barcelona ES	Can a clinical model be used to update a radiomics deep learning model for prediction of lung nodule malignancy?
P-B04-26	**David Van Klaveren** Rotterdam NL	Stronger penalties on treatment-covariate interactions improve the ability to predict treatment effect
P-B04-27	**Steven Wambua** Birmingham UK	Development and validation of postpartum cardiovascular disease (CVD) risk prediction model in women with a history of pregnancy incorporating reproductive and pregnancy-related candidate predictors.
P-B04-28	**Junfeng Wang** Utrecht NL	Comparison of different strategies in using Lasso in clinical prediction models for rare outcomes: A simulation study
P-B04-29	**Zhenwei Yang** Rotterdam NL	Predictive accuracy metrics in the context of interval censoring and competing risks
P-B04-30	**Lucinda Archer** Birmingham UK	Stop before you start: A checklist for those thinking about developing a clinical prediction model
P-B04-31	**Lucinda Archer** Birmingham UK	Uncertainty in clinical risk prediction: perspectives and approaches
P-B04-32 LB	**Amardeep Legha** Birmingham UK	Uncertainty-based sequential sample size calculations for developing, validating and updating clinical prediction models
POSTER SESSION B-05: DIAGNOSTIC AND SCREENING TESTS		
P-B05-01	**Alba M. Franco Pereira** Madrid ES	Inference on the symmetry point-based optimal cut-off point and associated sensitivity
P-B05-02	**Letizia Orsini** Stokholm SE	A joint model of tumour volume and mode of detection for a cohort of incident breast cancer cases
P-B05-03	**Tom Parry** London UK	Agreement of a latent class model with an expert panel for defining the reference standard in a diagnostic imaging study
P-B05-04	**Ioanna Tarelli** Thessaloniki GR	Reporting completeness of systematic reviews and meta-analysis of diagnostic test accuracy studies published in 2022 based on the PRISMA-DTA Reporting Guideline: An empirical study
P-B05-05	**Carsten Oliver Schmidt** Greifswald DE	Automated data screening and data quality checks using the dataquieR R package

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSION B-06: STATISTICAL ANALYSIS OF COMPLEX DATA STRUCTURES		
P-B06-01	**Hollie Hughes** Liverpool UK	Bayesian Spatial modelling of mental health incidence data: A comparison of approaches
P-B06-02	**Tina Košuta** Ljubljana SI	Data augmentation for improving parameter estimation in mixed models
P-B06-03	**Dae-Jin Lee** Madrid ES	Injurytools: A toolkit for sports injury data analysis
P-B06-04	**Anikó Lovik** Stockholm SE	A case study of combining cluster analyses in clinical trial data
P-B06-05	**Nicolas Sauvageot** Allschwil CH	Comparison of the statistical performance of generalized pairwise comparisons, LWYY and Cox models for composite endpoints including recurrent and terminal events in clinical trials
P-B06-06 CASc	**Shafayet Khan Shafee** Dhaka BD	Interval estimation of median odds ratio for measuring contextual effects in multilevel data using binary logistic model
P-B06-07	**Lubomir Stepanek** Prague CZ	Graphical inference tests revisited: Using two-dimensional grids in nonparametric testing for hypothesis evaluation and educational insights
P-B06-08	**George Vamvakas** London UK	Prediction of language growth scores from a two-stage population survey with longitudinal data
P-B06-09	**Yueyun Zhu** Galway IE	Estimation and application of derivative multivariate functional principal component analysis
POSTER SESSION B-07: VALIDATION OF SYNTHETIC DATA		
P-B07-01	**Hanning Yang** Freiburg DE	Calibrating representations of expert knowledge with patient data in latent spaces for synthetic trajectories
POSTER SESSION B-08: COMPETING RISKS AND MULTISTATE MODELS		
P-B08-01	**Caroline Chesang** London UK	Estimating treatment effects when there are competing risks using real-world data: Application to prostate cancer
P-B08-02	**Chantelle Cornett** Manchester UK	Penalisation methods for multi-state models: A comparative, simulation study
P-B08-03	**Ömer Faruk Dadaş** Izmir TR	A comparison between the joint model and landmark approach for dynamic prediction of competing risks in survival analysis with time-dependent covariates
P-B08-04	**Lorenzo Del Castello** Milan IT	An R function for data preparation for an acyclic multistate model with non-ordered intermediate states
P-B08-05	**Maryam Farhadizadeh** Freiburg DE	Enhancing healthcare understanding from clinical routine data by simplifying the representation of treatment pathways
P-B08-06	**Derek Hazard** Freiburg DE	Balancing speed and detail: Assessing multi-state prediction models in the face of emerging pathogens
P-B08-07	**Jonatan Hedberg** Mölndal SE	Marginal analysis of time spent hospitalised or exacerbating
P-B08-08	**Sandra Schmeller** Ulm DE	The challenge of time-to-event analysis for multiple events: A guidance

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSION B-09: INTEGRATIVE DATA ANALYSIS		
P-B09-01	**Camilo Broc** Lyon FR	Integrated analysis of humoral immune response across several COVID immunisation schemes in COVIDAuRA study
P-B09-02	**Md Mynul Islam** Dhaka BD	Child malnutrition at Upazila Level in Bangladesh: A small-area estimation approach
POSTER SESSION B-10: LONGITUDINAL DATA ANALYSIS		
P-B10-01	**Mélanie Guhl** Paris FR	Semi-Bayesian methods to compute standard errors in nonlinear mixed effects models for sparse data
P-B10-02	**Markus Schepers** Mainz DE	Bayesian mixed models approach to exploring resilience dynamics: Impact of stress on subjective health and affects over time during the COVID-19 pandemic
P-B10-03	**Birzhan Akynkozhayev** Stockholm SE	Longitudinal and joint models for PSA values and time to prostate cancer diagnosis
P-B10-04	**Yanjie Wang** Guangzhou CN	A dynamic prediction model for predicting the time to conversion to AD in patients with MCI Based on time-dependent covariates
P-B10-05	**Jamie Wong** London UK	Emulating hypothetical physical activity interventions to assess obesity risk reduction: Applying target trial emulation in the 1958 British Birth Cohort
P-B10-06	**Denitsa Grigorova** Sofia BG	Modelling longitudinal cognitive test data with ceiling effects and left skewness
P-B10-07	**Sofia Kaisaridi** Paris FR	A multivariate disease progression model for identifying subtypes in CADASIL
P-B10-08	**Maria De Martino** Udine IT	Use of interrupted time-series analysis to examine the progression of respiratory viruses amidst the SARS-CoV-2 pandemic
P-B10-09	**Juliette Ortholand** Paris FR	LEASPY: LEArning Spatiotemporal patterns in Python
P-B10-10	**Doug Thompson** Kilmacolm UK	Design considerations for continuous time models in progressive diseases
P-B10-11	**Kiana Farhadyar** Freiburg DE	Impact of different longitudinal data representations on transformer performance in small data applications
P-B10-12	**Komal Aryal** Hamilton CA	Identifying factors associated with physician assistance in dying for older adults: A cohort study
P-B10-13	**Bashayr Aldawsari** Liverpool UK	Longitudinal clustering of liver cancer biomarkers
P-B10-14	**Francesca Little** Cape Town ZA	Examining heterogeneity in longitudinal data using latent class mixed effect modelling
P-B10-15	**Marta Spreafico** Leiden NL	Cluster-based recurrent marked point process approach for longitudinal volume-outcome studies
P-B10-16 LB	**Ottavio Khalifa** Paris FR	Identifying clustering methods for longitudinal data with categorical features: A scoping review

WEDNESDAY, 24 JULY 2024

POSTER AREA
(Museum Hall)

POSTER SESSION B-11: ANALYSIS OF IMPERFECT DATA (INCLUDING MISSING DATA & MEASUREMENT ERRORS)		
P-B11-01	**Joydeep Basu** Coventry UK	Estimating treatment effect in longitudinal trials with monotone missingness for mixed observations
P-B11-02	**Vineet Kumar Kamal** Kalyani IN	Assessment and comparison of classical and machine learning approaches for imputing missing data in the framework of excess hazard analysis
P-B11-03	**Michail Katsoulis** London UK	A data driven approach to address missing data in the 1970 British birth cohort
P-B11-04	**Natasa Kejzar** Ljubljana SI	Biomarkers with limit of detection: Modelling composite event in patients with PAD
P-B11-05	**Matteo Petrosino** Milan IT	The use of Joint and Cox models to assess the association between a longitudinal marker and a time-to-event: A simulation study under different missing mechanisms and applications in ICU setting
P-B11-06	**Bohan Zhang** Manchester UK	Investigating the performance of missing data methods when applied to time-series environmental exposure data: A simulation study
P-B11-07	**Saghar Garayemi** Augsburg DE	Missing data imputation in the context of propensity score analysis: A systematic review
P-B11-08	**Emily Kawabata** Bristol UK	Scoping review of software implementations of quantitative bias analysis to informatively missing data
P-B11-09	**Neža Dvoršak** Bath UK	Accounting for missing data when calculating conditional/predictive power
P-B11-10	**Dominik Grathwohl** Lausanne CH	Quantifying the benefits of double measurements for anthropometric estimates in paediatric clinical trials
P-B11-11	**Stanislav Katina** Prague CZ	Methods of estimating parameters of skewed or truncated normal distribution in the presence of observations outside of measurable range
P-B11-12	**Tamae Kawasaki** Tokyo JP	Maximum likelihood estimation of the correlation coefficient for a trivariate normal distribution with missing data
P-B11-13	**Toshiro Tango** Tokyo JP	Can we estimate a risk without observing the relevant number of cases?
POSTER SESSION B-12: RARE DATA ANALYSIS OR CASE STUDIES IN MEDICAL STATISTICS		
P-B12-01	**Audrone Jakaitiene** Vilnius LT	Rare variants statistical tests for analysing genetic changes in 15Q13.3
P-B12-02	**Adnan Karaibrahimoglu** Isparta TR	On the comparison of Classical, Bayesian and Jackknife binary logistic regression models in paediatric epilepsy patients
P-B12-03	**Cinzia Anna Maria Papappicco** Padua IT	When the size matters: A simulation study evaluating stability of variable importance driving pancreas re-transplant graft survival in a very small dataset using machine learning techniques

SPONSORS ACKNOWLEDGEMENT

Silver Sponsor

Pfizer

Bronze Sponsors

Pharmathen
Innovation inspired by life

GSK

MSD

Exhibitor

TStat

Supporters

Bank of Greece
Eurosystem

90 ΧΡΟΝΙΑ
Παράγουμε Ελληνικό Φάρμακο

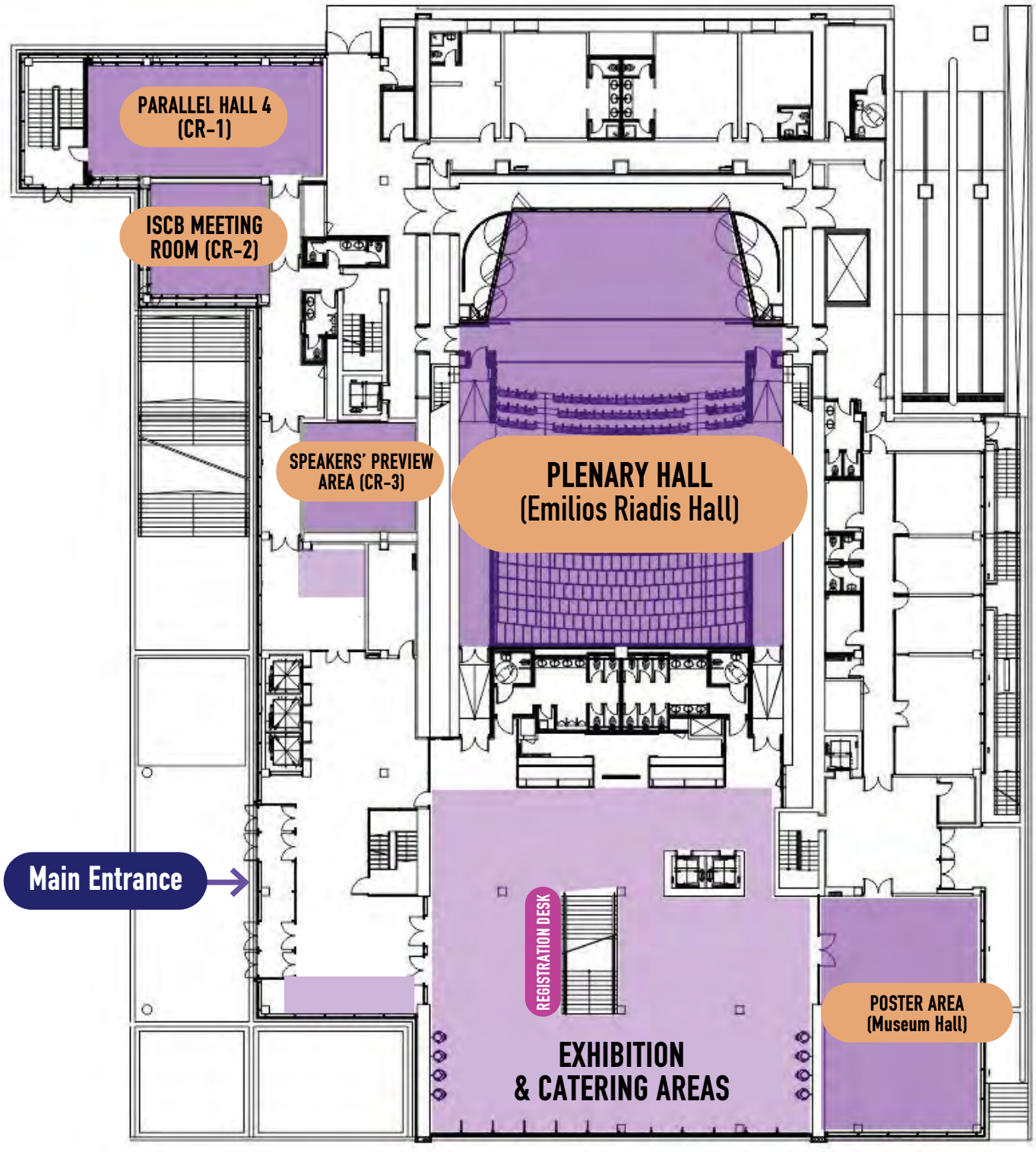
Πανελλήνια Ένωση Φαρμακοβιομηχανίας

Boehringer Ingelheim

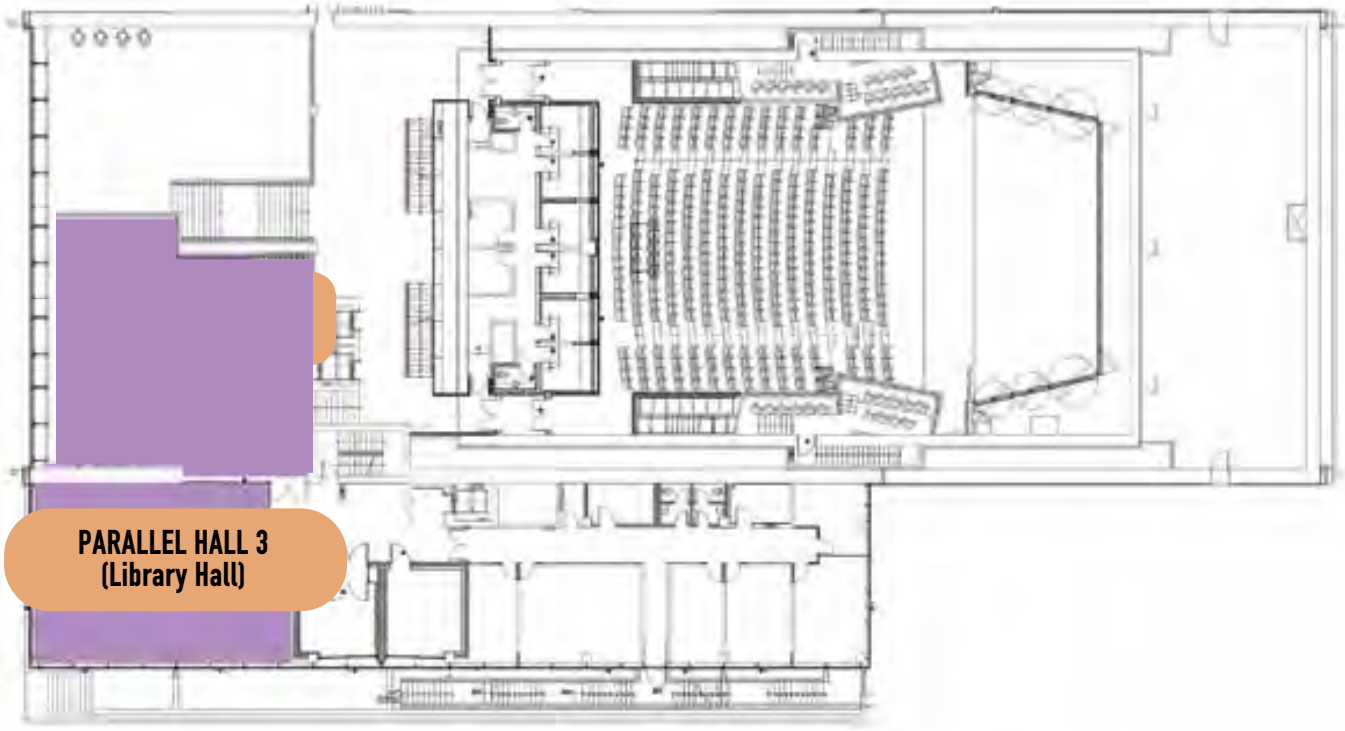
Printed Programme Publication

propobos
PUBLICATIONS

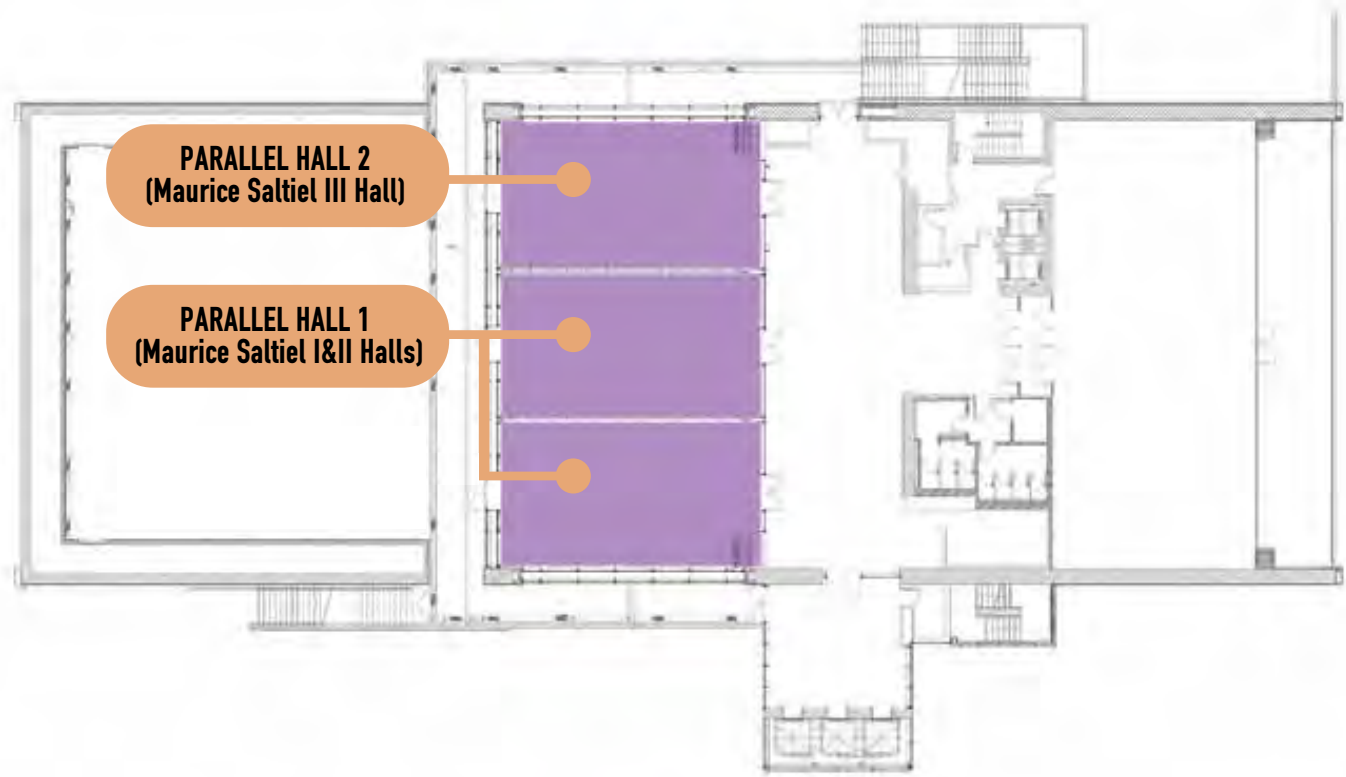
M2 Building - Ground floor



M2 Building - Level 2



M2 Building - Level 5



DOWNLOAD THE OFFICIAL CONFERENCE APP

1 **Conference App Download Instructions**
Use your device to scan the QR Codes

ANDROID DEVICES

GET IT ON Google play

iOS DEVICES

Download on the App Store

2 Download The Event App by EventsAIR

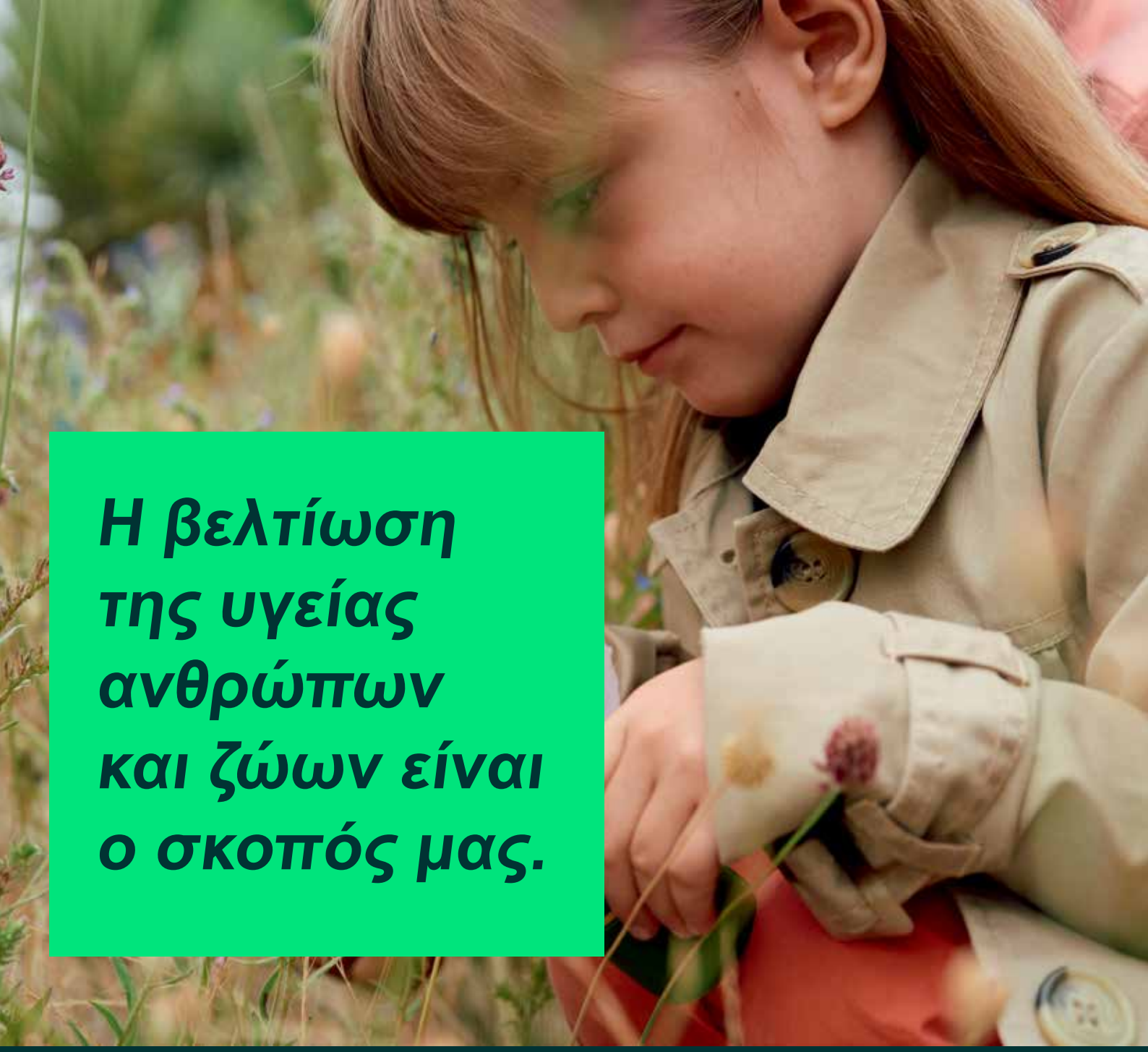
3 When you are prompted for a code upon launching the App, insert **iscb45**

In order to log in, tap the Log In icon and insert your email address and the 4-digit PIN that has been sent to you.

FREE WiFi
iscb45

➜ @ISCB_info

Official Hashtags ➜ #ISCB45



Boehringer Ingelheim

Life forward

Boehringer Ingelheim Ελλάς Μονοπρόσωπη Α.Ε, Αριθμός ΓΕΜΗ: 122178501000, Λεωφ. Ανδρέα Συγγρού 340, 176 73 Καλλιθέα. Τηλ.: 210 89 06 300
Γραφείο Μακεδονίας - Θράκης: Αντώνη Τρίτση 15-17 & Μαρίας Κάλλας 6, Πυλαία. 555 35 Θεσσαλονίκη. Τηλ.: 2310 424 618.
E-mail: info@ath.boehringer-ingelheim.com | www.boehringer-ingelheim.gr

90

ΧΡΟΝΙΑ

Παράγουμε Ελληνικό Φάρμακο

ΓΙΑ ΑΣΘΕΝΕΙΣ | ΚΟΙΝΩΝΙΑ | ΟΙΚΟΝΟΜΙΑ

για την

ΥΓΕΙΑ

- Το Ελληνικό Φάρμακο μπορεί να καλύψει άμεσα το 70% των αναγκών της **ΠΡΩΤΟΒΑΘΜΙΑΣ ΠΕΡΙΘΑΛΨΗΣ** και το 50% των αναγκών της **ΝΟΣΟΚΟΜΕΙΑΚΗΣ ΠΕΡΙΘΑΛΨΗΣ**
- Η ανάδειξη και στήριξη του Ελληνικού Φαρμάκου είναι **ΕΘΝΙΚΗ ΠΡΟΤΕΡΑΙΟΤΗΤΑ**, με μια **ΕΘΝΙΚΗ ΠΟΛΙΤΙΚΗ ΦΑΡΜΑΚΟΥ** και με **ΚΙΝΗΤΡΑ** σε Γιατρούς & Φαρμακοποιούς

Με αυτόν τον τρόπο μπορούμε να πετύχουμε:

- **ΜΕΙΩΣΗ** στις άσκοπες εισαγωγές ακριβών φαρμάκων
- **ΠΕΡΙΟΡΙΣΜΟ** του ελλείμματος στο ισοζύγιο εμπορικών συναλλαγών στα φάρμακα
- **ΕΞΟΙΚΟΝΟΜΗΣΗ** πόρων για τα καινοτόμα φάρμακα που πραγματικά χρειαζόμαστε

για την

ΚΟΙΝΩΝΙΑ

Το Ελληνικό Φάρμακο είναι:

- **ΑΠΟΤΕΛΕΣΜΑΤΙΚΟ** και **ΑΣΦΑΛΕΣ** γιατί πιστοποιείται από τους σημαντικότερους Οργανισμούς Φαρμάκου παγκοσμίως
- **ΠΟΙΟΤΙΚΟ** γιατί παράγεται σε ελληνικά εργοστάσια που ακολουθούν αυστηρά τα διεθνή πρότυπα διασφάλισης ποιότητας
- **ΔΙΕΘΝΩΣ ΑΝΑΓΝΩΡΙΣΜΕΝΟ** γιατί εκατομμύρια ασθενείς σε περισσότερες από 140 χώρες το εμπιστεύονται καθημερινά
- **ΠΡΟΣΙΤΟ** για τον Έλληνα ασθενή γιατί μειώνει το κόστος συμμετοχής των ασφαλισμένων

για την

ΑΝΑΠΤΥΞΗ

- Παράγεται σε 45 υπερσύγχρονα ελληνικά **ΕΡΓΟΣΤΑΣΙΑ**
- Επηρεάζει πάνω από **53.000 ΘΕΣΕΙΣ ΕΡΓΑΣΙΑΣ**
- Κατέχει την **2η ΘΕΣΗ ΣΤΙΣ ΕΞΑΓΩΓΕΣ** της Ελλάδας
- Εξάγεται σε **147 ΧΩΡΕΣ**
- Για κάθε €1.000 δαπάνης το **ΑΕΠ ΕΝΙΣΧΥΕΤΑΙ** κατά €3.420
- Συμβάλει συνολικά με **€2,8 δισ.** το χρόνο στο **ΑΕΠ**
- Η ελληνική φαρμακοβιομηχανία υλοποιεί την πενταετία 2021-2025 ένα μεγάλο επενδυτικό πρόγραμμα ύψους **1,2 δισ. Ευρώ** σε παραγωγή, έρευνα και ανάπτυξη.

ΠΑΝΕΛΛΗΝΙΑ ΕΝΩΣΗ ΦΑΡΜΑΚΟΒΙΟΜΗΧΑΝΙΑΣ

Καινοτομία εμπνευσμένη από τη ζωή

Στην Pharmathen αλλάζουμε το μέλλον της υγείας, επενδύοντας συνεχώς στην έρευνα και την ανάπτυξη τεχνολογιών αιχμής. Αναπτύσσουμε, παράγουμε και διανέμουμε σύνθετα καινοτόμα φαρμακευτικά σκευάσματα σε περισσότερες από 90 χώρες, βελτιώνοντας την ποιότητα ζωής των συνανθρώπων μας σε ολόκληρο τον κόσμο.

Pharmathen

Innovation inspired by life

www.pharmathen.com

KEYNOTE LECTURES

KEYNOTE SPEAKER

David Spiegelhalter
University of Cambridge, United Kingdom

Trustworthy Communication of Statistical Evidence: What is it, and how we can get more of it

The recent pandemic has emphasised the key role played by evidence based on data. But how do we decide whether to trust all the claims that are made, whether in mainstream or social media? Are the numbers being used to manipulate us? What questions should we ask?

I will look at the way that statistics can be used to try and persuade audiences to think or act in a certain way, and contrast this with efforts to make communication 'trustworthy', by presenting balanced information that seeks to inform rather than persuade.

I will also talk about my own experience in trying to communicate statistics in the media, including both successes and disasters.

PRESIDENT'S INVITED SPEAKER

Chris Holmes
University of Oxford, United Kingdom

Biostatistics in the AI era

Artificial intelligence (AI) has emerged as valuable tool in data science and of growing influence in medical research, with an accelerating pace of innovation. The rapid development of AI is driven, in part, by the ongoing expansion in computer power and data availability. However, the same features that make AI such a valuable resource make it vulnerable from a statistical perspective. This contradiction is particularly pertinent for medical science. Techniques that are adequate for targeted advertising to consumers or that enhance weather prediction may not meet the rigorous demands needed for risk prediction or diagnosis in medicine.

I will discuss the statistical challenges in developing and targeting AI to biomedical data analysis and the delicate balance that researchers face in wishing to learn as much as possible from data while ensuring that data-driven conclusions are accurate, robust, and reproducible. I will emphasise examples where statistical analysis benefits from using AI, and in response, where the field of AI benefits from statistical thinking.

INVITED SESSIONS

INVITED SESSION 1 — Causal Inference and Machine Learning

IS1-1 Double robust: A great asset not to be constructed lightly

Erin Gabriel
University of Copenhagen, Denmark

Double robustness has been a vehicle through which machine learning entered the causal inference literature - even for randomized clinical trials. There are now many options for doubly robust estimation. However, there is a concerning trend in the applied literature where authors believe that any combination of a propensity score and an adjusted outcome model automatically results in a doubly robust estimator. This is particularly true for inverse probability of treatment (propensity score) weighted (IPTW) Cox PH models, which is the method used in several high-impact applied papers. We will demonstrate that IPTW PH models, Cox or parametric, are not doubly robust away from the null. This is not to say that there are no simple doubly robust estimators. Canonical link generalized linear models (GLM) fit via inverse probability of treatment (propensity score) weighted maximum likelihood estimation followed by standardization (the g-formula) for the average causal effect, which we call IPTW GLM, is a doubly robust estimation method. Understanding how IPTW GLM is doubly robust and how IPTW Cox, or IPTW non-canonical link GLM, is not, requires clarity on several concepts that are often assumed known in the causal inference literature and thus confused in the applied literature. We aim to provide a clear outline of these concepts, in addition to simulated and real data examples.

References: *[1] Statistics in Medicine 2023: Gabriel EE, et al. Inverse probability of treatment weighting with generalized linear outcome models for doubly robust estimation. https://arxiv.org/abs/2310.16207. [2] Gabriel EE, et al. Propensity score weighting plus an adjusted proportional hazards model*

IS1-2 Causal machine learning with DoubleML: An introduction and applications

Martin Spindler
Universität Hamburg, Germany

Causal Machine Learning (CausalML) has emerged as new field combining causal inference and modern machine learning methods. The so-called Double / Debiased Machine Learning (DML) allows for valid inference in high-dimensional complex settings using ML. In this talk an introduction to the DML framework and the DoubleML package (R/Python) is given. Moreover, applications and use cases in health economics are presented and an outlook to recent developments is given.

INVITED SESSIONS

IS1-3 Efficient and robust machine-learning-based approaches for simple, cluster randomized, and sequential multiple assignment randomised trial analysis: Illustrations from HIV trials in East Africa

Maya Petersen
University of California, Berkeley, United States

Covariate adjustment in randomized trials can improve precision and power while maintaining rigorous inference. Targeted maximum likelihood estimation (TMLE) is a general semi-parametric efficient approach for optimizing precision gains from covariate adjustment in trials by combining 1) flexible, machine learning-based outcome regression; and, 2) additional covariate adjustment in the treatment mechanism. This general method can be used across a range of trial designs, including those employing cluster, restricted, sequential, and adaptive randomization schemes, and can be extended to account for loss-to-follow-up, missing data, competing risks, and intercurrent events.

Using examples from completed HIV trials in Kenya and Uganda that used TMLE as the primary pre-specified analysis, we illustrate several design-specific TMLE estimators. We first discuss the SEARCH Study (NCT01864603), a trial in Uganda and Kenya of universal HIV testing and person-centered multi-disease treatment to reduce HIV incidence and improve community health. Using SEARCH, we illustrate implementation of TMLE for the analysis for cluster randomized data, with and without pair-matching. In particular, we illustrate how "adaptive-pre-specification", an empirical efficiency maximization strategy, can be employed to achieve meaningful precision gains even in the context of cluster-randomized trials with few independent units.

We then use the ADAPT-R Study (NCT02338739), a trial of sequential behavioral intervention strategies to improve retention in HIV care among people with HIV in Kenya, applying TMLE to sequential multiple assignment randomized trials (SMARTs). Adaptive intervention strategies (or dynamic regimes), in particular those that modify treatment based on a participant's own response, are a core component of precision health approaches. SMART designs are growing in popularity and are specifically designed to facilitate the evaluation of sequential adaptive strategies. We present a robust and efficient approach using TMLE for estimating and contrasting expected outcomes under the dynamic regimes embedded in a SMART, together with generating simultaneous confidence intervals for the resulting estimates. We contrast this method with two alternatives (G-computation and inverse probability weighting estimators).

References: *[1] Balzer LB, van der Laan M, Ayieko J, Kamya M, Chamie G, Schwab J, Havlir DV, Petersen ML. Two-Stage TMLE to reduce bias and improve efficiency in cluster randomized trials. Biostatistics. 2023 Apr 14;24(2):502-517. doi: 10.1093/biostatistics/kxab043. PMID: 34939083; PMCID: PMC10102904. [2] Montoya LM, Kosorok MR, Geng EH, Schwab J, Odeny TA, Petersen ML. Efficient and robust approaches for analysis of sequential multiple assignment randomized trials: Illustration using the ADAPT-R trial. Biometrics. 2023 Sep;79(3):2577-2591. doi: 10.1111/biom.13808. Epub 2022 Dec 22. PMID: 36493463; PMCID: PMC10424093.*

INVITED SESSIONS

INVITED SESSION 2

Recent advances in survival analysis with complex data structures

IS2-1 Joint modelling of (un)bounded longitudinal markers, competing risks, and recurrent events in Cystic Fibrosis data

Eleni-Rosalina Andrinopoulou
Erasmus University, the Netherlands

Joint models for longitudinal and survival data have become a popular framework for studying the association between repeatedly measured biomarkers and clinical events. Nevertheless, addressing complex data structures remains a challenge. In particular, even though many applications consist of recurrent and competing event times, those are commonly not analysed using a single model. Moreover, existing frameworks rely mainly on a Gaussian distribution for continuous markers, which may be unsuitable for bounded biomarkers, resulting in biased estimates of associations. To address these limitations, we propose a Bayesian shared-parameter joint model that simultaneously accommodates multiple (possibly bounded) longitudinal markers, a recurrent event process, and competing risks. We use the beta distribution to model responses bounded within any interval without sacrificing the interpretability of the association. The model offers various forms of association, discontinuous risk intervals, and both gap and calendar timescales. We analyse the US Cystic Fibrosis Foundation Patient Registry to study the associations between changes in lung function and body mass index, and the risk of recurrent pulmonary exacerbations, while accounting for the competing risks of death and lung transplantation. Our comprehensive approach provides new insights into cystic fibrosis progression.

IS2-2 Semiparametric estimation of misclassified semi-competing risks data under gamma-frailty conditional Markov model

Ying Zhang
University of Nebraska Medical Center, United States

Joint models for longitudinal and survival data have become a popular framework for studying the association between repeatedly measured biomarkers and clinical events. Nevertheless, addressing complex data structures remains a challenge. In particular, even though many applications consist of recurrent and competing event times, those are commonly not analysed using a single model. Moreover, existing frameworks rely mainly on a Gaussian distribution for continuous markers, which may be unsuitable for bounded biomarkers, resulting in biased estimates of associations. To address these limitations, we propose a Bayesian shared-parameter joint model that simultaneously accommodates multiple (possibly bounded) longitudinal markers, a recurrent event process, and competing risks. We use the beta distribution to model responses bounded within any interval without sacrificing the interpretability of the association. The model offers various forms of association, discontinuous risk intervals, and both gap and calendar timescales. We analyse the US Cystic Fibrosis Foundation Patient Registry to study the associations between changes in lung function and body mass index, and the risk of recurrent pulmonary exacerbations, while accounting for the competing risks of death and lung transplantation. Our comprehensive approach provides new insights into cystic fibrosis progression.

INVITED SESSIONS

IS2-3 Bayesian semiparametric modeling of spatially-referenced multistate current status data

Dipankar Bandyopadhyay
Virginia Commonwealth University, United States

Assessment of multistate disease progression is commonplace in biomedical research, such as, in periodontal disease (PD). However, the presence of current status (a severe form of interval-censoring) endpoints, where a single snapshot of study subjects transitioning through a sequence of well-defined disease states at random inspection times, complicates the inferential framework. In addition, these endpoints can be clustered (tooth-level event status within subjects), and spatially associated, where a group of proximally located teeth may experience similar PD status compared to distally-located teeth. Failure to adjust for the aforementioned complexities may lead to biased and imprecise inference. Motivated by a clinical study recording multistate event time progression of PD, we propose a Bayesian semiparametric accelerated failure time model with a Wishart proposal for accommodating (spatial) random effects, and flexible errors that follows a Dirichlet process mixture of Gaussians. For elegant clinical interpretation, the systematic component of the event times is modeled using a monotone single index model, whose (unknown) link function is estimated via a novel integrated basis expansion, with basis coefficients enjoying constrained Gaussian process priors. In addition to promising parameter identifiability, we present scalable computing via a combination of elliptical slice sampling, fast circulant embedding techniques, and smoothing of hard constraints, leading to straightforward computation of the parameter estimates, state occupation, and transition probabilities from posterior estimates. Using synthetic data, we study the finite sample properties of our Bayesian estimates, and its performance under model misspecification. We also illustrate our methodology via application to a real dataset recording PD status of Type-2 diabetic African-Americans living in coastal South Carolina.

INVITED SESSIONS

INVITED SESSION 3 — Innovative Complex Adaptive Designs for Confirmatory Clinical Trials with Multiple Primary Research Questions

IS3-1 Graph Based Adaptive MAMS Designs for Trials with Multiple Endpoints

Cyrus Mehta
President, Cytel Inc.; Harvard University, Cambridge, MA, United States

The graph-based approach is an extremely powerful and intuitive tool for designing trials that have multiple endpoints. This method enables a study team to represent clearly its priorities for hierarchical testing of the endpoints, and for propagating the available type-1 error from rejected hypotheses to hypotheses yet to be tested. While originally developed for single stage non-adaptive designs it has recently been extended to two-stage designs that permit adaptive sample size re-estimation, dropping of hypotheses, and changes in the hierarchical testing strategy at the end of stage one. Two approaches are available for preserving the family wise error rate (FWER) in the presence of these adaptive changes; the p-value combination (PV) method, and the conditional error rate (CER) method. In this session we will present the statistical methodology underlying each approach and will compare the operating characteristics of the two methods in a large simulation experiment.

IS3-2 Pairwise and familywise error rate control in platform trials: Impact on sample size, trial timelines and analysis

Babak Choodari-Oskooei, Mahesh (KB) Parmar[1], David Robertson[2]
1 MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, London, United Kingdom
2 MRC Biostatistics Unit, Cambridge University, Cambridge, United Kingdom

The operating characteristics of confirmatory randomised clinical trials are required to be controlled at prespecified levels. Traditionally, the overall type I error rate has been used to quantify the probability of false positive conclusions. It is often a key quantity of interest for both regulators and reviewers since it ensures the generalisability and reproducibility of trial results.

In platform trials and designs with a master protocol, there are two main measures of type I error rate: pairwise type I error rate; and familywise type I error rate. Recently, other performance measures such as the false discovery rate and "online" versions of these measures have been proposed for platform trials. However, there is no clear consensus in the statistical community about which of these quantities should be controlled in such trials. Further, the impact of the chosen measure on design and analysis of trials with master protocols have not been fully explored.

This talk provides an overview of the proposed performance measures and offers practical guidelines for selecting the most appropriate performance measure for the specific platform trial setting. Using the flagship STAMPEDE platform trial, it explores the impact of the chosen measure on the trial efficiency, i.e., sample size and timelines, as well as the analysis of the outcome measures.

The definitions of the false discovery rate, pairwise and familywise type I error rates as well as the concept of online error rates in trials with master protocol will be presented. We explain when each measure should be controlled at the pre-specified level, and what factors drive their values. We illustrate the degree of efficiency loss if an inappropriate measure of type I error rate is controlled. For online error rate control, the choice of method and the consideration of the order in which treatment arms enter the trial are crucial.

In platform trials, whenever possible, multiple distinct research questions should be addressed. This provides the rationale for targeting control of pairwise error rate. It is challenging to control the familywise error rate for all pairwise comparisons in a trial that adds new research arms. Either the number of new research arms should be limited, or in many situations a high price should be paid in terms of efficiency if new research arms are not added.

INVITED SESSIONS

IS3-3 Using Bayesian methods to include non-concurrent controls in the analysis of platform trials: benefits and limitations

Annette Kopp-Schneider, Vivienn Weru, Manuel Wiesenfarth, Silvia Calderazzo
German Cancer Research Center, Heidelberg, Germany

An appealing feature of platform trials is the fact that the trial can include one shared control arm for several treatment arms that enter over time. If a treatment arm is included after the start of the platform trial, this implies that not only concurrent but also non-concurrent controls recruited before the start of this arm are available for comparison. One approach to deal with this situation is to consider the controls as historical/external data that inform the comparison to the concurrent controls (see, e.g. Bofill Roig et al 2023). Many approaches for borrowing from external data have been proposed. Even though these methods are mainly based on Bayesian approaches by incorporating external information into the prior for the current analysis, frequentist operating characteristics of the analysis strategy are of interest. In particular, type I error and power at a prespecified point alternative are in the focus. For a fair comparison of test procedures without and with borrowing, the tests are calibrated to achieve the same type I error rate (Kopp-Schneider et al. 2024). We will consider approaches that dynamically borrow information according to the similarity of current and external data, e.g. the power prior approach that incorporates external data in the prior used for analysis of the current data. This prior is proportional to the likelihood of the external data raised to the power of a weight parameter. An Empirical Bayes approach for the estimation of the weight parameter from the similarity of external and current data has been proposed by Gravestock et al. (2017). We will also consider the robust mixture prior approach (Schmidli et al, 2014), a popular method that uses a weighted mixture of an informative and a more dispersed prior to address potential prior-data conflict and robustify the analysis. In the frequentist framework, power gains are not possible when borrowing external control data to the current trial, a finding that had been proven in general before (Kopp-Schneider et al. 2020). In fact, we have observed that the power in a comparison including non-concurrent controls may even lead to power losses compared to the test calibrated to borrowing.

References: *[1] Bofill Roig M et al (2023). Trials, 24(1), 408. [2] Gravestock I, Held L et al (2017). Pharmaceutical Statistics 16:349-360. [3] Kopp-Schneider A et al (2020). Biom J 62(2): 361-374. [4] Kopp-Schneider A et al (2024). Pharm Stat 23(1):4-19. [5] Schmidli H et al (2014) Biometrics 70(4), 1023-1032.*

INVITED SESSIONS

INVITED SESSION 4 — Optimal individualized treatment rules

IS4-1 Robust Sample Weighting to Facilitate Individualized Treatment Rule Learning for a Target Population

Menggang Yu
University of Michigan, United States

We consider a setting when a study or source population for individualized-treatment-rule (ITR) learning can differ from the target population of interest. We assume subject covariates are available from both populations, but treatment and outcome data are only available from the source population. Existing methods use "importance" and/or "overlap" weights to adjust for the covariate differences between the two populations. We develop a general weighting framework that allow a better bias-variance trade-off than existing weights. Our method seeks covariate balance over a non-parametric function class characterized by a reproducing kernel Hilbert space. Our weights encompass the importance weights and overlap weights as special cases. Numerical examples demonstrate that our weights can improve many ITR learning methods for the target population that rely on weighting.

IS4-2 Policy learning with distributional welfare

Sukjin Han
University of Bristol, United Kingdom

In this paper, we explore optimal treatment allocation policies that target distributional welfare. Most literature on treatment choice has considered utilitarian welfare based on the conditional average treatment effect (ATE). While average welfare is intuitive, it may yield undesirable allocations especially when individuals are heterogeneous (e.g., with outliers)---the very reason individualized treatments were introduced in the first place. This observation motivates us to propose an optimal policy that allocates the treatment based on the conditional quantile of individual treatment effects (QoTE). Depending on the choice of the quantile probability, this criterion can accommodate a policymaker who is either prudent or negligent. The challenge of identifying the QoTE lies in its requirement for knowledge of the joint distribution of the counterfactual outcomes, which is generally hard to recover even with experimental data. Therefore, we introduce minimax policies that are robust to model uncertainty. A range of identifying assumptions can be used to yield more informative policies. For both stochastic and deterministic policies, we establish the asymptotic bound on the regret of implementing the proposed policies. In simulations and two empirical applications, we compare optimal decisions based on the QoTE with decisions based on other criteria. The framework can be generalized to any setting where welfare is defined as a functional of the joint distribution of the potential outcomes.

INVITED SESSIONS

IS4-3 Functional additive models for interaction effects between a treatment and functional covariates

Hyung Park
New York University, United States

A primary goal of precision medicine is to make efficient use of data gathered at the time a patient presents for treatment, including imaging and other high-dimensional data, to select the optimal treatment for each patient. We present a functional additive regression model, uniquely constrained to represent the effect of the interaction between a categorical treatment variable and a potentially large number of pretreatment functional covariates on a response variable, while allowing the marginal effects of the covariates to remain unspecified. This method simultaneously selects functional/scalar treatment effect modifiers that exhibit possibly nonlinear interactions with the treatment indicator and that are relevant for making optimal treatment decisions. We will also consider an extension to incorporate matrix-valued pretreatment covariates in addition to functional-valued covariates. The methods will be illustrated with data from a depression clinical trial with electroencephalogram functional data as patients' pretreatment covariates..

INVITED SESSIONS

INVITED SESSION 5 — Machine Learning Algorithms for Survival Analysis

IS5-1 Machine Learning Procedures in Survival Analysis

Malka Gorfine
Tel Aviv University, Israel

The recent popularity of survival neural networks (NNs) is notable. Almost every new advancement in NNs is rapidly adapted for survival analysis. While most existing studies assess performance using the well-known c-index, a high c-index does not necessarily imply a low bias. In other words, the survival estimator derived from the NN might significantly deviate from the actual survival value. Furthermore, an area that remains inadequately addressed is the provision of confidence bands for the survival curve derived from NN analysis. This work contributes in two significant ways: (1) We compare existing survival NN packages, evaluating them in terms of both bias and c-index, demonstrating that Cox-Time often yields the best performance. (2) We introduce various methods for generating confidence bands in any NN-based survival analysis.

IS5-2 How to leverage the success of ChatGPT for longitudinal and time-to-event data

Harald Binder
University Medical Center, Germany

Large language models (LLMs), such as GPT, contain considerable information on biomedical relations. In particular, these models excel in extrapolating sequences of events. I will demonstrate how typical longitudinal and time-to-event data can be transformed to leverage the sequence information in LLMs to improve prediction performance and generate realistic synthetic data.

IS5-3 Hypothesis Testing for the Deep Cox Model

Qixian Zhong
School of Economics, Xiamen University, China

Deep learning has become enormously popular in the analysis of complex data, including event time measurements with censoring. To date, deep survival methods have mainly focused on prediction. Such methods are scarcely used in matters of statistical inference such as hypothesis testing. Due to their black-box nature, deep- learned outcomes lack interpretability which limits their use for decision-making in biomedical applications. This paper provides estimation and inference methods for the nonparametric Cox model – a flexible family of models with a nonparametric link function to avoid model misspecification. Here we assume the nonparametric link function is modeled via a deep neural network. To perform statistical inference, we utilize sample splitting and cross-fitting procedures to get neural network estimators and construct test statistic. These procedures enable us to propose a new significance test to examine the association of certain covariates with event times. We establish convergence rates of the neural network estimators, and show that deep learning can overcome the curse of dimensionality in nonparametric regression by learning to exploit low-dimensional structures underlying the data. In addition, we show that our test statistic converges to a normal distribution under the null hypothesis and establish its consistency, in terms of the Type II error, under the alternative hypothesis. Numerical simulations and a real data application demonstrate the usefulness of the proposed test.

INVITED SESSIONS

INVITED SESSION 6 — Bayesian methods in clinical development

IS6-1 Applying Bayesian methods in clinical trials: opportunities and challenges

Becky Turner
University College London, United Kingdom

This talk will begin with a review of confirmatory trials which have used Bayesian methods in their primary analysis, in papers published over the last 5 years. I will review trial characteristics, approaches taken to design, analysis and decision making, and the use of informative priors. Next, I will present three case studies of clinical trials in which Bayesian methods were used to borrow information.

The ODYSSEY trial evaluated dolutegravir-based antiretroviral therapy regimens for children and adolescents living with HIV. The main trial recruited 707 children weighing 14kg or more, while a smaller group of 85 children weighing <14kg was recruited 12 months later following a lead-in pharmacokinetics study. The treatment effect in younger children needed to be estimated separately, to avoid delaying presentation of results from the main trial, but a standalone analysis of their data would not be adequately powered. Preplanned Bayesian methods were used to enable borrowing of information from the larger subgroup of older children, when estimating the treatment effect in younger children, to increase power and precision. The degree of borrowing was informed by elicited clinical opinion about similarity of treatment effects between the two weight cohorts, obtained before the main trial results were available.

VQUIN and TB-CHAMP were separate randomised placebo-controlled phase 3 trials in adults and children respectively, both evaluating levofloxacin as tuberculosis (TB) preventive treatment in household contacts of individuals with multidrug-resistant TB. While the trials were ongoing, investigators realised that TB event rates were lower than expected and each trial would therefore likely be underpowered. In addition to conventional individual patient data meta-analysis, Bayesian methods were planned, ahead of trial completion, to borrow information between trials and estimate the efficacy in each trial with more precision. To define the weights given to borrowed information, we elicited expert opinion on how efficacy was expected to differ between adults and children, informed by results from relevant natural history studies and meta-analyses of observational studies. Bayesian methods enabled us to estimate treatment efficacy with more precision, by borrowing evidence from a comparable external trial, with evidence-informed recognition of key differences between relevant study populations. This approach was accepted by WHO and informed revised guidelines for prevention of multidrug-resistant TB in children and adults.

INVITED SESSIONS

IS6-2 Bayesian approaches in clinical development: methods and case studies

Nicky Best
GlaxoSmithKline, United Kingdom

There is growing interest in use of Bayesian clinical trials designs with informative prior distributions in settings where it may be advantageous to combine multiple sources of evidence, such as extrapolation of adult data to pediatric populations, or to borrow control data from previous clinical trials to augment a new randomised clinical trial (RCT) in the same population and indication.

When incorporating external data into the analysis of an RCT, a key concern is the potential for bias due to "drift" between the external data and the true response in the new trial. This drift could bias the assessment of control response or treatment effect, leading to misguided decisions (e.g. the approval of ineffective therapies or false non-approval). In addition to careful selection of the external dataset, the use of dynamic Bayesian borrowing methods (e.g. [1]) has also become common practice to help mitigate the risk of bias by dynamically down-weighting the external prior information when the outcome data are inconsistent with the observed outcomes in the new study.

Recently, there have been methodological contributions that seek to combine methods that adjust for imbalances in measured baseline covariates between the external and new trial subjects (e.g. propensity score matching/weighting or regression adjustment) and dynamic borrowing to account for unmeasured confounding not captured by measured baseline characteristics (e.g. [2]).

In descriptions of Bayesian designs it is common to find an evaluation of the associated classical (i.e. frequentist or conditional) type I error. However, it is known that this type I error cannot be strictly controlled and depending on several factors, it can be above, below or equal to its nominal level. Additional alternative operating characteristics have recently been proposed [3] to complement classical type 1 error and power to take an informed decision about the risk of false positive and false negative conclusions when using Bayesian clinical trial designs.

In this talk we illustrate some of these recent methodological developments by describing the use of inverse probability weighted robust mixture priors and their application to an ongoing clinical trial that utilizes a hybrid control arm comprised of internal and dynamically borrowed external controls. We will then describe an expanded set of metrics, including the average type I error [3], that can be used to evaluate the operating characteristics of the proposed design, and allow optimizing the trade-off between potential bias versus efficiency gain of Bayesian designs using external data.

References: *[1] Schmidli et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014;70(4):1023-1032. doi:10.1111/biom.12242. [2] Callegaro et al. Dynamic borrowing of historical controls adjusting for covariates in vaccine efficacy clinical trials. Pharmaceutical Statistics. 2024; doi:10.1002/pst.2384. [3] Best et al. Beyond the classical type I error: Bayesian metrics for Bayesian designs using informative priors. Statistics in Biopharmaceutical Research. 2024; doi.org/10.1080/19466315.2024.2342817.*

INVITED SESSIONS

IS6-3 Regulatory considerations for the acceptability of Bayesian methods

Andrew Thomson
European Medicines Agency, Netherlands

This talk will discuss some of the thinking that goes into the acceptability of Bayesian methods for regulatory decision making. Specific focus will be on paediatric extrapolation, where external data is explicitly leveraged, and other designs such as Bayesian platform designs which may not leverage external information but decision making during and at the end of the trial is conducted within the Bayesian paradigm. The importance attached to Type 1 error control by regulatory agencies is well known, and the transference of this concept into the Bayesian paradigm, whilst ensuring regulatory standards are met, will be discussed.

INVITED SESSIONS

INVITED SESSION 7 — Regulators' view of randomized and non-randomized evidence in drug development

IS7-1 The magic of randomization versus the myth of Real-World Evidence

Richard Peto
University of Oxford, United Kingdom

Non-randomized observational analyses of large electronic patient databases are being promoted as an alternative to randomized clinical trials as a source of "real-world evidence" about the efficacy and safety of treatments. For drugs or procedures that are already being used widely, such observational studies may involve exposure of large numbers of patients. Consequently, they have the potential to detect rare adverse effects that cannot plausibly be attributed to bias, generally because the relative risk is large (e.g. rhabdomyolysis associated with the use of statin therapy). Non-randomized clinical observation may also suffice to detect sudden beneficial effects when good outcomes would not otherwise be expected (eg, control of diabetic ketoacidosis with insulin, or sudden tumour shrinkage).

However, because of the potential biases inherent in observational studies, they cannot generally be trusted when — as is often the case — the effects of the treatment are actually null, or only moderate (ie, less than a twofold difference in the incidence of the health outcome between using and not using the treatment). In those circumstances, large observational studies may yield misleading associations of a treatment with health outcomes that are statistically significant but non-causal, or that are mistakenly null when the treatment really does have clinically important effects. Instead, randomized, controlled trials of adequate size are generally required to ensure that realistically moderate benefits or moderate harms of a treatment are assessed reliably enough to guide patient care appropriately.

The solution to the problems caused by the bureaucratic burdens that have been increasingly imposed on randomized trials during the past 25 years is not to replace randomization with unreliable non-randomized database analyses. Instead, unnecessary obstacles to the reliable assessment of the efficacy and safety of treatments in randomized trials of appropriate size need to be removed.

Terminology: It has unfortunately become common usage for "real-world evidence" to mean non-randomised evidence, and that is the usage in this abstract. The term "real-world data" has been used in recent guidelines merely to mean data not specifically collected for research (eg, from electronic health records); with this usage, both randomised and non-randomised studies use "real-world data".

Reference: *[1] Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. NEJM 2020; 382: 674-8.* *Acknowledgement: This abstract was drafted by R Collins.*

INVITED SESSIONS

IS7-2 FDA and Real-World Evidence

John Concato
FDA, United States

A "randomized trials vs. observational studies" dichotomy oversimplifies what is more appropriately described as a spectrum of study designs ranging from randomized controlled trials (RCTs) to externally controlled trials to observational (non-interventional) studies. At the same time—and despite referring to traditional sources of data and types of study design—the terms real-world data (RWD) and real-world evidence (RWE) have become popular. In this context, the U.S. Food and Drug Administration's (FDA's) Center for Drug Evaluation and Research (CDER) uses its existing evidentiary standard when evaluating randomized or non-randomized studies of drug-outcome associations.

As historical background, FDA has for decades encountered what we now call RWD/RWE (such as data from single-arm trials compared to historical controls). Increased interest in "RWE" is attributable to multiple factors, including improved access to detailed clinical information in the era of big data as well as research showing that non-interventional studies—despite various threats to validity—can generate results that emulate randomized trials. An FDA RWE Program, launched in 2018, has formally integrated assessments of RWD/RWE into day-to-day operations. Corresponding efforts reflect an understanding that evidence from non-randomized studies can serve as an addition to—not a replacement of—RCTs.

Numerous challenges exist when assessing drug-outcome associations using non-randomized comparisons and/or routinely collected clinical data. Nonetheless, trustworthy evidence can be generated when reliable and relevant data are analyzed using a rigorous study design. For example, a Cochrane report (https://doi.org/10.1002/14651858.MR000034.pub2) stated "on average, there is little evidence for significant effect estimate differences between observational studies and RCTs" and "factors other than study design per se need to be considered when exploring reasons for a lack of agreement between results of RCTs and observational studies." An emulation of 32 RCTs using claims data and observational cohort designs (https://jamanetwork.com/journals/jama/fullarticle/2804067) found that although often not achievable, such studies "can reach similar conclusions as RCTs when design and measurements can be closely emulated."

As a notable regulatory example of RWE—based on a non-interventional study comparing data from a well-established registry with data from historical controls—CDER approved tacrolimus (Prograf®) in combination with other immunosuppressants for prevention of organ rejection in patients receiving lung transplants. Three additional issues are relevant: point-of-care RCTs generate RWE when outcomes are identified using RWD; FDA is actively supporting efforts to update clinical trial regulations and promote clinical trial innovation; RCTs will continue to be the main approach to generating evidence for drug approvals.

INVITED SESSIONS

IS7-3 Use of Real-World Evidence in EU regulatory decision making

Andrew Thomson
European Medicines Agency, Netherlands

This talk will discuss some of the thinking that goes into the acceptability of Bayesian methods for regulatory decision making. Specific focus will be on paediatric extrapolation, where external data is explicitly leveraged, and other designs such as Bayesian platform designs which may not leverage external information but decision making during and at the end of the trial is conducted within the Bayesian paradigm. The importance attached to Type 1 error control by regulatory agencies is well known, and the transference of this concept into the Bayesian paradigm, whilst ensuring regulatory standards are met, will be discussed.

IS7-4 The use of Real World Evidence within the context of Open-Label Extensions

Adrian Mander
GSK, United Kingdom

Open label extension (OLE) studies are conducted following a randomised controlled trial to obtain long term safety and efficacy data. A commonly used OLE design is one that offers the experimental treatment to all trial participants after the randomised period of the trial. The lack of randomisation means that it is impossible to obtain a long-term causal estimate of the treatment effect. The OLE design can be improved by adding an external control arm that acts as a comparator. Propensity score weighting or matching are used to help with estimating the causal treatment effect, however, this ignores the overlapping period when the randomised control arm is observed at the same time as the external control arm. The methods presented in this talk will introduce a longitudinal Bayesian Dynamic Borrowing (BDB) approach that helps downweigh the external control arm when in disagreement with the placebo arm. The methodology extends standard BDB approaches to handle multivariate outcomes, both when the variance is known or unknown. Careful choice of multivariate vague distributions and trajectories give good operating characteristics that are evaluated via simulation

INVITED SESSIONS

INVITED SESSION 8 — Combining RWD and randomized clinical trials

IS8-1 Fusing experimental and observational data with multi-task Gaussian processes to estimate heterogeneous treatment effects

Brieuc Lehmann
Stats Dept, UCL, United Kingdom

Bridging the gap between internal and external validity is crucial for heterogeneous treatment effect estimation. Randomised clinical trials (RCTs), favoured for their internal validity due to randomisation, often encounter challenges in generalising findings due to strict eligibility criteria. Observational studies on the other hand, provide external validity advantages through larger sample sizes but suffer from compromised internal validity due to unmeasured confounding. Motivated by these complementary characteristics, we propose a novel Bayesian nonparametric approach leveraging multi-task Gaussian processes to integrate data from both RCTs and observational studies. In particular, we introduce a parameter which quantifies the degree of borrowing and can prevent the observational dataset from dominating the estimation. Our approach outperforms state-of-the-art methods in point predictions across the covariate support of the observational study and provides a measure of uncertainty for the estimated treatment effects. We demonstrate the robust performance of our approach in diverse scenarios through multiple simulation studies and a real-world education randomised trial.

IS8-2 Using observational data for patient characteristics to obtain estimated in target decision population

Nicky Welton
Bristol University, United Kingdom

Heath technology assessment (HTA) compares the effectiveness and cost-effectiveness of treatment options to inform national healthcare decisions. This requires estimates of an objective function (for example net-benefit) that reflects the target population and any specific subgroups that decisions are being made for. Economic models can be complex functions of multiple parameters, and there can be heterogeneity in many of the model inputs.

We show how different ways of summarising heterogeneity reflects different decision questions, and that integrating the objective function over joint covariate distributions is the appropriate summary for national decision-making. We illustrate heterogeneity in non-treatment effect parameters using a HTA of intravenous immunoglobulin (IVIG) for patients with severe sepsis, using data from the Case-Mix Programme national database from the Intensive Care National Audit and Research Centre (ICNARC) to define the target population for the decision.

For heterogeneity in multiple relative treatment effect parameters multi-level network meta-regression can be used to estimate a network meta-regression model when individual patient data is available for at least one RCT. This can be integrated over a population-specific joint covariate distribution to obtain population average estimates in a defined target population. We illustrate with synthetic data from RCTs of treatments for plaque psoriasis to obtain population average estimates in two different target populations with characteristics matching observational datasets.

Observational evidence plays a crucial role in characterising the case-mix of target populations for decision-making. Integrating the objective function, such as net-benefit, over the joint covariate distribution provides the appropriate estimate to inform national healthcare decisions..

INVITED SESSIONS

IS8-3 Evidence synthesis of different types of studies: bias-correction methods, and the hierarchical meta-regression approach

Pablo Verde
University Hospital Düsseldorf, Germany

Introduction: Bayesian nonparametric (BNP) meta-analysis models have been developed to relax assumptions of random effects distributions. These types of models account for possible clustering of the random effects and are attractive to complex settings like real-world evidence synthesis. However, real-world evidence synthesis involves the combination of results from observational studies that can be susceptible to multiple forms of biases, and BNP models do not explicitly address bias corrections.

Methods: In this work, we present the bias-corrected BNP model (BC-BNP). This model is built upon two sub-models: a parametric meta-analysis model of interest, and a model of bias correction. The model of bias follows a Dirichlet-Process that handles nonparametrically studies' internal validity bias. We extend the BC-BNP model by adding a meta-regression model to explain systematic variation in real-world evidence synthesis.

Results: We illustrate the BC-BNP model with two real case studies. The first one assesses potential risk factors for COVID-19 patients, and the second one aims to bridge efficacy and effectiveness in diabetic patients.

Conclusions: We show that the BNP meta-analysis models that ignore bias correction could be misleading when studies of different quality and types are combined. In addition, when bias correction involves non identifiable parameters, subjective information plays an important role in sensitivity analysis of results.

References: *[1] Verde PE. (2021) A bias-corrected meta-analysis model for combining, studies of different types and quality. Biom J. 2021 Feb;63(2):406-422. doi: 10.1002/bimj.201900376. Epub 2020 Sep 30. PMID: 32996196. [2] Verde PE (2023). jarbes: Just a Rather Bayesian Evidence Synthesis_. R package version 2.1.2, <https://CRAN.R-project.org/package=jarbes>. [3] Verde PE, and Rosner GL (2024) A Bayesian Nonparametric Approach and the Bias-Corrected Meta-Analysis Model for Combining Disparate Studies. Biometrical Journal (under review).*

ORAL CONTRIBUTED SESSION 01
Clinical Trials I

OC01-1 SEMAP-curvature: A model-free Bayesian approach for analysis of dose-finding trials

Linxi Han[1], Qiqi Deng[2], Zhangyi He[3], Feng Yu[1]
1 University of Bristol, United Kingdom
2 ModernaTX Inc, Cambridge, United States
3 The George Institute for Global Health, London, United Kingdom

Background/Introduction: The Multiple Comparison Procedure and Modelling (MCPMod) is an efficient statistical method for the analysis of Phase II dose-finding trials, it requires specialised expertise to pre-specify plausible candidate models along with model parameters, which can be problematic given limited knowledge of the agent/compound being studied. Misspecification of candidate models and model parameters can severely degrade its performance.

Methods: To circumvent this challenge, in this work, we develop SEMAP-curvature, a novel Bayesian approach for the detection of the dose-response signal and estimation of the dose-response relationship and target doses. Our method utilises a Bayesian hierarchical framework with the sigmoid Emax model as the default response function for characterising dose-response relationships. By incorporating the total curvature of the dose-response curve as a prior parameter, SEMAP-curvature avoids the requirement of a set of pre-specified candidate models. We construct a test statistic and use simulation to determine the threshold for achieving Proof of Concept. The responses are estimated using maximum a posteriori (MAP). We then estimate the dose-response relationship and target doses of interest using simple interpolation. Results Through extensive simulations, the performance of SEMAP-curvature is evaluated in terms of power to detect dose-response signals as well as estimates of the dose-response curve and target doses of interest. We show that SEMAP-curvature has comparable performance to MCPMod if the true underlying dose-response model is included in the candidate model set of MCPMod. Otherwise, SEMAP-curvature can achieve performance superior to that of MCPMod, especially when the true dose-response model deviates drastically from candidate models in MCPMod. SEMAPcurvature also consistently outperform smoothing spline, especially in complex dose-response relationships.

Conclusion: SEMAP-curvature provides a model-free Bayesian approach for analysis of Phase II dose-finding trials, which not only does not require pre-specification of candidate model curves, but also demonstrates superior performance compared to MCPMod and smoothing spline.

ORAL CONTRIBUTED SESSION 01
Clinical Trials I

OC01-2 Bayesian predictive probability of success for interim monitoring of clinical trials with competing risks data

Chiara Micoli, Alessio Crippa, Martin Eklund, Andrea Discacciati
I-SPY COVID Network Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Introduction: The use of Bayesian predictive probabilities of success (PPoS) for interim monitoring of clinical trials enables the prediction of the trial's success probability if additional data are collected. This allows for potential early cessation of the trial if the PPoS is too low (discontinuation for futility). PPoS in trials with time-to-event (survival) endpoints is complex, especially compared to continuous or binary endpoints, and its applications are scarce. Most importantly, no specific methodology is available for the common situation where the trial outcome is subject to competing events. To fill this gap, we expanded existing methodology to compute the PPoS for time-to-event endpoints in the presence of competing events. We illustrate it using data from two randomized trials, the I-SPY COVID (NCT04488081) and the STHLM3 (ISRCTN84445406) trial, which motivated our work.

Methods: The key steps are: (i) using baseline and outcome data collected up to the interim monitoring, model the joint distribution (T,D) of time-to-event (T) and event causes (D) as a function of cause-specific hazards with Bayesian survival models; (ii) predict new outcome data for future patients by sampling from the joint posterior predictive distribution of (T,D), taking into account the right-censored event time of patients still in the trial as a left-truncation time when predicting their updated outcome; (iii) analyse observed and new predicted data together to determine if the trial will reach a "success", as defined in the protocol. The PPoS is obtained by repeating steps (ii) and (iii) a large number of times and taking the proportion of "successful" simulated trials.

Results: We illustrate the methodology and apply it to two randomized clinical trials featuring time-to-event endpoints affected by competing events. We present different model specifications for the joint distribution of time-to-event and event causes and demonstrate how the choice of the priors can be used to evaluate the PPoS under different assumptions about the treatment effect, as well as how PPoS results can support the DSMB's decision-making process.

Conclusion: Our work addresses a need in the literature by enabling the interim monitoring of clinical trials using PPoS when the outcome is subject to competing events. This approach allows informed decisions on trial continuation or early termination for futility, ensuring patients are not treated with ineffective or inferior therapies and reallocating resources (financial, trial staff) to alternative therapies when needed.

ORAL CONTRIBUTED SESSION 01
Clinical Trials I

OC01-3 A real-time Bayesian re-design of a prostate cancer detection randomised controlled trial

Alexander Ooms[1], Ioana R. Marian[1], Jane Holmes[1,2], Matthew J. Parkes[1,2,3], Alastair D. Lamb[4,5], Richard J. Bryant[4,5]

1 Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom
2 Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom
3 Centre for Biostatistics, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester, United Kingdom
4 Nuffield Department of Surgical Sciences, University of Oxford, United Kingdom
5 Department of Urology, Oxford University Hospitals NHS Foundation Trust, United Kingdom

Introduction: The TRANSLATE randomised controlled trial (RCT) examines two different biopsy techniques, Transrectal (TRUS) and Local Anaesthetic Transperineal (LATP) biopsy, to diagnose clinically significant prostate cancer (csPCa) in men being investigated for this malignancy. The primary outcome measure is detection of csPCa, typically available within seven days of the biopsy procedure. In parallel to the main trial, designed using a conventional framework, the trial was re-designed using an adaptive Bayesian methodology to explore potential savings (of sample size and time) of running the trial using an adaptive Bayesian design.

Methods: The short time to outcome response and large study sample size, 1042, randomised 1:1, lends TRANSLATE to potential adaptation and early stopping. Early stopping for success and futility were incorporated. Success was evaluated via direct posterior probabilities of one-way superiority (LATP superior to TRUS), futility via the posterior predictive probability of success. In keeping with the aims of the main study, the trial design that generated a one-sided Type-I error rate of 2.5% (equivalent to two-sided 5% error rate) and 90% power was used for the Bayesian trial. Trial designs were assessed via simulation, using Beta priors and Binomial outcome data for conjugacy. Priors of Beta(1,1) for both groups were used. The Bayesian design was finalised whilst the trial was ongoing before the primary analysis of the trial. To our knowledge this is the first time an adaptive Bayesian RCT re-design has been performed in 'real-time'.

Results: The Bayesian design yields a 2.29% Type-I error rate and 88.58% Power, with a mean final sample size of 709 across the four reasonable effect size scenarios studied, based on 100,000 simulated trials per effect size. With an early success probability threshold of 0.999, final analysis success probability threshold of 0.976, early futility probability threshold of 0.05 and four planned interims at sample sizes of 520, 640, 760, and 880 (approximately every two months based on expected recruitment rate in the main study). Maintaining both error rates without inflating the final sample size is not possible when introducing interim looks, so the authors elected to preserve Type-I error and allowed a small loss of power.

Conclusion: Based on our simulations, had TRANSLATE been designed as an adaptive Bayesian design, it is anticipated the trial would have been more efficient in terms of time and patients treated than the frequentist design employed. The main trial will be re-analysed under this design when the data becomes available.

ORAL CONTRIBUTED SESSION 01
Clinical Trials I

OC01-4 Evaluating the performance of Bayesian cumulative logistic models in randomised controlled trials: A simulation study

Christopher Selman[1], Katherine Lee[1,2], Robert Mahar[1,3]
1 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, Australia
2 Department of Paediatrics, University of Melbourne, Parkville, Australia
3 Centre for Epidemiology and Biostatistics, University of Melbourne, Parkville, Australia

Background: The cumulative logistic model assuming a proportional odds ratio across categories is the most commonly used method for ordinal in randomised controlled trials. However, departures from proportional odds have the potential to obscure treatment effects, posing challenges in defining a clinically meaningful and appropriate estimand. Although potentially useful extensions to the proportional odds model exist, a comprehensive comparison of their performance when the odds are and are not proportional is limited.

Methods: This study aimed to address this gap by describing and comparing the performance of the logistic, proportional, and unconstrained and constrained partial proportional odds models, within the framework of a two-arm randomised controlled trial. The models were evaluated via simulations and illustrated using empirical data, leveraging insights from the Australasian COVID-19 Trial, a platform trial with multiple secondary ordinal endpoints. Our simulations covered a range of distributional, proportionality, sample size and effect size scenarios across the levels of the ordinal outcome.

Results: Our findings indicate that the proportional odds model outperforms all other models when the data is generated assuming proportional odds, but can introduce substantial bias and poor coverage in the presence of non-proportional odds, particularly at extreme cut-points of the ordinal scale, increasing effect size, sample size, and number of categories in the ordinal outcome. The cumulative odds ratios estimated using the partial proportional odds and separate logistic regression models in the presence of non-proportional odds had negligible bias and good coverage across most scenarios. In particular, the constrained partial proportional odds model had better efficiency and bias, especially at the extreme ends of the ordinal scale, if the cumulative odds ratios were generated as a function of the ordinal scale cutpoints.

Conclusion: The proportional odds model breaks down, to various degrees, when the proportional odds assumption is not satisfied. This can lead to invalid inference of a treatment effect, though this is less of a concern for a smaller number of categories. Alternatives, such as the logistic or partial proportional odds models, offer a useful workaround to provide an unbiased and efficient estimate of the treatment effect for each cut-point of the ordinal scale in most scenarios, but each approach requires different estimand/s that may or may not be relevant to the clinical context.

ORAL CONTRIBUTED SESSION 01
Clinical Trials I

OC01-5 Approximate Bayesian-frequentist power in platform trial design

Robert Mahar[1,2], Julie Simpson[1,3], Katherine Lee[1,2]
1 University of Melbourne, Australia
2 Murdoch Children's Research Institute, Parkville, Australia
3 Oxford University, United Kingdom

Introduction: Evaluating adaptive platform trial designs is challenging because M trials must be simulated for S scenarios for models with many parameters $b!$. A traditional 'piecewise' design approach seeking to control power for S hypothetical scenarios is computationally burdensome, because the number of configurations of $b!$, and hence S, can be infeasibly large. Commonly and often arbitrarily, only a handful of scenarios is selected for piecewise evaluation. A less arbitrary alternative is a 'hybrid' Bayesian-frequentist design approach which draws the S scenarios from a design prior distribution on $b!$. For both the 'piecewise' and 'hybrid' approaches, typically $S << M$ (e.g. $S = 10$, $M = 1000$), thus the computation burden is approximately the same. We propose an 'approximate' hybrid approach where $M <<< S$ (e.g. $M = 1$, $S = 1000$), where power is modelled from the S trials/scenarios using logistic regression, reducing computational burden and allowing power to be estimated for any reasonable configuration of $b!$, without additional computation.

Methods: Using the multifactorial Staphylococcus aureus Network Adaptive Platform Trial as an example, we defined $S = 3$ reference scenarios corresponding to a null, a moderate, and a mixed effect scenario for the 7 treatment effect parameters. Our target/s of estimation was the power for each 7 treatment contrasts (for a trial size of 6000, with 5 interim analyses), within the 3 reference scenarios, estimated using the piecewise approach with $M = 2500$. Our approximate approach simulated $M = 1$ trial for each of $S = 500$, 1000, and 2000 draws of the design prior. Posteriors for simulated trials were computed using Stan. Bias and empirical standard deviation of repetitions of the approximate approach was compared to the piecewise simulation for the 3 reference scenarios. Graphical summaries (with 95% confidence intervals), and predictive distributions, of estimated power over many configurations of $b!$ were provided for a randomly selected instance of the approximate approach.

Results: The approximate approach has relatively small bias and variation at $S = 1000$ (depending on the intended application), and apparent consistency. Graphical representations provided a simple way of assessing sampling and/or parameter uncertainty around the estimates of power, over any reasonable configuration of $b!$.

Conclusion: Our results show that the task of designing a multifactorial adaptive platform trial may be both simplified and expedited, particularly at the design prototyping stage, using an approximate approach.

ORAL CONTRIBUTED SESSION 02
Joint Models I

OC02-1 PLS structural equation modelling and quantile composite-based path modelling for medical and healthcare research

Pasquale Dolce[1], Cristina Davino[2], Domenico Vistocco[3]
1 Department of Translational Medical Science, University of Naples Federico II, Italy
2 Department of Economics and Statistics, University of Naples Federico II, Italy
3 Department of Political Science, University of Naples Federico II, Italy

Introduction: Partial Least Squares Structural Equation Modeling (PLS-SEM) explores complex relationships between composites (linear combinations of observed variables) guided by prior theoretical knowledge. Despite its significant impact in psychometrics, education, environmental, and business research [1], its potential in medical research remains largely untapped. PLS-SEM exploits iterative alternating Ordinary Least Squares (OLS) algorithms and focuses on conditional means of the distributions of the outcome variables. Models focusing on conditional expectations often lack effectiveness, especially in case of skewed response variables, heteroscedasticity, and outliers. For these scenarios, we proposed a Quantile Composite-based Path Modeling (QC-PM) approach [2] to complement PLS-SEM and to explore the entire distribution of outcomes.

Methods: We emphasized the utility of PLS-SEM in health science and medical research through a secondary analysis of a study on health-related quality of life (HRQoL) in patients with neuroendocrine neoplasms (NENs), facilitated by the data being partitioned into blocks, each representing a composite. The application makes it possible to highlight the capabilities of QC-PM to uncover heterogeneity in the relationships between variables and its advantages as a complementary approach to PLS-SEM.

Results: PLS-SEM findings confirmed the significant impact of clinical severity on NENs patients' HRQoL, while clinical heterogeneity showed no direct significant effect. However, resilience significantly mediates these relationships. In contrast, QC-PM reveals the strengthening impact of clinical severity on HRQoL as quality of life decreases and highlights the significant effect of clinical heterogeneity at higher HRQoL percentiles, suggesting different influences of these factors based on HRQoL levels. At lower HRQoL levels, the mediating role of resilience diminishes.

Conclusion: PLS-SEM offers relevant potential for medical and health research. However, as OLS regression, caution is advised against its indiscriminate use to avoid misleading conclusions. QC-PM, by examining heterogeneous effects across outcome distributions, provides a richer data understanding and nuanced insights. We advocate for PLS-SEM and QC-PM's exploration to answer complex research questions, promoting their wider adoption in medical research.

References: *[1] Ciavolino, E., Aria, M., Cheah, J.H., et al. (2022). A tale of PLS Structural equation modelling: Episode I - a Bibliometrix Citation Analysis. Soc. Indic. Res. 164, 1323–1348. [2] Dolce, P., Davino, C. & Vistocco, D. (2022). Quantile composite-based path modeling: algorithms, properties and applications. Adv Data Anal Classif 16, 909–949. [3] Modica, R., Scandurra C., Maldonato N.M., Dolce P., et al. (2022) Health-related quality of life in patients with neuroendocrine neoplasms: a two-wave longitudinal study. J Endocrinol Invest. 45(11):2193-2200.*

ORAL CONTRIBUTED SESSION 02
Joint Models I

OC02-2 Bayesian joint location-scale model for time-to-event and multivariate longitudinal data with association based on within-individual variability

Marco Palma[1], Ruth H. Keogh[2], Siobhán B. Carr[3,4], Rhonda Szczesniak[5,6], David Taylor-Robinson[7], Angela M. Wood[8], Graciela Muniz-Terrera[9,10], Jessica K. Barrett[1]

1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom
3Royal Brompton Hospital, part of Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
4 National Heart and Lung Institute, Imperial College London, United Kingdom
5 Divisions of Biostatistics and Epidemiology and Pulmonary Medicine, Cincinnati Children's Hospital Medical Centre, OH, United States
6 Department of Paediatrics, University of Cincinnati, OH, United States
7 Department of Public Health, Policy and Systems, University of Liverpool, United Kingdom
8 Cardiovascular Epidemiology Unit, University of Cambridge, United Kingdom
9 Ohio University Heritage College of Osteopathic Medicine, Athens OH, United States
10 University of Edinburgh, United Kingdom

Background/Introduction:The quantification of within-individual variability of the measurement around the mean trajectory in longitudinal data is attracting interest in many biomedical applications. As an alternative to using summary statistics (like the standard deviation of the measurements for each individual), which rely on sample size and overlook the longitudinal nature of the observations, a novel approach based on distributional mixed models has been recently proposed [1]. Nevertheless, neither approach accounts for informative drop-out (e.g., due to death). In addition, when within-individual variability is used as a covariate in a separate time-to-event model, the estimates of the hazard ratios suffer from regression dilution bias [2].

Methods: We propose a Bayesian joint model to study the association between a time-to-event outcome and the within-individual variability in multiple longitudinal biomarkers. The longitudinal outcomes are jointly modelled via a multivariate mixed-effects location-scale regression, where random effects capture the correlation between outcomes and their within-individual variability. A proportional hazards model with flexible baseline hazard is proposed to study the association of both the mean and the within-individual variability of the biomarkers with the time-to-event outcome. The model is implemented in Stan.

Results: We apply the model on annual encounter data from the UK Cystic Fibrosis Registry. The aim is to quantify within-individual variability of FEV1 (forced expiratory volume in 1 second, a measure of lung function) and body mass index (BMI) and evaluate their association with mortality in adults with cystic fibrosis, adjusted for common demographic and clinical confounders. Evidence of within-individual variability within sub-populations is found in both FEV1 and BMI measurements. Higher FEV1 within-individual variability is associated with an increase in the risk of death.

Conclusion: Joint models with the association based on within-individual variability provide a more complete overview of the relationship between multivariate longitudinal biomarkers and time-to-event outcomes. The output of these models could inform the design of dynamic risk prediction tools that incorporate within-individual variability as a predictor.

References: *[1] Palma, M., Keogh, R. H., Carr, S. B., Szczesniak, R., Taylor-Robinson, D., Wood, A. M., Muniz-Terrera, G., & Barrett, J. (2023). Factors associated with within-individual variability of lung function for people with cystic fibrosis: a longitudinal registry study. medRxiv, 2023-05. [2] Barrett, J. K., Huille, R., Parker, R., Yano, Y., & Griswold, M. (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. Statistics in medicine, 38(10), 1855-1868.*

ORAL CONTRIBUTED SESSION 02
Joint Models I

OC02-3 Using a Multi-Parameter Estimation of Prevalence (MPEP) model to estimate the prevalence of opioid dependence in Scotland

Andreas Markoulidakis[1,2], Matthew Hickman[1], Andrew McAuley[2,3], Lee R. Barnsdale[2], Nicky J. Welton[1], Megan Glancy[3], Tara Shivaji[2], Craig Collins[2], Jaroslaw Lang[2], Femke de Wit[2], Gordon Hunt[2], Levin Wilkinson[2], Rosalyn Fraser[3], Alan Yeung[3], Kirsten Horburgh[4], Saket Priyadarshi[5], Sharon J. Hutchinson[2,3], Hayley E. Jones[1]
1 University of Bristol, United Kingdom
2 Public Health Scotland, Edinburgh, United Kingdom
3 Glasgow Caledonian University, United Kingdom
4 Scottish Drugs Forum, Glasgow, United Kingdom
5 Alcohol and Drug Recovery Services, NHS Greater Glasgow and Clyde, United Kingdom

Background/Introduction: As the number of drug-related deaths in Scotland increased, and estimates of prevalence using "capture-recapture" approach used in the past are not feasible due to some datasets are no longer available, we deployed a Multi Parameter Estimation of Prevalence (MPEP) modelling approach. The MPEP approach involves administrative record linkage between opioid agonist treatment (OAT) prescriptions and adverse events (e.g., deaths and hospitalisations) data, using a Bayesian statistical model with simultaneous regressions on aggregated data to estimate prevalence.

Methods: The MPEP approach operates under the assumptions: 1) the adverse events modeled (opioid-related deaths and hospitalizations) are specific to the population of interest (i.e. people with opioid dependence), 2) the baseline cohort includes everyone receiving Opioid Agonist Treatment (OAT), and 3) everyone in the baseline cohort is opioid-dependent. Simultaneous Poisson regressions were fitted to the counts of deaths and hospitalizations within the observed cohort, stratified by age group, sex, region, year, and treatment status. A regression structure was fitted to latent prevalence on the log-odds scale, under the key assumption that rates of adverse events among the unobserved part of the population are equal to the rates observed among the baseline cohort during periods not on OAT. The model was extended to account for the fact that part of the target population will die during the year. In this presentation, we will discuss how we approached our application of the MPEP approach to estimate the number of people with opioid dependence in Scotland from 2014 to 2019. Given that our prevalence estimates relied on joint modeling of opioid-related deaths and hospitalisations, we conducted sensitivity analyses by systematically excluding each of these two data sources.

Results: In 2019/20, the estimated number of individuals with opioid dependence in Scotland was 47,100 (95% Credible Interval, CrI, 45,700 to 48,600), with a prevalence of 1.32% (95% CrI 1.28% to 1.37%). In general, prevalence estimates derived from a single data source exhibited minimal divergence from estimates combining both sources, except for males aged 15 to 34 years, where variations ranged from 0.31% to 0.54%. Our findings suggest a slight decrease of -0.07% (95% CrI - 0.14% to 0.00%) in prevalence between 2014/15 and 2019/20.

Conclusions: The flexibility of the MPEP approach allows for adaptation to local settings and circumstances, incorporating a built-in test for the consistency of information regarding the size of the known population and drug-related harm.

ORAL CONTRIBUTED SESSION 02
Joint Models I

OC02-4 Joint modelling of longitudinal and competing-risk data under failure cause and non-informative right censoring misclassification

Christos Thomadakis[1], Loukia Meligkotsidou[2], Constantin Yiannoutsos[3], Giota Touloumi[1]
1 Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Greece
2 Department of Mathematics, National and Kapodistrian University of Athens, Greece
3 Department of Biostatistics and Health Data Science, Indiana University, Indianapolis IN, United States

Introduction: In our motivating example of jointly modelling CD4 count evolution and competing risks (death and disengagement from care) in a resource-constrained setting, death misclassification is concerning. This challenge can be addressed by incorporating data from double sampling, a random sample of disengaged patients with their true vital status actively ascertained. Complicating matters, transfer to other programs (assumed non-informative censoring) is a potential outcome among doubly sampled individuals, with "silent" transfers not reported to the original program. We aim to devise an approach to jointly model longitudinal marker and competing-risk data while accommodating misclassified failure causes (death/disengagement) and censoring indicators ("silent" transfer).

Methods: A linear mixed model (LMM) was assumed for the marker, $y_i(t)=m_i(t)+\varepsilon_i(t)$, where $m_i(t)$ depends on both the fixed/random effects. For the competing risks, we simultaneously modelled the cumulative incidence functions (CIFs) through the transformation1, $F_{ik}(t)=1-[1+c_k\int h_{0k}(s)\exp\{\gamma_k^T w_{ik}+\alpha_k m_i(s)\}]\}ds_{t0}]^{-1/c_k}$, depending on $m_i(t)$ and baseline covariates, with $c_k\to 0$ leading to proportional subdistribution hazards. To address misclassification, we assumed that the true event type, K_i, is available only in doubly sampled individuals. A reported failure cause, $\tilde{K}_i$, is assumed to be always available instead. This necessitates modelling the misclassification probabilities $\pi_{jk}=P(\tilde{K}_i=j|\ K_i=k)$ alongside the joint model. Unlike assuming correctly classified censoring indicators1, we proved that the conditional probability of failure causes for the non-doubly sampled individuals depends on the hazard of censoring on top of the CIFs and π_{jk}'s. Thus, a proportional hazard model was adopted for the censoring distribution. Bayesian inference for the unified approach employed a hybrid MCMC algorithm using data augmentation.

Results: A simulation study, mimicking our motivating example, was conducted, with marker data generated by an LMM and two competing risks based on the proposed model. We also simulated non-informative censoring times (transfers). Misclassification was introduced for the failure causes and censoring indicators, with 20% of relevant individuals included in the double sample. The proposed model, being correctly specified, exhibited negligible biases with nominal coverage rates for all parameters (92.4%-96.4%). However, ignoring the "silent" transfers among doubly sampled individuals resulted in moderately biased association parameters α_k and suboptimal coverage rates (88.6%-90.6%). Moreover, population-averaged CIFs were substantially biased when disregarding transfers. Applying the above-mentioned approaches to real data from the East Africa IeDEA collaboration yielded similar conclusions.

Conclusion: In a joint modelling framework, ignoring a non-informative right censoring process under potentially misclassified event type and censoring indicator can lead to invalid inferences.

Reference: *[1] Thomadakis et al., Biostatistics, https://doi.org/10.1093/biostatistics/kxac043 z*

ORAL CONTRIBUTED SESSION 02
Joint Models I

OC02-5 Bayesian joint modelling of foetal growth measurements and their associations with adverse neonatal outcomes

Rehema Ouko[1], Mavuto Mukaka[2], Eric Ohuma[1]
1 London School of Hygiene and Tropical Medicine, United Kingdom
2 University of Oxford, United Kingdom

Introduction: Foetal and child growth trajectories are essential health and development indicators and may be associated with several maternal and neonatal outcomes. Proper growth monitoring of foetuses and children using appropriate methods is crucial for the early identification and management of at-risk children and mothers. Growth measurements are a multidimensional construct, but most models treat them separately, despite the measurements being highly correlated. Therefore, it would be better to represent these models with models that estimate and incorporate various aspects of correlation between these growth measurements simultaneously.

Methods: In this study, we propose a joint Bayesian model of foetal biometric measurements, i.e., head circumference and femur length, to characterise foetal growth trajectories during pregnancy for prediction of adverse neonatal outcomes such as small-for-gestational age, large for gestational age, preterm birth, and low birth weight. A shared random-effects model is proposed for the foetal biometric measurements to incorporate the correlation structure across the longitudinal measurements. We will use fractional polynomial regression to capture the nonlinear structure in the longitudinal trajectories. The Bayesian posterior computation will be done by Markov chain Monte Carlo sampling. We will compare the predictive ability of the Bayesian joint model to the univariate modelling of growth measurements.

Results: A preliminary analysis modelling the height and weight trajectories of children under 5 years jointly based on the Jenss-Bayley growth function using a Bayesian approzach showed promising results; hence, we plan to explore other flexible models such as fractional polynomials for foetal growth biometry data. Thus, the joint model to be developed will be applied to other longitudinal datasets that are correlated in nature and can be modelled simultaneously. This will enable us to investigate the associations between different growth measures over time.

Conclusion: In conclusion, the joint modelling approach proposed in this study will help in incorporating the correlation between measurements, thus leading to accurate parameter estimates of foetal growth trajectories and enabling early identification of risks that children and mothers might face. Hence, it can be used by health professionals to identify early foetuses or children who are at risk and to provide effective interventions that could improve their health outcomes in the long run.

ORAL CONTRIBUTED SESSION 03
Methods for High Dimensional Data

OC03-1 Personalised protein-protein interaction networks via weighted conditional GGMs

Alessia Mapelli[1,2], Gianmauro Cuccuru[2], Emanuele Di Angelantonio[2,3,4,5,6], Francesca Ieva[1,2]
1 Laboratory for Modelling and Scientific Computing (MOX), Department of Mathematics, Politecnico di Milano, Italy
2 Health Data Science Centre (HDS), Human Technopole, Milan, Italy
3 Blood and Transplant Research Unit in Donor Health and Behaviour, Cambridge, United Kingdom
4 Department of Public Health and Primary Care, University of Cambridge, United Kingdom
5 British Heart Foundation Centre of Research Excellence, University of Cambridge, United Kingdom
6 Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, United Kingdom

Background: / Introduction: Biological networks are frequently inferred using Gaussian graphical models (GGMs) based solely on expression data. GGMs identify conditional dependence by estimating a precision matrix between biological entities. However, conventional GGM approaches applied to protein expression often overlook existing knowledge about protein-protein interactions (PPI) and the influence of covariates or confounding factors on these interactions. Recent advancements, such as the weighted graphical Lasso (wGlasso), have highlighted the benefits of incorporating PPI information (Zhuang et al., 2022). While some approaches have considered incorporating covariates in the estimation with conditional Gaussian graphical models (Augugliaro et al., 2023), jointly addressing both aspects in network estimation must still be explored.

Methods: This study introduces a novel weighted conditional graphical Lasso model to estimate personalized PPI networks. To enhance computational efficiency, we adopt a two-step procedure. First, neighborhood selection employing Meinshausen and Buhlmann's method (Meinshausen & Buhlmann, n.d.) is utilized for dimensionality reduction, considering confounding factors. Subsequently, a conditional Gaussian graphical model is estimated on the selected features, inducing sparsity through lasso penalization and incorporating prior biological knowledge.

Results: The method is applied to the UK Biobank cohort (Sun, 2023), comprising 54,219 participants recruited in the United Kingdom, with measurements of 2,923 unique proteins using antibody-based Olink technology. This personalized representation enables the capture of individualized imprints of current health status and simultaneous representation of behavioral traits such as smoking and the potential risk of diseases such as diabetes. Conclusions: This study introduces a novel personalized representation method that integrates biological knowledge from databases into an estimation-based approach, addressing the impact of confounding factors. The resulting representation can be leveraged for subsequent statistical analyses, ranging from differential graph analysis to disease prediction.

References: *[1] Augugliaro, L., Sottile, G., Wit, E. C., & Vinciotti, V. (2023). cglasso: An R Package for Conditional Graphical Lasso Inference with Censored and Missing Values. Journal of Statistical Software. [2] Meinshausen, N., & Buhlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. The Annals of Statistics, Ann. Statist. [3] Sun, B. B. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. Nature 622. [4] Zhuang, Y., Xing, F., Ghosh, D., Banaei-Kashani, F., Bowler, R. P., & Kechris, K. (2022). An Augmented High-Dimensional Graphical Lasso Method to Incorporate Prior Biological Knowledge for Global Network Learning. Frontiers in Genetics, 12*

ORAL CONTRIBUTED SESSION 03
Methods for High Dimensional Data

OC03-2 Sparse canonical correlation for multiple measurements with latent variable trajectories

Nuria Senar Villadeamigo, Mark van de Wiel, Koos Zwinderman, Michel Hof
Amsterdam UMC, Netherlands

Introduction: Canonical Correlation Analysis (CCA) is a popular multivariate method in omics research for high-dimensional data integration. CCA is effective in discovering underlying mechanisms via linear projections of the observed features (canonical variables) which maximally correlate these datasets. Standard CCA assumes observations are independent, and thus it cannot properly handle repeated measurements or longitudinal data. Moreover, this data may be sparsely and irregularly observed. Current CCA extensions dealing with repeated measurements use one of two approaches, (1) perform CCA on summarised data (i.e.: estimated random effects), or (2) estimate canonical vectors for each measurement. While these techniques indeed factor in the correlation across measurements, they are sub-optimal in high-dimensional analysis and in exploiting the temporal aspects of the data.

Methods and Results: We propose a new extension of CCA where we add random effects at the latent variable level to both correct for the correlation of the repeated measurements and draw the longitudinal latent trajectory over the measurements. Canonical weights are fixed over all measurements and can be regularised. That is, we focus on changes in correlation structures for each component. We implement an $\ell-1$ penalty which fixes sparsity levels to aid interpretability and computational efficiency. We estimate the longitudinal paths using random effects in the latent variables, which are low-dimensional. As a result, we avoid estimating many random effects from the original data. Through this design, we leverage the clustered information of the high-dimensional datasets and study their correlated dynamics. The sparse canonical weights, yield interpretable outcomes on variable contribution to the estimated correlations. Additionally, modelling time in the latent dimensions reduces the computational burden. We use simulated data to validate our model's performance and use data from the Human Microbiome Project to illustrate its real world applications on high-dimensional and sparsely and irregularly observed data.

Conclusion: This approach leads to efficient estimation of the canonical correlations over measurements for clustered data. Compared to the currently existing methods, we ease computational times in high-dimensional analysis. Last, the sparse canonical weights as well as the longitudinal trajectories provide interpretable results.

ORAL CONTRIBUTED SESSION 03
Methods for High Dimensional Data

OC03-3 Biomarker discovery for cancer drug-sensitivity screens via multi-output Gaussian Processes

Leiv Rønneberg
University of Cambridge, United Kingdom

Background: High-throughput drug sensitivity screens enable rapid in-vitro testing of compounds on cancer cell lines, in order to determine the efficacy of a certain treatment. Coupled with various "omics" characterisations of the cancer cell lines, these experiments provide the ingredients necessary to discover biomarkers of treatment effect. However, in-vitro cell viability measurements are frequently corrupted by measurement error due technical error sources and natural biological variability, making the precise quantification of treatment efficacy difficult. Furthermore, the process of biomarker discovery is usually completely disentangled from the noisy reality of the raw viability measurements, utilising crude summary measures of treatment efficacy, estimated with no uncertainty quantification.

Methods: We propose a model that jointly estimate cell line-specific dose-response functions and at the same time performs biomarker discovery. The model is based on a multioutput Gaussian Process (MOGP) formulation, that allows the estimation of non-linear dose-response relationships, while at the same time incorporating high-dimensional "omics" measurements for biomarker discovery. Selection of potential biomarkers is achieved by placing a horseshoe prior within the MOGP kernel construction.

Results: We show on simulated and real data that the model can capture non-linear dose-response relationships and correctly identify known biomarkers of drug efficacy. We further establish a link between our MOGP model and a version of function-on-scalar (FoS) regression, suggesting further applications of the model in e.g. longitudinal data analysis. Viewing FoS-regression as a MOGP further facilitates using all the machinery developed to scale GPs to large datasets, potentially speeding up inference in these models.

Conclusion: We propose a new model for biomarker detection in high-throughput drug sensitivity screens that combines non-linear regression using Gaussian Processes, and variable selection with sparse priors. Fitting dose-response curves and detecting biomarkers in a joint model circumvents the need for noisy two-stage approaches based on crude summary measures of treatment efficacy, and the Bayesian framework offers principled uncertainty quantification.

OC03-4 Distance correlation methods for genome-wide association studies

Dominic Edelmann[1], Fernando Castro-Prado[2], Jelle Goeman[3]
1 German Cancer Research Centre, Heidelberg, Germany
2 University of Santiago de Compostela, Spain
3 Leiden UMC, Netherlands

Background: Distance Correlation is a powerful concept for measuring dependence across various data types. Recently the concept of distance covariance has been unified with the Hilbert-Schmidt-Independence Criterion and the Global Test framework.

Methods: Leveraging these insights, we develop distance covariance methods specifically tailored for genome-wide association studies (GWAS). We propose a family of tests to examine the association of single nucleotide polymorphisms (SNPs) with quantitative responses. Using the equivalence with the Global Test framework, we establish that specific versions of distance covariance equate to locally most powerful tests under certain statistical models. This result outlines the scenarios in which these tests are particularly effective. Closed form expressions for the asymptotic and the finite-sample distributions of the test statistics are derived.

Results: Simulation studies confirm the good theoretical power properties of the proposed methods outperforming the conventional linear model in a wide range of scenarios.

Conclusions: Considering the robust theoretical underpinnings and the excellent performance in simulation studies, our proposed methods merit recognition as a standard analytical tool in the association testing of single SNPs with quantitative responses.

ORAL CONTRIBUTED SESSION 03
Methods for High Dimensional Data

OC03-5 Comparison of the LASSO and Integrative LASSO with Penalty Factors (IPF-LASSO) methods for multi-omics data: Variable selection with error control

Charlotte Castel[1], **Zhi Zhao**[1,2], **Magne Thoresen**[1]
1 Centre for Biostatistics and Epidemiology (OCBE), Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Norway
2 Institute for Cancer Research, Department of Cancer Genetics, Oslo University Hospital, Norway

Variable selection in relation to regression modeling has constituted a methodological problem for more than 60 years. Especially in the context of high-dimensional regression, developing stable and reliable methods, algorithms, and computational tools for variable selection has become an important research topic. Omics data is one source of such high-dimensional data, characterized by diverse genomic layers, and an additional analytical challenge is how to integrate these layers into various types of analyses. While the IPF-LASSO1 model has previously explored the integration of multiple omics modalities for feature selection and prediction by introducing distinct penalty parameters for each modality, the challenge of incorporating heterogeneous data layers into variable selection with Type I error control (false positives) remains an open problem. To address this problem, stability selection was used as a method for variable selection with false positives control. Both regular LASSO and IPF-LASSO were integrated into the stability selection framework. The objective of this study was to compare the LASSO algorithm with the IPF-LASSO, investigating whether introducing different penalty parameters per omics modality could improve statistical power while controlling false positives. Two high-dimensional data structures were investigated, one with independent data and the other with correlated data. Three cut-off values were tested for stability selection in both regular LASSO and IPF-LASSO. In the simulation studies, both methods control the false positives, but the IPFLASSO increases the power compared to standard LASSO, particularly in scenarios with a small true model size, a large absolute difference between proportions of relevant variables in the two modalities, and a small ratio between smaller modality size and larger modality size. The different models were also illustrated using data from a study on breast cancer treatment, where the IPFLASSO model was able to select some highly relevant clinical variables. To our knowledge, this work is the first to treat the question of integrating several different correlated omics data modalities into the same regression model while at the same time controlling the number of false positive selections. The IPF-LASSO integrated into the stability selection framework appears advantageous for variable selection with false positive control while increasing the power of the model.

References: *[1] Anne-Laure Boulesteix, Riccardo De Bin, Xiaoyu Jiang, Mathias Fuchs, et al. Ipf-lasso: integrative-penalized regression with penalty factors for prediction based on multi-omics data. Computational and mathematical methods in medicine, 2017, 2017*

ORAL CONTRIBUTED SESSION 04
Statistical Analysis for Complex Data Structures

OC04-1 Unveiling the impact of social and environmental determinants of health on lung function decline in cystic fibrosis through data integration using the US Registry

Eleni-Rosalina Andrinopoulou[1,2], **Emrah Gecili**[3,4], **Rhonda Szczesniak**[3,4,5]
1 Department of Biostatistics, Erasmus MC, Rotterdam, Netherlands
2 Department of Epidemiology, Erasmus MC, Rotterdam, Netherlands
3 Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Centre, OH, United States
4 Department of Paediatrics, University of Cincinnati Children's Hospital Medical Centre, OH, United States
5 Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Centre, OH, United States

Background: Integrating different data resources is a challenging task, but it provides a more comprehensive view of the patient information that could improve healthcare decision-making compared to relying solely on electronic health records. For example, in Cystic Fibrosis (CF), which is a genetic disorder mainly affecting the lungs, biomarkers that track lung function decline serve as important predictors for assessing disease progression and predicting hospitalization, transplantation, and mortality outcomes. It has been shown that including data sources incorporating location-specific social and environmental determinants significantly improves the accuracy of disease progression prognostication, particularly in characterizing lung function decline.

Methods: Even though people with CF lung disease represent heterogeneous socioeconomic groups from different geographic regions, lung function decline could differ in these groups. To explore these disparities in the progression of lung disease, we integrate patient registry data from the US Cystic Fibrosis Foundation with information on social and environmental health. In particular, we focus on the relation between lung function, measured as forced expiratory volume in 1 s of % predicted (FEV1) and the community deprivation index, a marker derived from six variables of the American Community Survey. Given that both outcomes are time-dependent, our methodology is built upon an extension of multivariate mixed-effects models. This approach is designed to model multiple longitudinal outcomes, incorporating varied functional forms to establish their connections. We use the area under the deprivation index curve specified at different periods in the submodel of FEV1. Examining various periods would enable us to investigate whether this relationship differs based on the duration of patients' exposure to areas with a high deprivation index. Additionally, considering the geographical variations in both lung function decline and community deprivation, we explore this relationship within each state of the US.

Results: A strong association is observed between FEV1 and the area under the deprivation index curve across all states. The strength of this association diminishes when considering a two-year time window preceding the FEV1 measurement, as opposed to the patient's entire medical history. Furthermore, we conducted a sensitivity analysis to explore different ways of linking these outcomes.

Conclusion: Incorporating environmental and socioeconomic markers in clinical decision-making strategies is expected to provide more insights into the progression of the disease.

ORAL CONTRIBUTED SESSION 04
Statistical Analysis for Complex Data Structures

OC04-2 Automating functional data analysis in real time: An application to pressure sensor data in the treatment of venous leg ulcers

Nastaran Sharifian[1], Darren Burke[2], Emer Garry[2], Andrew Cameron[2], Andrew Simpkin[1]
1 University of Galway, Ireland
2 FeelTect Limited, Galway, Ireland

Background: Venous Leg Ulcers (VLUs) afflict approximately 11.5 million individuals globally, stemming from impaired leg vein function. Compression therapy is the conventional treatment for VLUs, but ensuring appropriate pressure remains a challenge. The novel medical device, Tight Alright developed by FeelTect, can measure bandage pressure 20 times per second at three relevant physiological points on the leg, ensuring effective venous return along the length of the leg and promoting VLU healing. The Tight Alright connected health platform provides measurement of sub-bandage pressure when applying compression therapy, ensuring safe and efficacious target pressure is consistently delivered. These pressure data are transmitted wirelessly (via Bluetooth) to a user-friendly Mobile App, leading to multivariate real-time sensor data. Pressure data are available from the same individual using the pressure sensing device and app under two pressure settings while performing three exercises for 60 seconds each.

Methods: Leveraging Functional Data Analysis (FDA), we smoothed these data to derive a finite-dimensional representation for model fitting, facilitating Multivariate Functional Principal Component Analysis (MFPCA). MFPCA, akin to traditional Principal Component Analysis (PCA) but tailored for multivariate functional data, decomposes variability into orthogonal functional principal components (FPCs), enabling dimension reduction and visualisation. Each pressure wave is attributed a multivariate score for each component of variation, and these scores are then modelled using longitudinal data analysis to investigate pressure change over time and compare between exercise states.

Results: We develop an algorithm to automate pressure wave detection, segmentation, registration, and MFPCA, culminating in multivariate principal component scores. These scores compute variation across the three sensor positions for each wave. We use longitudinal data analysis to model these scores over time to find differences between exercise and starting pressure.

Conclusions: This approach accommodates the influx of pressure sensor data, ensuring adaptability to new datasets and facilitating real-time analysis for individuals utilising the monitoring device so as to harness and harmonise these multimodal data into useful user facing visualisations on the app. This will ultimately advance monitoring and treatment decision making in VLU wound

treatment, which will advance the use of connected-health technology with precision compression therapy to expedite wound healing.

ORAL CONTRIBUTED SESSION 04
Statistical Analysis for Complex Data Structures

OC04-3 Machine learning algorithm to predict biomarker levels using metabolomics data

Fabiola Del Greco Miglianico[1], Daniele Giardiello[2], Chiara Volani[1,3], Giuseppe Paglia[2]
1 Institute for Biomedicine (affiliated to the University of Lübeck, Germany), Eurac Research, Bolzano/Bozen, Italy
2 School of Medicine and Surgery, University of Milano-Bicocca, Italy
3 Department of Biosciences, Università degli Studi di Milano, Italy

Introduction: Metabolomics aims to measure small molecules (called metabolites) in cells, tissues, biofluids, that represent intermediates and/or end-products of biochemical/cellular processes. As such, metabolomics has shown to be useful for making prediction about disease risks. Given the large complexity and size of the data, the Machine learning (ML) approach represents an appropriate statistical and computational tool for building predictive models.

By a data-driven investigation, the aim of this work is to provide a ML workflow to predict biomarker levels by using targeted metabolomics data from a general population. Specifically, we identify and discuss specific ML issues that could influence the results.

Method: We explore a supervised ML algorithm based on the feature selection using hypothesis testing, paying attention on the presence of: confounding; non-independent observations (typical of family- or pedigree-based studies); high dimensionality of the dataset (feature space). The impact of those issues on the results is investigated through a dataset of 154 serum targeted metabolites determined by liquid chromatography (LC)–electrospray ionization–tandem MS and flow injection electrospray ionization–tandem MS profiling (AbsoluteIDQ p180 kit, Biocrates Life

Sciences AG), and serum ferritin levels, as an example of biomarker, from a subsample of the Cooperative Health Research In South Tyrol (CHRIS) study with around 5,000 participants.

Results: Inclusion of confounders in the ML algorithm leads to more accurate predictions, in terms of lower loss function. None adjustment could lead to false positive associations. For the feature selection, the Variance Inflation Factor selection with Lasso regression is pretty stable for any confounding approach. The presence of non-independent observations is not influencing the algorithm performance.

Conclusions: This workflow should provide a framework for greater integration of ML within metabolomics studies, taking into account confounding, high large and complex dataset, and presence of family related data. Family-based study has been simulated mimicking the real-data scenario.

ORAL CONTRIBUTED SESSION 04
Statistical Analysis for Complex Data Structures

OC04-4 Generalised partial credit model with covariates: Comparing the efficacy of two implementations in a Bayesian framework

Cristina Ehrmann, Carla Sabariego
Swiss Paraplegic Research, Nottwil, Switzerland

Introduction: The Generalized Partial Credit Model (GPCM), used in many applications in health and medical sciences, is a polytomous Item Response Theory (IRT) model that estimates the position of individuals and the items along a unidimensional latent continuum. Previous research has shown that the implementation of GPCM in a full Bayesian framework, computed using Markov Chain Monte Carla (MCMC) methods in the STAN software package allows the inclusion of covariate effects as an additional effect allowing simultaneous estimation of regression and item parameters. The GPCM with covariates in a Bayesian framework responds to the need to explain the effect of covariates when hierarchical health data are available. Although this model has shown a good prediction of the latent trait score for the new individual, it does not account for the effect of the covariates on the item parameters, i.e. the differential item functioning, an important IRT assumption. This project implements the hierarchical GPCM model in the STAN software, where the item parameters vary across groups defined by the covariates and the abilities of the individuals are assumed to be random across the groups. In addition, the fit of the model and the predictive power for new individuals will be tested and compared with the GPCM with additional covariate effect.

Methods: We used two simulated studies specifically designed to validate the accuracy of parameter estimation for the implementation of the hierarchical GPCM in the STAN software. We then compare the predictive performance of this model with the GPCM with additional covariate effect through an analysis of real health data from the International Spinal Cord Injury (InSCI) Survey in terms of confusion matrix, classification error, and perplexity.

Results: The simulated data consisted of 10 polytomous items and 500 individuals for each covariate group. Symmetric item thresholds were considered for one simulated data and asymmetric item thresholds for the other. Most of the discrepancies between the estimated and generated parameters were about zero, indicating that the STAN implementation of the hierarchical GPCM successfully recovers the true parameters. The hierarchical GPCM on the has been preferred by the Watanabe information criteria and the prediction indicators for InSCI data.

Conclusion: The hierarchical GPCM model can be used to compare the abilities of individuals when large hierarchical health data are available and all IRT assumptions are to be considered. However, it takes longer to compute convergence than the GPCM with additional covariate effects.

ORAL CONTRIBUTED SESSION 04
Statistical Analysis for Complex Data Structures

OC04-5 Domain adaptation approaches for harmonising multi-site tabular data

Patric Tippmann[1,2], **Harald Binder**[1,2]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Centre, University of Freiburg, Germany
2 Freiburg Centre for Data Analysis and Modelling, University of Freiburg, Germany

Introduction: Increasing sample size by modeling clinical data across different sites often involves the challenge of different data distributions, sometimes also described as different domains in the methods literature. Domain discrepancies are typically mitigated by manually harmonizing datasets. Yet, manual harmonization is a laborious process. Therefore, an automated approach would be attractive for harmonization. As candidates, so-called domain adaptation approaches have been developed in computer science. There, variables characterizing patients across domains are learned, either through explicit alignment methods or by learning domain-invariant representations, e.g. in domain adaptation approaches using Variational Autoencoders (VAEs). Even so, tabular medical data, where patients are often characterized by a combination of continuous and categorical variables, can be challenging for neural network approaches, including VAEs. While there has been a recent focus on adapting neural network approaches for tabular prediction tasks, the corresponding insights still have to be fully transferred to representation learning. In addition, a challenge is posed by the need to harmonize potentially multiple sites, where domain adaptation approaches so far have predominantly focused on a single source and a single target domain.

Methods: We specifically focus on the use of Variational Autoencoders (VAEs) for harmonizing tabular medical data. There, a shared representation for different domains is created by learning a low-dimensional latent space. We account for domain-specific, domain-invariant, and residual components by incorporating corresponding structure into the representation. Furthermore, we evaluate different options for a harmonization approach in a comprehensive simulation study. As an extension, we consider the use of an existing reference dataset to inform the representation.

Results: We find that an adequate representation of tabular medical data with a mixture of continuous and categorical variables (such as diagnoses, procedures, and medications) can be obtained. The use of these representations for harmonization is seen to depend on an adequate choice of tuning parameters to balance representations of variables of different type. This particularly affects the choice of metrics for improving the overlap between datasets from different sites.

Conclusion: As seen from these results, VAEs present a promising approach for harmonizing tabular medical data via domain adaptation. This enables leveraging data from different sites to enhance the generalization of predictive models in clinical settings, which could ultimately contribute to improved patient care and healthcare outcomes.

ORAL CONTRIBUTED SESSION 05
Clinical Trials II

OC05-1 Flexible borrowing of historical information in a basket trial via the exchangeability-nonexchangeability (EXNEX) model

Libby Daniells[1], Pavel Mozgunov[2], Helen Barnett[1], Alun Bedding[3], Thomas Jaki[2,4]
1 Lancaster University, Lancaster, United Kingdom
2 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
3 Roche Products Ltd, Welwyn Garden City, United Kingdom
4 Faculty of Informatics and Data Science, University of Regensburg, Regensbury, Germany

Background: With the advances in personalized medicine, two-stage adaptive enrichment designs for clinical trials assessing targeted therapies have received increased attention. In this design, the subpopulation that is likely to benefit from the treatment is identified based on the data from the first stage. In the second stage, the recruitment is restricted to the selected subpopulation. Data from both stages are then used to perform the confirmatory analysis for the treatment effect in the selected population. This "select and test" structure of adaptive enrichment designs leads to statistical challenges in hypothesis testing, point estimation and confidence interval calculation. Several methods concerning hypothesis testing and point estimation have been proposed. However, the question of constructing confidence intervals (CIs) has drawn less attention. In this work, we develop CIs that adjust for trial adaptation.

Method: We consider the adaptive enrichment design proposed by Jenkins et al. [1], which is realistic in practice, and the case when the outcome is normally distributed. Our approach combines the uniformly minimum variance conditional unbiased estimator (UMVCUE) and bootstrap sampling procedures for diagnostic tests proposed by Pepe et al. [2], which simulate the adaptivity of the trial, to construct bias-adjusted confidence intervals for the treatment effect in the selected population.

Result: We demonstrate that confidence intervals generated by our bootstrap algorithm achieve coverage probability very close to nominal level and are approximately symmetric with probability above and below true effect approximately equal. Using these metrics, it outperforms existing methods, including double bootstrap confidence intervals and duality confidence intervals.

Conclusion: The new interval estimation method shows a superior capability to correct selection bias and achieve nominal coverage probability over existing methods. It can be extended to other adaptive designs with binary or time-to-event endpoints when the distribution of the effect size estimate is asymptotically normal.

Refereneces: *[1] Jenkins, M., Stone, A. & Jennison, C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical statistics 10, 347–356 (2011) [2] Pepe, M. S., Feng, Z., Longton, G. & Koopmeiners, J. Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. Statistics in Medicine 28, 762–779 (2009).*

ORAL CONTRIBUTED SESSION 05
Clinical Trials II

OC05-2 Dual criteria approach to early conditional approval in time-to-event group sequential trials via historical borrowing

Marco Ratta[1], Gaëlle Saint-Hilary[3], Donia Skanji[2], Valentine Barboux[2], Mauro Gasparini[1], Pavel Mozgunov[4]
1 Politecnico Di Torino, Torino, Italy
2 International Research Institute Servier (IRIS), Paris, France
3 Saryga, France
4 Cambridge University, MRC Biostatistics Unit, Cambridge, United Kingdom

Background: The urgency of delivering novel effective treatments against life threatening diseases to patients has brought various health authorities to allow for accelerated approvals. Accelerated approval is the "fast track" program where promising treatments are evaluated based on surrogate (short term) endpoints likely to predict clinical benefit to get a conditional approval subject to providing further evidence of efficacy, for example, on the primary (long term) endpoint. Despite this procedure being quite consolidated, a number of conditionally approved treatments do not obtain full approval. This implies a need of improvement of the criteria (and their quantification) to control the risk of conditional approvals for non-effective treatments and contextually maximize the chance of conditional approvals for effective ones.

Methods: In this work, we first propose a novel adaptive design that includes a dual criteria "early conditional approval" analysis, where efficacy on a surrogate endpoint is tested jointly with a predictive criterion based on the primary endpoint. Secondarily we explore how the latter may be impacted by historical information borrowing, in particular: (i) the estimated historical relationship between the surrogate and the primary endpoints, and (ii) historical control data on the primary endpoint. Results We propose various metrics characterizing the risk of correct and incorrect early conditional approvals and demonstrate how the proposed design allows to explicitly control them, with particular attention to the FWER (family-wise error rate). The methodology is then evaluated through a simulation study motivated by a Phase III trial in metastatic colorectal cancer (mCRC).

Conclusion: In this work we demonstrate that - in the context of trials designed to include a conditional approval request - adding a predictive criterion is useful in controlling the family-wise error rate (FWER). Moreover, we show that borrowing historical information may be beneficial in improving the operating characteristics of the study if there is fair accordance between concurrent and historical data.

References: *[1] Roychoudhury S, Neuenschwander B. Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. Statistics in Medicine 2020 3; 39:984–995. [2] Saint-Hilary G, Barboux V, Pannaux M, Gasparini M, Robert V, Mastrantonio G. Predictive probability of success using surrogate endpoints. Statistics in Medicine 2019 5; 38:1753– 1774..*

ORAL CONTRIBUTED SESSION 05
Clinical Trials II

OC05-3 Multimodal outcomes in N-of-1 trials: A simulation study

Juliana Schneider[1], Thomas Gärtner[1], Stefan Konigorski[1,2]
1 Hasso Plattner Institute, Potsdam, Germany
2 Hasso Plattner Institute for Digital Health at Mount Sinai, New York City, United States

N-of-1 trials are (possibly randomized) multi-crossover trials in individuals aimed at investigating presumable treatment effects. Participants alternate between periods of two or more interventions. This trial design is especially useful for rare or chronic diseases and personalized analyses. Multimodal N-of-1 trials are a variant that use multimedia such as audio or image data directly to measure health outcomes of interest. Here, we present an N-of-1 trial simulation study examining the effect of facial treatments on acne assessed on synthetically generated images of faces, in order to evaluate an unsupervised deep learning approach for analysing multimodal N-of-1 trials. Simulated images were created by inserting a number of white spots randomly into face images according to an underlying simulated acne severity. We simulated two sets of scenarios on 1000 individuals each, (i) under the null hypothesis (H0) where there was no treatment effect on acne severity i.e. there were on average equally many white spots across treatment phases, and (ii) scenarios with an immediate treatment effect on the skin (H1). Additionally, we varied the size of the inserted spots in order to investigate our model's ability to identify also minor shifts in the skin's appearance. We applied several statistical tests to evaluate for power and type I error on the first principal component from the simulated image embeddings of an autoencoder as a proxy for acne severity (see [1]). The results showed the capability of our model to detect a treatment effect when there was one, while controlling the type I error rates. More specifically, there was a moderate to high power for H1 and a type I error rate around the alpha level of 0.05 for varying spot sizes. Subsequent hypothesis tests validated that the model caught large as well as small changes with high power in simulations with decreasing spots magnitudes. This research presents an innovative design of multimodal simulation studies in order to evaluate methods for multimodal N-of-1 trials. In future simulations, we aim for a more realistic image generation of acne severity such as from diffusion models. Further improvements on our approach may lead to more robust results on empirical data. Moreover, we strive to improve the statistical efficiency and validity of our analyses by incorporating the multivariate nature of multimodal outcomes.

References: *[1] Juliana Schneider, Thomas Gärtner, and Stefan Konigorski. Multimodal outcomes in n-of-1 trials: Combining unsupervised learning and statistical inference, 2023.*

ORAL CONTRIBUTED SESSION 05
Clinical Trials II

OC05-4 Improving interim decisions for single arm trials by adjusting for baseline covariates and short-term endpoints

Eline Anslot, Kelly Van Lancker
Ghent University, Ghent, Belgium

Background/Introduction: Researchers must endeavour to minimize exposure to harmful and/or ineffective treatments and maximize exposure to effective ones. One approach to achieve this is using multi-stage designs that allow stopping an ongoing trial for futility and/or efficacy. Commonly used designs for single arm trials include group sequential designs with error spending functions, and the Simon two-stage design for binary outcomes which only allows stopping for futility. In both designs, when continuously recruiting, the interim analysis might be inefficient as they usually ignore information in baseline covariates and/or short-term endpoints.

Methods: Since participants enter the study at different times, some may not reach their (long-term) primary endpoint during the interim analysis, but they will possess other pertinent information (e.g., baseline covariates and short-term endpoints). Handling this as a missing data problem, prediction models are constructed to forecast the long-term endpoint of these subjects based on the available information. The interim estimator has the appealing property of being asymptotically unbiased, even when the prediction models are misspecified. When the prediction models are correct, the interim estimator is asymptotically efficient. A simulation study and data analysis were conducted to assess the operating characteristics of the proposed method.

Results: Simulation results indicate that the Type I error and the probability to correctly stop the trial for futility and/or efficacy were controlled for, even when the prediction models were misspecified. Additionally, the results show a gain in power and a decrease in the chance of incorrectly stopping the trial for futility and/or efficacy. Specifically, benefits are obtained with a) more partial information available relative to the total number of participants at interim, b) earlier adjusted variable measurement, and c) higher predictivity of the adjusted variable(s).

Conclusion: The proposed method outperformed the standard methods. Important assets are the feature that it uses all pertinent information available at interim, its robustness against model misspecification, and its capability to be easily expanded (e.g., to multiple short-term endpoints).

ORAL CONTRIBUTED SESSION 05
Clinical Trials II

OC05-5 Adjusting for time-dependencies in N-of-1 trials: A comparative analysis ofstatistical methods

Thomas Gaertner[1], Juliana Schneider[1], Stefan Konigorski[1,2]
1 Hasso Plattner Institute, Potsdam, Germany
2 Hasso Plattner Institute for Digital Health at Mount Sinai, New York City, United States

N-of-1 trials present a high individualization of clinical studies, especially helpful for personalized medicine. They are multi-crossover controlled trials, where each patient is their own control group. As the patients are exposed to multiple treatments over time, the trial data may be considered as time series data. Previous research has shown that adjusting statistical methods for time-dependencies, especially for carry-over, is challenging. Nevertheless, it is suggested that the statistical analysis can be improved by appropriate adjustment for time dependencies. Building upon our previous work on time adjustments for carry-over [1] , this study ventures further to investigate the impact of diverse time dependencies on the biases and interpretations of statistical estimates in simulated N-of- 1 trials. This includes novel investigations into Estimands when carry-over or stochastic processes are present. In a series of simulated N-of-1 studies, we compare different statistical models with adjustments for time dependencies, identify potential biases of certain estimates, and discuss model interpretations. To tackle this complex landscape, we simulate a large array of realistic scenarios featuring varied carry-over structures, time dependencies such as AR-1, and seasonality through weekdays as an effect on the outcome. A comprehensive suite of parametric statistical tests, including linear and generalized linear models (with mixed and autoregressive variants), is employed to assess the effectiveness and biases of different models, both with and without sophisticated time adjustments. The results indicate that our simulated framework facilitates a comprehensive evaluation of statistical methods, uncovering insights into Type-I and Type-II errors, effect size biases, and the suitability of models for specific scenarios. Some time dependencies can significantly influence model performance, potentially leading to biases. This experimental approach enabled us to delineate the strengths and weaknesses of various statistical methods in a controlled setting. In summary, our analysis shows that time-dependent data in N-of-1 trials can lead to biases of treatment effect estimates when the models are misspecified with respect to the time dependencies. Furthermore, statistical methods can increase their power by adding adjustments for the time-dependencies. With the systematic comparison of models in scenarios with complex time-dependencies, this research contributes to the ongoing improvement of statistical methodologies for N-of-1 trials, offering a pathway towards more personalized and precise medical interventions.

References: *[1] Thomas Gärtner, Juliana Schneider, Bert Arnrich, and Stefan Konigorski. Comparison of bayesian networks, g-estimation and linear models to estimate causal treatment effects in aggregated n-of-1 trials with carry-over effects. BMC Medical Research Methodology, 23(1), August 2023.*

ORAL CONTRIBUTED SESSION 06
Meta-analysis

OC06-1 multinma: A comprehensive R package for network meta-analysis of survival outcomes with aggregate data, individual patient data, or a mixture of both

David Phillippo[1], Ayman Sadek[1], Hugo Pedder[1], Sofia Dias[2,1], Tony Ades[1], Nicky Welton[1]
1 University of Bristol, Bristol, United Kingdom
2 University of York, York, United Kingdom

Introduction: Survival or time-to-event outcomes are commonplace in disease areas such as oncology. Healthcare decision makers require estimates of relative efficacy between different treatment options, however treatments of interest are frequently not all compared in head-to-head randomised controlled trials, and so indirect comparison and network meta-analysis (NMA) methods are required to synthesise evidence from a connected network of trials and treatments. An extension of NMA, multilevel network meta-regression (ML-NMR), is increasingly used to account for differences in effect modifiers between populations where individual patient data are available from one or more trials. However, to date there has been no user-friendly software package that can perform NMA or ML-NMR with survival outcomes; instead analysts have needed to rely on complex bespoke modelling code.

Methods: A recent update to the multinma R package provides a user-friendly suite of models and tools for synthesising survival outcomes from multiple trials, with aggregate data, individual patient data, or mixtures of both. Models are fitted in a Bayesian framework using Stan. A full range of parametric proportional hazards and accelerated failure time survival distributions are implemented, along with flexible baseline hazard models via M-splines or piecewise exponential hazards with a novel random walk shrinkage prior that avoids overfitting. Shape parameters may be stratified or regressed on treatment arm and/or covariates to relax proportionality. Right, left, and interval censoring, and delayed entry are all supported.

Results: We present analyses of two case studies using the multinma package.

First, we performed a NMA of published aggregate data from a network of treatments for advanced non-small cell lung cancer using flexible M-spline baseline hazards. We introduced treatment effects onto the spline coefficients to account for non-proportional hazards, and produced estimated survival curves in a target population required for further economic modelling.

Second, we performed a ML-NMR using a mixture of individual patient data and aggregate data from a network of treatments for newly-diagnosed multiple myeloma. We adjusted for effect-modifying covariates, and produced population-adjusted estimates for target populations of interest to decision-making. Covariate adjustment removed evidence for non-proportional hazards that was present in unadjusted models.

Conclusions: The multinma package makes NMA and ML-NMR methods accessible to a broad audience. The latest update to include a suite of functionality for survival analysis facilitates application of these methods to widespread settings such as oncology, where until now there was no user-friendly software available.

ORAL CONTRIBUTED SESSION 06
Meta-analysis

OC06-2 Path-based approach for detecting and assessing inconsistency in network meta-analysis: A novel method

Noosheen Rajabzadehtahmasebi, Adriani Nikolakopoulou, Theodoros Papakonstantinou
1 Institute for Medical Biometry and Statistics (IMBI), Freiburg, Germany

Introduction: Network Meta-Analysis (NMA) plays a pivotal role in synthesizing evidence from various sources and comparing multiple interventions. An essential aspect of NMA is assessing the consistency of evidence, ensuring that the direct (head-to-head comparisons) and indirect (comparisons through a common comparator) sources align.

Methods: In this presentation, we will examine two primary facets: (1) a concise review of established methods used to detect and assess inconsistency in NMA, and (2) a proposal for a novel path-based method to quantify the degree in which different sources of evidence agree when obtaining a specific network estimate. To achieve these goals, we showcase a toy example of a network with four treatments and five observed comparisons. In the first part, we explore some of the methods that are often used, including side-splitting (Dias, 2010), loop-specific (Lu, 2004), and the design-by-treatment interaction (Higgins, 2012), which are commonly utilized for identifying local and global inconsistencies in evidence networks. These approaches employ statistical tests to study inconsistencies between different sources of evidence, offering valuable insights into where inconsistency may exist. However, in some cases disparities between sources of evidence evade detection. Particularly, when the average discrepancy in effects of different pathways of a comparison is relatively small but their variance is high, these methods sometimes fall short to properly pinpoint the inconsistency in the network.

Results: In the second part, we emphasize the necessity for a path-based approach. Our proposed path-based approach offers a solution by examining the complete web of evidence. Furthermore, we introduce a measure based on the square of differences to quantitatively capture the extent of inconsistency within the network for each comparison. This measure aims to increase the accuracy of assessment and interpretability of inconsistency.

Conclusion: In conclusion, by adopting a pathway-centric perspective, our proposed method, accompanied by a novel inconsistency measure, holds the potential to enhance thorough examination of potential disagreements, thereby contributing to more informed healthcare decisions and policy recommendations.

ORAL CONTRIBUTED SESSION 06
Meta-analysis

OC06-3 Extensions of threshold analysis in network meta-analysis

Theofanis Ioannidis[1], Adriani Nikolakopoulou[1], Dimitrios Mavridis[2], Anna-Bettina Haidich[1]
1 Aristotle University, Thessaloniki, Greece
2 University of Ioannina, Ioannina, Greece

Background/Introduction: Network meta-analysis is a statistical methodology that extends pairwise meta-analysis, generating estimates of the relative effectiveness between multiple treatments. Threshold analysis systematically examines how changes in effect estimates from individual studies impact the network meta-analysis estimates, and subsequently the recommendation, based on a prespecified criterion (1). Such a criterion could be the probability of a treatment being best. Then, intervals within which the criterion remains unchanged are called invariant intervals. Despite the utility of invariant intervals, the method of threshold analysis lacks information about the heterogeneity and inconsistency within these predefined ranges. The aim of this work is to calculate and to indicate potential heterogeneity and inconsistency across the range of given invariant intervals for each study.

Methods: We provide a framework to derive how changes in treatment effects within invariant intervals impact on heterogeneity and inconsistency. Within a simulation framework, we monitor the heterogeneity parameter for pairwise and network meta-analysis, the inconsistency factor from the separating indirect and direct evidence method and the associated p-value and the p-value from the design by treatment interaction model within the invariant intervals. The proposed framework serves as an extension of threshold analysis to enhance the understanding of uncertainty and contribute to more informed recommendations.

Results: To illustrate the framework, we use an example with dummy data with six studies and three treatments. For the interpretation of results, we created line plots to indicate the evolution of explored metrics for each study within the boundaries of the invariant intervals.

Conclusion: The presented method aims to act as tool to inform clinicians, decision makers and researchers about the potential heterogeneity and inconsistency within invariant intervals in threshold analysis, thereby enhancing the reliability and validity of treatment recommendations. Additionally, in cases of excessive heterogeneity or inconsistency within the given range, it may suggest lack of reliability when conducting a network meta-analysis with the given variations in study level treatment effects across studies.

References: *[1] Phillippo DM, Dias S, Welton NJ, Caldwell DM, Taske N, Ades AE. Threshold analysis as an altern ative to grade for assessing confidence in guideline recommendations based on network meta-analyses. Ann Intern Med. 2019;170(8):538–46*

ORAL CONTRIBUTED SESSION 06
Meta-analysis

OC06-4 Meta-analysis of time-to-event data under model misspecification, based on median ratios

Fotios Siannis[1], Emmanouil Androulakis[2], Dimitrios Stogiannis[3]
1 National and Kapodistrian University of Athens, Athens, Greece
2 University of Piraeus, Piraeus, Greece
3 National Documentation Center, Athens, Greece

Introduction: Model misspecification in survival analysis usually has inferential consequences, such as loss of power and bias in parameter estimates. Inferences based on misspecified survival models when treatment effect is small was a topic of interest for Solomon [1] and Hutton and Solomon [2]. There seems to be an attractive analogy between the proportional hazards (PH) and the accelerated failure time (AFT) estimates when treatment effect is small, irrespectively of the nature and the form of the data. This offers a useful flexibility in model selection for the analysis. These results can be very important in a meta-analysis framework. Small treatment effects are of the essence, since the driving force of a meta-analysis is to bring together independent studies with small and/or statistically not significant results in an attempt to draw a more solid and conclusive result.

Methods: In this work, we extend the results of Solomon and Hutton regarding model misspecification, and we investigate this effect under the individual patient data meta-analysis framework of time-to-event endpoints. We consider the median survival times ratio (MR) as the quantity of interest, in an attempt to move away from modeling restriction. Our aim is to examine whether inferences made from misspecified models are aligned with inferences from the true model. For this purpose, we generated data from a multiplicative mixture model, combining a PH and an AFT model, and we fit the data using a number of standard models. We provide analytical expressions of the corresponding probability densities and likelihoods, for specific underlying distributions. An extensive simulation study is carried out in order to explore the different modeling scenarios and examine the impact of the model misspecification.

Results/Conclusions: It turns out that the MR is an informative and robust measure, and it can be utilized irrespectively of the modeling assumptions. Simulations based on small treatment effects showed that results based on the wrong models perform well. The same appears to be true even when the treatment effect is not small. Coverage probabilities and mean biases are utilized as means of measuring the misspecification impact.

References: *[1] Solomon, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. Biometrika, 291-298. [2] Hutton, J. L., & Solomon, P. J. (1997). Parameter orthogonality in mixed regression models for survival data. Journal of the Royal Statistical Society Series B: Statistical Methodology, 125-136.*

ORAL CONTRIBUTED SESSION 06
Meta-analysis

OC06-5 Extending multilevel network meta-regression to disconnected networks and single-arm studies: A case study on plaque psoriasis treatments

Samuel Perren, David Phillippo, Hugo Pedder, Nicky Welton
University of Bristol, Bristol, United Kingdom

Introduction: Single-arm studies and disconnected evidence networks are becoming common in health technology assessments. This has led to increasing use of unanchored population-adjustment methods for indirect treatment comparisons, which rely on strong assumptions that all prognostic factors and effect modifiers have been accounted for. Current methods have limitations: they only apply to pairwise indirect comparisons (one individual patient data (IPD) and one aggregate data (AgD) study), and can only produce estimates for the AgD population which may not represent the decision target population. Multilevel network meta-regression (ML-NMR) is an extension of network meta-analysis that overcomes some of these limitations to coherently incorporate both IPD and AgD, whilst adjusting for prognostic factors and effect modifiers in anchored population-adjusted analyses. However, ML-NMR has not yet been extended to unanchored analyses.

Methods: We extend ML-NMR to incorporate single-arm studies and disconnected networks, by predicting baseline response on a reference treatment for these disconnected studies using either other studies in the network or external evidence. We demonstrate with a network of treatments for moderate-to-severe plaque psoriasis, removing treatment arms to artificially disconnect the network, then reconnecting using different sources of evidence to predict reference treatment response, including external AgD from the PROSPECT and Chiricozzi 2019 cohort studies. We compare estimates of relative treatment effects and absolute outcome probabilities from the reconnected networks against those derived from the original connected network, treating the original connected analysis as the 'truth'. Analyses were performed in a Bayesian framework in R and Stan using the multinma R package.

Results: Analysis across multiple reconnected networks indicated variability in estimated treatment effects and absolute outcomes across study populations, depending on the evidence source used for reconnection. The PROSPECT reconnected network generally provided estimates close to those from the original connected network, whereas estimates from the Chiricozzi reconnected network deviated more substantially. This can be attributed to the similarity of the PROSPECT population to the disconnected study population in terms of prognostic and effect modifying factors, whereas the Chiricozzi population was more different.

Conclusion: We have demonstrated, for the first time, how ML-NMR may be extended to single-arm studies and disconnected networks. Performance of the reconnected network varied by the study population used for reconnection, emphasizing the need for careful evidence selection methods to assess bias. Future research should continue to explore new methods for network reconnection and methods for validating the strong assumptions required.

ORAL CONTRIBUTED SESSION 07
Statistics in Epidemiology I

OC07-1 Modelling of global estimates: An application of Bayesian modelling for low birthweight and preterm births

Ellen Bradley[1], Alexandra Lewin[2], Hannah Blencowe[1], Joy Lawn[1], Eric Ohuma[1]
1 Department of Infectious Disease Epidemiology and International Health, London School of Hygiene and Tropical Medicine, United Kingdom
2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

Introduction and Objective(s):Global estimates are used for setting priorities and global agendas, and are important for tracking progress of key health indicators at national, regional and globally. However, the data which the estimates are dependent on often has important data quality challenges and sparsity. The modelling of global estimates has to be rigorous and flexible to account for the data gaps and data quality. Currently no standard framework is available to model these vital statistics consistently.

We aimed to develop a statistical modelling framework that accounts that is flexible and can account for differing data sources, data quality, data gaps and other biases. Using available time-series data, we aimed to use this framework to estimate low birthweight (LBW) and preterm birth time-series for 195 UN member states at the national, and regional and global level.

Method(s): We developed a novel Bayesian hierarchical regression framework which we tailored to produce country, regional and global estimates for LBW and preterm birth and small vulnerable newborns. The models incorporated hierarchical country-specific intercepts, covariates, and non-linear time trends into the logistic regressions.

Two different sources of data were included in the estimates, which both had data availability, quality and biases that differed across the regions. We developed a data quality categorisation (DQC) to systematically group data points based on a pre-specified criteria of indicators of data quality. This allowed us to include all data points into a single model, which strengthened the estimates for countries without data, and accounted for the data quality differences and biases using weighting and bias shifts.

Results: We developed a flexible Bayesian modelling framework that can be used and adapted for other global estimates that accounts for various data quality issues, data gaps, varying data sources, and data temporality. We estimated national, regional and global estimates for low birthweight and preterm birth. Further, we expanded the framework to include conditional modelling for specific strata for LBW and preterm i.e., subgroups of pre-terms and birthweight strata, and the estimation of small vulnerable newborns.

Conclusions: Global estimates are essential for the setting and tracking of global health targets such as the sustainable development goals. Modelling global estimates have numerous challenges i.e., data sparsity, data gaps, and data quality issues. Developing a flexible Bayesian framework allowed us to use all the data appropriately and produce global estimates for LBW and preterm which are essential for tracking of international global targets.

ORAL CONTRIBUTED SESSION 07
Statistics in Epidemiology I

OC07-2 Estimating smooth age-specific contact rates in the population of Greece during the COVID-19 pandemic

Vasiliki Engeli, Vana Sypsa
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Background/Introduction: Social contacts play a key role in understanding how infectious diseases spread. Estimating age-mixing patterns, through social contact surveys, is important to determine disease spread between age groups [1], parameterise mathematical models and assess the impact of age-targeted interventions. The age of contacted individuals is usually coarsely reported, resulting in aggregated age-mixing mapping. Although one-year age intervals could lead to more precise results, methodological challenges arise due to data scarcity. Here, we aim to estimate age-specific contact rates through smooth contact surfaces, using as motivating example social contacts collected during COVID-19.

Methods: We analyzed data on mixing patterns collected through a social contact survey in Greece (September-October 2021, N=1,200). We recorded the exact age of the participant and the exact age or the age range of the contacted person. If an age range was specified, we imputed the exact contact age using (a) a discrete uniform distribution, or (b) the age distribution of the population in Greece within this range. We mapped the observed log-contact rates by one-year age intervals. Nonparametric regression using a local linear fit to age-specific contact data was applied to obtain social contact surfaces. Reciprocity of contacts was considered. As sensitivity analysis, we artificially increased the contact rates of children/adolescents in the main diagonal of the smoothed contact matrix to account for possible underestimation due to smoothing [2].

Results: No contacts were observed for several combinations of ages of the participants and contacts. We dealt with scarcity of detailed data by mapping smooth contact surfaces, in which different contact rates by age were depicted. Both imputation methods for the exact contact age yielded similar contact surfaces. Indicatively, the highest contact rates by one-year of age were estimated among young individuals (5-24 years) with their peers (0.5-0.8 daily contacts/individual –up to 1.4 after artificial increase). Adults aged 25-55 years had 0.2-0.5 daily contacts/individual with adults of similar age, while the elderly (60+) had low contact rates with their peers (<0.2 daily contacts/individual).

Conclusion: Smooth contact surfaces can provide estimates for contact rates for various combinations of ages and can improve the performance of age-structured models for infectious diseases in the absence of detailed data

References: *[1] Wallinga et al., American Journal of Epidemiology, 2006, 164.10:936-944. [2] Vandendijck et al., Biostatistics, 2023, kxad005.*

ORAL CONTRIBUTED SESSION 07
Statistics in Epidemiology I

OC07-3 Dissecting COVID-19's spread in the Basque Country: A Bayesian spatio-temporal analysis

Lander Rodriguez[1], Irantzu Barrio[1,2], José Maria Quintana-López[3], Nere Larrea[3], Miguel Ángel Martínez-Beneito[4]
1 Applied Statistics, Basque Center for Applied Mathematics, Spain
2 Department of Mathematics, University of the Basque Country (UPV/EHU), Spain
3 Research Unit of the Galdakao-Usansolo University Hospital, Osakidetza Basque Health Service, Spain
4 Department of Statistics and Operations Research, University of Valencia (UV), Spain

Background: The transmission dynamics of the COVID-19 disease are possibly driven by the interaction of sociodemographic and environmental factors, leading to heterogeneous spatiotemporal patterns. Small area estimation techniques are appropriate to investigate the interaction of these factors and to obtain a more profound understanding of the disease transmission in a specific spatio-temporal frame. In this work, we assess the evolution of the COVID-19 pandemic in terms of cases, hospitalizations, deaths and reinfections in the Basque Country using Bayesian hierarchical spatio-temporal techniques.

Methods: This was a retrospective cohort study of patients diagnosed with COVID-19 in the Basque Country from March 1, 2020 until January 9, 2022. The 4-type spatio-temporal interaction model of Knorr-Held [1] was adopted along with the Integrated Nested Laplace Approximation (INLA) to obtain reliable estimates of cases, hospitalizations, deaths and reinfections on a weekly basis in all the primary care units (PCU) of the Basque Country. In addition, the influence on the different outcomes of the type of area (urban, mixed or rural), a deprivation index and the number of tests per 1000 inhabitants of each PCU was studied.

Results: Relative risks of cases, hospitalizations, deaths and reinfections were obtained on a weekly basis for all PCUs. The number of tests performed was relevant on all outcomes, which can be explained considering the test was a necessary and prior condition to diagnose any of these outcomes. Rural areas experienced a lower risk of cases, hospitalizations and reinfections while the type of area was not relevant in terms of deaths. Actually, the virus probably circulated to a greater extent in more populated areas and the increased distance from rural areas to hospitals may have prevented hospitalizations from these regions. Finally, the deprivation index was also relevant in all of the outcomes.

Conclusions: In this study we have assessed the evolution of the COVID-19 pandemic in the Basque Country with Bayesian spatio-temporal hierarchical models. Rural areas have been related to a lower risk of COVID-19 positive cases, hospitalizations and reinfections whereas the more deprived areas have been related to a higher risk of cases, hospitalizations, deaths and reinfections. These conclusions may be considered in future pandemics when planning intervention strategies to provide improved medical attention to the most vulnerable individuals.

References: *[1] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. Stat Med. 2000 Sep 15-30;19(17-18):2555-67.*

ORAL CONTRIBUTED SESSION 07
Statistics in Epidemiology I

OC07-4 A data-driven approach for studying the interplay between psoriasis and multiple conditions: A phenome-wide association and Mendelian randomisation study

Christos Chalitsios[1], Alexandros-Georgios Asimakopoulos[1], Christos Papagiannopoulos[1], Georgios Markozannes[1,2], Kostas Tsilidis[1,2], Ioanna Tzoulaki[2,3]

1 Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece
2 Department of Biostatistics and Epidemiology, School of Public Health, Imperial College London, London, United Kingdom
3 Centre for Systems Biology, Biomedical Research Foundation, Academy of Athens, Athens

Background: Psoriasis often co-exists with other conditions, but existing studies test the association between psoriasis and a pre-defined set of conditions. In this hypothesis-free, data-driven phenome-wide association (PheWAS) and Mendelian randomisation (MR) study, we determined the possible causal interplay between psoriasis and multiple conditions.

Methods: We used genome-wide data from the largest GWAS meta-analysis to construct a polygenic risk score (PRS) of 46 variants related to psoriasis. Then, we performed a PheWAS testing the association between the psoriasis-PRS with phenotypes using multivariable logistic regression adjusted for age, sex, and first 15 principal components. We then implemented a bidirectional 2-sample MR analysis to investigate the causal relevance between psoriasis and the conditions identified and passed stringent correction for multiple testing from the PheWAS using inverse variance-weighted (IVW) MR. We also did several sensitivity analyses to assess pleiotropy and the robustness of estimation.

Results: Our study population comprised 373,776 UK Biobank participants (mean [SD] age, 57.2 [7.9] years; 203,523 females [54.5%]), and analyses were done for 1,136 unique phecodes from 17 different disease categories. Psoriasis-PRS showed the strongest signals that passed the FDR threshold for ankylosing spondylitis (p=3.52E-05), celiac disease (p=5.39E-05), and skin cancer (p=8E-05); however, only one association was supported by the MR analyses. Specifically, genetically predicted ankylosing spondylitis was associated with a higher risk of psoriasis (ORIVW=2.629, 95%CI 1.453-4.758, p=1.39E-03) with all the sensitivity analyses supporting this finding. Using the validation population (using data from the FinnGen GWAS), this association was also confirmed from both primary and sensitivity analyses.

Conclusions: To our knowledge, this is the first study to comprehensively and rigorously assess the potential bidirectional effects of a wide range of conditions on psoriasis. Strong evidence was found for ankylosing spondylitis as a new risk factor for psoriasis.

ORAL CONTRIBUTED SESSION 07
Statistics in Epidemiology I

OC07-5 Heterogeneity in the acquisition of multiple infections: A mathematical and statistical modelling perspective

Alessandro Grosso[1], Niel Hens[2,3], Steven Abrams[1,2]
1 University of Antwerp - Department of Family Medicine and Population Health - Global Health Institute, Anwerp, Belgium
2 Hasselt University - Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt, Belgium
3 University of Antwerp - Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute, Antwerp, Belgium

Background: Disease dynamics are driven by heterogeneities in risk factors, transmissibility, and susceptibility to infection, often unaccounted for when estimating epidemiological parameters. Mathematical compartmental models could account for them by clustering individuals in different risk groups, expanding the number of compartments. Statistical models can encompass heterogeneity with latent random variables, termed 'frailties', typically acting multiplicatively on a baseline infection hazard. Frailty models have been applied in the context of multivariate serological data to gauge the strength of association in the acquisition of infections and elucidate common elements in transmission routes. We aim to provide an explicit connection between mathematical and statistical approaches when describing infection dynamics for two pathogens in the presence of unobserved heterogeneity in infection risk and associations determined by, for instance, shared transmission routes.

Methods: We consider the use of serological data, leading to current status data, generated from a bivariate Susceptible - Infected - Recovered (SIR) compartmental model mimicking disease spread at the population level. Age at time of cross-sectional data collection and immunological status regarding the pathogens under consideration are simulated from the model. Heterogeneity in acquisition and transmissibility of infection is induced by considering group-specific hazard functions (forces of infection in infectious disease epidemiology) featuring frailty terms, from a discrete K-point distribution, acting multiplicatively on the infection-specific hazard. Frailty models, differing in the amount of information available regarding risk group membership, are fit to these current status data and compared in their ability to capture underlying disease dynamics. Complete lack of information was dealt with using an Expectation-Maximization algorithm.

Results: Ignoring heterogeneity leads to biased and inconsistent estimates for the infection hazard. Acknowledging heterogeneity, even under increasing imperfect information regarding risk group membership, leads to unbiased and consistent estimates of the infection rates. Misspecification of the number of risk groups does not invalidate estimates related to the forces of infection. We show how information criteria can be used to compare different models to obtain an estimate of the number of risk groups. Moreover, using a continuous frailty distribution provides unbiased and consistent estimates of the frailty variance, at the cost of not being able to elucidate the risk group partition in the population.

Conclusion: We demonstrate that the use of shared frailty models for bivariate current status data is a sensible approach when estimating the infection hazard in the presence of individual- and risk-group specific heterogeneity in the acquisition of infections.

ORAL CONTRIBUTED SESSION 08
Diagnostics

OC08-1 Improving hypotheses testing with external information knowledge for achieving higher efficiency in non-normal population

Sunil Mathur[1,2], Ethan Burns[2], Shreya Mathur[3], Kai Sun[2], Ravi Pingali[3], Jenny Chang[2]
1 Weill Cornell Medical College, Houston, United States
2 Houston Methodist Neal Cancer Center, Houston, United States
3 Boston Medical Center, Boston, United States

Background/Introduction: Triple-negative breast cancer (TNBC) poses a formidable challenge due to its aggressive and resistant nature. For individuals with locally advanced unresectable or metastatic TNBC (mTNBC), the median overall survival (OS) ranges between 8 to 13 months, underscoring the urgent need for effective treatments. Assessing drug efficacy for TNBC treatment stands as a formidable challenge in cancer research, pivotal for informed medical decisions, healthcare planning, and empowering patients in their healthcare choices. The consequential improvements in care quality, error avoidance, adverse event reduction, enhanced efficiency, cost-benefit optimization, and heightened satisfaction among providers and patients underscore the significance of efficacious medicine. Due to the nature of diseases, the patient population seldom follows a normal distribution.

Methods: To address this challenge, we propose a test based on ranked-set empirical distribution functions using external information knowledge and construct a distribution-free competitor to hypotheses testing procedures under nonnormal conditions which would overcome some of the restrictions imposed on earlier tests. Here we use the power divergence between two empirical distribution functions to define the proposed test statistic.

Results: Our test statistic, which is based on a power divergence between empirical distribution functions derived from two independent samples, proves particularly potent in comparison to alternatives under both heavy-tailed and light-tailed distributions. Employing the permutation principle fortifies the test's implementation. Utilizing the Monte Carlo method, empirical power computations demonstrate the superiority of our test across heavy-tailed, light-tailed, and even elliptically asymmetric population distributions. We show the implementation of the proposed test using real data.

Conclusion: In contrast to Hotelling's T^2 and Chatterjee and Sen's bivariate Wilcoxon rank sum test, the innovative approach we propose eliminates the need for estimating a dispersion matrix. This presents a significant advantage over existing methods, as the sample dispersion matrix may become singular, rendering its inverse non-existent and thereby hindering the implementation of the test. Unlike the Kolmogorov-Smirnov test, which utilizes the empirical distribution function in a univariate setting for both one and two-sample problems, its multivariate extension is limited to one sample. To the best of our knowledge, there is currently no test available for two samples, apart from our proposed test, that employs the empirical distribution function, addressing a notable gap in the existing literature. In summary, our proposed test consistently exhibits superior power, regardless of the underlying population characteristics, offering a robust solution to drug efficacy assessment, and the proposed test can be applied to other situation.

ORAL CONTRIBUTED SESSION 08
Diagnostics

OC08-2 Construction and evaluation of optimal diagnostic tests

Ainesh Sewak, Sandra Siegfried, Torsten Hothorn
University of Zurich, Zurich, Switzerland

Introduction: Accurate diagnostic tests are crucial to ensure effective treatment, screening, and surveillance of diseases. However, the limited accuracy of individual biomarkers often hinders comprehensive diagnostic assessment. The heterogeneity of many diseases, particularly cancer, calls for the use of several biomarkers together into a composite diagnostic test. There is longstanding knowledge from signal detection theory that the likelihood ratio function yields optimal biomarker combinations.

Methods: We present a novel multivariate model that optimally combines multiple biomarkers using the likelihood ratio function. The model has various levels of flexibility that can be tuned for specific applications, each having built-in performance measures such as ROC curves and AUC. Our method allows for reliable predictions even in scenarios where specific biomarker measurements are unavailable and can guide the selection of biomarker combinations under resource constraints. The method is applicable for both screening and as an adjunct diagnostic tool alongside imaging modalities, guaranteeing enhanced accuracy in disease detection.

Results: We demonstrate the efficacy of our method by constructing an optimal diagnostic test for hepatocellular carcinoma (HCC), a cancer type known for lacking a single ideal marker. While the serum marker alpha-fetoprotein (AFP) remains widely used for HCC screening, its utility is limited due to elevated levels appearing in other benign liver conditions. Utilizing data from a retrospective case-control study, we construct and evaluate an optimal diagnostic test incorporating AFP and additional biomarkers for HCC detection.

Conclusion: The combination of multiple biomarkers offers the potential for significant performance improvements over individual markers. Our method stands on par with machine learning models while remaining characterized by simple entities which have direct interpretations in terms of diagnostic accuracy metrics. Further, our model assumptions can be assessed in a data-driven manner. An accompanying R implementation is provided for reproducibility of all results.

References: *[1] Ainesh Sewak, Sandra Siegfried, and Torsten Hothorn. Construction and evaluation of optimal diagnostic tests with application to hepatocellular carcinoma diagnosis. 2024. URL https://arxiv.org/abs/2402.03004. arXiv:2402.03004 [stat.ME]*

ORAL CONTRIBUTED SESSION 08
Diagnostics

OC08-3 Development of a time-dependent AUC estimator for competing risks models

Leire Garmendia Berges[1,2], Irantzu Barrio[1,2], Guadalupe Gómez Melis[3]
1 Basque Center for Applied Mathematics, Spain
2 Department of Mathematics, University of the Basque Country UPV/EHU, Spain
3 Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Spain

Background/Introduction: Competing risks situations often appear in survival analysis when the endpoint of interest (e.g., breast cancer) is precluded by another event (e.g., death). Those models help clinicians to understand the disease process of a patient and are commonly used to predict the course of the disease for a new patient. A high predictive capacity measure, commonly given by the area under the time-dependent ROC curve (AUC(t)), is a useful aid when making clinical decisions. To date, there are no proposals to measure the time-dependent predictive capacity of a competing risks model.

Methods: In our work, we provide an estimator for the global AUC(t) for a competing risk model. We propose a global AUC(t) as a weighted mean of the partial AUC(t) of each transition (AUCk(t) k = 1, …, K), with each AUCk(t) being weighted by the probability of experiencing that event k before time t. Since the competing risks model can be conceptualised as a combination of transitions where different survival models are fitted, we can use estimators proposed in the literature for survival models to estimate each AUCk(t). The proposed AUC(t) evaluates the predictive capacity of the entire competing risk model at any time t of interest.

Results: A multicentre study of almost 700 hospitalized adult COVID-19 patients from five Catalan hospitals during the fifth wave of the Spanish pandemic was conducted. The aim of this study was to analyse the evolution of hospitalized patients towards more severe states, such as Non-Invasive Mechanical Ventilation (NIMV) or Invasive Mechanical Ventilation (IMV), using a competing risks model and to make predictions for new individuals based on this model. We analyse the discrimination ability of this model using the proposed global AUC(t) of the competing risks model. Our findings indicate that the competing risks model has a good predictive capacity with promising results observed for the proposed AUC(t).

Conclusion: The proposed estimator provides a measure for the global predictive capacity of a competing risks model.

ORAL CONTRIBUTED SESSION 08
Diagnostics

OC08-4 Community-driven Bayesian inference for sociocultural questionnaire instrument validation

Nia Kang
McGill University, Montreal, Canada

Introduction: Cognitive assessment instruments (CAIs) are widely used in medical practice to assess cognitive function in individuals who experience a potentially pathological decline in cognition that impedes everyday functioning. The validation of CAIs usually involves a reference population, but dynamics in population characteristics due to factors such as immigration can interfere with the instrument's measurement properties, potentially rendering the instrument's diagnostic performance as suboptimal. With the emerging diversity in sociodemographic profiles around the world, there remains untapped opportunity to leverage representative expert knowledge and opinions to safeguard and promote the appropriate adaptation of instruments such as CAIs for underrepresented populations.

In contrast to conventional Bayesian applications, 'Community-Driven Bayesian Inference' (CDBI) prioritizes knowledge holder input by intentionally exaggerating the weight of the prior information (i.e., expert voices) in comparison to its empirical counterpart (i.e., likelihood function). We propose an accessible parametric prior distribution that grants lay-experts agency in endorsing the evident and latent attributes of a CAI, and their level of certainty in doing so.

Methods: We apply analytical methods to derive appropriate expert prior parameterizations for initially validated CAIs that employ inverse probability weighting by adapting a Horvitz–Thompson estimation approach. We conservatively assume modest to low certainty to achieve a clinically meaningful discrepancy in the maximum likelihood estimate and the resulting highest posterior density estimate. Through this process, we ensure that the voices of underrepresented end-users have sufficient weight to potentially challenge currently established evidence of construct validity in CAI.

Using a Monte Carlo Simulation study, we demonstrate the application of the proposed CDBI approach in the context of two widely used CAIs, the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA). The simulated CAI data reflect currently established construct validity metrics reported for the MMSE and MoCA in the literature, alongside hypothetical expert prior input data (varying numbers of experts with varying levels of item endorsements and respective certainties).

Results: Our research study demonstrates that even in the presence of large empirical validation studies for CAIs, Community-Driven Bayesian Inference enables relatively low numbers of experts ($n<50$) to provide robust evidence on the validity (and applicability) of the CAIs in an underrepresented target population.

Conclusion: Bayesian instrument validation methods incorporating expert input are effective in formally appraising the sociocultural appropriateness of clinical decision-making tools.

ORAL CONTRIBUTED SESSION 08
Diagnostics

OC08-5 Methods for comparing ROC curves under the presence of covariates

Juan Carlos Pardo-Fernandez[1], Aris Fanjul-Hevia[2], Wenceslao Gonzalez-Manteiga[3]
1 Universidade de Vigo, Spain
2 Universidad de Oviedo, Spain
3 Universidade de Santiago de Compostela, Spain

Introduction: The Receiver Operating Characteristic (ROC) curve is a graphical tool routinely used for evaluating the performance of a binary classification procedure based on a continuous marker. The ROC curve is constructed by depicting, for each potential threshold in the support of the marker, the rate of positive subjects (those with the characteristic/disease under study) correctly classified as positive, or sensitivity, against the rate of negative subjects (those without the characteristic/healthy) incorrectly classified as positive, that is, 1 - specificity. In many practical applications, covariates related to the marker are available. Under these circumstances, it is of interest to evaluate the influence that those covariates might have in the performance of the marker in terms of classification ability. Two extensions of the ROC curve have been proposed in the literature: the covariate-specific ROC curve, which is defined in terms of conditional distributions, and the covariate-adjusted ROC curve, which can be seen as a sort of average of the covariate-specific curves.

Methods: In this talk we will present two methods to compare ROC curves. First, a test to compare covariate-specific ROC curves will be discussed. In practice, this test would allow to decide if, for a given value of the covariate, the classification capabilities of several markers differ. Second, a method for testing the equality between the ordinary ROC curve and the covariate-adjusted ROC curve will be introduced. This test can be employed to evaluate the convenience of incorporating the covariate to the ROC analysis. The proposed methodologies rely on nonparametric estimation of the involved ROC curves and bootstrap resampling plans to approximate the null distribution of the test statistics.

Results: Simulations show the that proposed tests are correctly calibrated under the null hypothesis and reach reasonable power under the alternative. Applications to real data (in particular, to a dataset of patients with pleural effusion) show that the tests are useful to analyse possible differences in the diagnostic properties of the markers when taking covariates into account.

Conclusion: The proposed methods are shown to be useful in practice to analyse possible differences in the classification capability of markers when considering covariates.

ORAL CONTRIBUTED SESSION 09
Joint Models II

OC09-1 Combining and optimising dynamic predictions with Super Learner

Nina Van Gerwen, Dimitris Rizopoulos, Sten Willemsen, Bettina Hansen
Erasmus Medical Center, Rotterdam, Netherlands

Background: Dynamic predictions for longitudinal and time-to-event outcomes have become a versatile tool in precision medicine. Our work is motivated by the application of dynamic predictions in the decision-making process for primary biliary cholangitis patients (PBC) and the Global PBC Study Group database. For these patients serial biomarker measurements (e.g., bilirubin and cholesterol levels) are routinely collected to inform the risk of liver failure and guide treatment decisions. Two popular statistical approaches to derive dynamic predictions are joint modeling and landmarking. However, recently machine learning techniques have also been proposed. Each of these approaches have their merits and flaws, and no single method yet exists to outperform all the others. Consequently, obtaining the best possible survival estimates is challenging. Therefore, we develop a novel framework inspired by Super Learner to combine dynamic predictions from different models and procedures. Super Learner is an ensemble method that allows users to combine several different prediction algorithms into one where the candidate algorithms can be quite different. It uses cross-validation to build the optimally weighted combination of predictions from a library of candidate algorithms. An advantage of Super Learner is that it allows for different objective functions to combine the predictions that best suit the specific application context (e.g., minimising the mean squared error or maximising the area under the receiver operating characteristic curve). In our work, we pay special attention to appropriate functions for Super Learner to obtain the most optimal weighted combination of dynamic predictions.

Methods: We illustrate super learning for dynamic prediction on the Global PBC Study Group database. We apply an ensemble of six models to obtain accurate dynamic survival curves of patients suffering from PBC and we compare these results to the survival curves of each model fit separately. For the meta-learner, we investigate three loss functions: the dynamic Area Under the ROC Curve, Inverse Probability Censoring Weighted Brier Score and Integrated Brier Score. In a simulation study, we further validate the performance of our proposed method.

Results: In the Global PBC Study Group database, we noted small differences in predictive performance of the ensemble and the models fit separately. Depending on the metric, the ensemble sometimes showed improved predictive accuracy compared to the single models. The simulation study showed that the method works properly.

Conclusion: Combined dynamic predictions through super learning are a worthwhile technique for researchers to employ when there is high emphasis on accurate predictions and uncertainty in model selection when working with longitudinal and time-to-event data.

ORAL CONTRIBUTED SESSION 09
Joint Models II

OC09-2 Personalised monitoring schedules for multiple types of measurements with application in chronic heart failure

Teun Petersen, Isabella Kardys, Dimitris Rizopoulos
Erasmus MC, Rotterdam, Netherlands

Background: In the setting of chronic heart failure (CHF), routine periodic outpatient visits are commonly performed to monitor disease progression. During these visits, various measurements and tests can be performed, such as blood biomarker measurements, and echo- and electrocardiograms. We propose an adaptive strategy to decide for a given patient when the next visits should be planned and what to measure, balancing patient-specific risk of HF exacerbation, information gained on prognosis and burden on the patient and the healthcare system.

Methods: Consider a monitored CHF patient at time t, followed up using k types of longitudinal outcomes. We want to determine a schedule specifying when and how many times to measure every longitudinal outcome in the near future. The strategy consists of a two-step approach. First, a time frame is chosen in which the measurements are optimized, starting at time t and ending when the estimated cumulative risk of HF exacerbation given the patient's longitudinal history, reaches a prespecified cut-off. This cumulative risk is estimated using joint models for longitudinal and time-to-event data. Second, within this time frame, times and types of measurements are optimized based on information gained on the patient's prognosis. This information gain can be quantified by comparing the estimated distributions of time of HF exacerbation at the end of the time frame with and without the new measurements via their expected Kullback-Leibler divergence. Performance of the adaptive strategy, compared to standard fixed periodically scheduled measurements, will be assessed using simulations based on a cohort study of CHF patients.

Results: Prior results indicate that similar adaptive scheduling strategies for single longitudinal outcomes show a significant reduction in measurements, without compromising on monitoring quality. We expect to see similar results in our approach to scheduling multiple longitudinal outcomes.

Conclusion: Our method could lead to more efficient use of healthcare resources without compromising on the standard of care. Future research should be aimed at validating this approach and making it accessible for use by clinicians.

ORAL CONTRIBUTED SESSION 09
Joint Models II

OC09-3 Dynamically predicting survival benefit to prioritise liver transplantation in hepatocellular carcinoma patients

Pedro Miranda Afonso[1], Hao Liu[2], Michele Molinari[2], Dimitris Rizopoulos[1]
1 Erasmus University Medical Center, Rotterdam, Netherlands
2 University of Pittsburgh Medical Center, Pittsburgh, United States

Background/Introduction: Liver transplantation provides the best survival outcomes for patients with hepatocellular carcinoma (HCC). However, as the demand far exceeds the number of available organs, it is crucial to identify the patients who will benefit most from liver transplantation. The α-fetoprotein (AFP) level, the radiographic tumour burden score (TBS), and the model for end-stage liver disease (MELD) score are routinely measured to monitor HCC progression and liver dysfunction to inform transplantation decisions. The transplant-related survival benefit has been proposed as a comprehensive metric combining post-transplant and waiting list life expectancies. Despite recent advances, existing methods of estimating this metric disregard the full longitudinal information, the observational nature of the data, and the influence of time-varying confounders. Our primary goal is to dynamically predict the individualized transplant-related survival benefit in HCC patients to improve transplant prioritization. We analyse data from 27,000 patients listed in the US Scientific Registry for Transplant Recipients (SRTR) between March 2002 and March 2022.

Methods: We propose a joint model that associates the risk of death, before and after transplantation, with the pre-transplant AFP level, the left-censored TBS, and the MELD score trajectories. We establish the assumptions to unbiasedly estimate the causal transplantation effects from the observational data at hand. Using the postulated joint model, for each patient at risk at a given time, we predict their pre- and post-transplant five-year survival probabilities. Then, we determine the causal transplantation benefit as the difference between them. This can be used as a score to discern which patients would derive the greatest benefit from an available organ. The individual scores can be updated as new marker measurements become available, serving as a dynamic scoring system. The model is available in the R package JMbayes2.

Results: Our results demonstrate that AFP, TBS, and MELD each have distinct associations with the risk of death before and after transplantation, confirming the value of the new approach. Despite the observational nature of the SRTR data, our model can unbiasedly estimate the causal effect of transplantation on survival, leading to unbiased individualized transplant-related benefit scores.

Conclusion: Our dynamic benefit score ensures a fair allocation of the limited number of liver transplants, optimizing the use of available organs and improving the overall survival of all waitlisted patients. Our scoring system will be easily accessible through a web-based tool to assist clinical decision-making.

ORAL CONTRIBUTED SESSION 09
Joint Models II

OC09-4 Impact of baseline hazards in jointly modelling longitudinal and terminal outcomes in cancer survival

Anand Hari, Divya Dennis, Jagathnath Krishna KM
Regional Cancer Centre, Thiruvananthapuram, Thiruvananthapuram, India

Introduction: Joint modelling is a powerful statistical framework enabling a simultaneous modelling of longitudinal covariate and time-to-event outcome. The complexity of the model makes the computation more intensive and time consuming. This motivated us to look into the role of baseline hazards in joint modelling. Thus in the present study we considered four joint models with different choices of baseline hazards. The present study highlights the importance of choosing the baseline hazards to get more accurate risk estimation and consistency was assessed using a simulation study.

Methods: Here we considered joint models with four different baseline hazards [Piecewise-constant Gauss-Hermite (GH), Piecewise-constant pseudo-adaptive GH, Weibull Accelerated Failure time model with GH & B-spline GH]. We also conducted a comprehensive simulation study to assess the model consistency with varying sample size (N=100,250,500) and censoring (20%, 50%) scenarios, with 50 replications. The performance of the models was evaluated using the Akaike information criteria (AIC) and Bayesian information criteria (BIC), which measures the goodness of fit while penalizing the model complexity. We also used Bias, Root Mean Square Error (RMSE), and coverage probability (CP) to evaluate consistency and precision of parameter estimates. Furthermore, computational efficiency was measured by runtime. The analyses were performed using R Software and illustrated using colon cancer data.

Results: In colon cancer patient data we obtained the Piecewise-PH-aGH as the optimal model based on the information criteria (AIC and BIC). Based on the Piecewise-PH-aGH model composite stage and Carcinoembryonic antigen (CEA) values were found to be significant predictors for survival. Further the consistency of the models was assessed using simulation study resulting in Piecewise-PH-aGH as the best model with least AIC and BIC values, and higher CP across all sample sizes. While the Bias, and RMSE for all the models showed a competitive performance. However, Piecewise-PH-aGH has shown better bias and RMSE in most cases and has taken the shortest computation time, which shows its computational efficiency.

Conclusion: The present study identified the Piecewise-PH-aGH model as the optimal choice for baseline hazards in joint model. The study identified composite stage as a prognostic factor for time-to-event and the longitudinal outcome, CEA as a dynamic predictor for overall survival in colon cancer patients. Further the simulation study established the computational efficiency of Piecewise-PH-aGH model over other joint models. This study is the first of its kind to discuss on the choice of baseline hazards.

ORAL CONTRIBUTED SESSION 09
Joint Models II

OC09-5 Diagnosis of cardiovascular diseases using interpretable cardiac magnetic resonance-derived latent factors

Andrea Mario Vergani[1,2], Soumick Chatterjee[1], Sara Ometto[1], Carlo Andrea Pivato[1,3], Alessia Visconti[1,4], Craig A. Glastonbury[1], Marco Masseroli[2], Francesca Ieva[1,2], Emanuele Di Angelantonio[1,5]
1 Human Technopole, Milan, Italy
2 Politecnico di Milano, Milan, Italy
3 IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
4 Università di Torino, Turin, Italy
5 University of Cambridge, Cambridge, United Kingdom

Background/ Introduction: Magnetic resonance imaging (MRI) accurately assesses cardiac structure and function, representing the gold standard for diagnosing certain cardiovascular diseases (CVD). In recent years, deep learning has improved the performance of automated diagnosis of CVD using heart MRI. To guarantee interpretability, the models may rely on segmentation to extract intermediate cardiac features; hence, these supervised approaches require large quantities of annotated MRI for training. To overcome this limitation, this study leverages deep representation learning to extract unsupervised features from cardiac MRI. We propose a framework to interpret these latent factors and use them to enhance CVD diagnosis, aiming to improve the classification performance of traditionally extracted cardiac measures.

Methods: We extracted latent vectors of size 128 from four-chamber long-axis cardiac MRI with a Diffusion Autoencoder, trained on magnetic resonance acquisitions from about 60,000 UK Biobank participants. We then compared the classification performance of our latent factors versus traditional cardiac measures (e.g., ventricular measures) for the presence of different subsets of CVD (e.g., cardiomyopathy, left ventricular failure); specifically, our classifiers were trained on balanced sets (50% cases – 50% controls, matched on age and sex). Additionally, we leveraged latent space arithmetic techniques and genetic data to enhance the explainability of our deep representation learning model, colocalizing Genome-Wide Association Study (GWAS) hits of the latent factors and known cardiac phenotypes.

Results: Our Diffusion Autoencoder achieved state-of-the-art reconstruction quality, with an average structural similarity index measure of 0.95. The classification models with MRI-derived latent vectors outperformed the same models with traditional cardiac measures as covariates, with improvements in accuracy of up to 15%; the impact of latent factors proved to be more evident for atrial arrhythmias and structural heart diseases. Through predictivity analyses with sparse models and latent manipulation, we interpreted the clinical meaning of more than 50% of our latent features; additionally, we mapped about 30% of the latent factors to genome regions significantly associated with ventricular and aorta measures, pulse rate, atrial fibrillation, and other CVD subtypes.

Conclusion: We propose an explainable deep learning approach to extract latent features from heart MRI, which proved to be more predictive of the presence of CVD than traditional cardiac measures. Our framework combines latent representations, cardiac phenotypes, and genetic data to enhance interpretability and clinical translation. To the best of our knowledge, this is the first work involving explainable unsupervised deep representation learning of cardiac MRI, also considering genetic information in the interpretability framework, to improve CVD diagnosis.

ORAL CONTRIBUTED SESSION 10
Survival Analysis I

OC10-1 Regression modelling in an extended multi-state model using relative survival

Damjan Manevski[1], Hein Putter[2], Maja Pohar Perme[1], Eva A. E. Koster[2], Johanneke E. A. Portielje[3], Liesbeth C. de Wreede[2]

1 Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
2 Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
3 Department of Medical Oncology, Leiden University Medical Center, Leiden, Netherlands

Background/Introduction: In survival analysis, multi-state models have become a staple tool for considering intermediate events, apart from death as the main event of interest. As for many diseases the number of deaths not related to the disease in question is non-negligible, our interest lies in distinguishing between disease-specific (excess) and other (population) mortality occurring after intermediate events. The field of relative survival deals with this distinction when the cause of death information is not provided. In this work, we are interested in applying the relative survival approach in the multi-state model framework for estimating covariate effects and calculating subject-specific predictions.

Methods: In a recent paper [1], we have proposed an extended multi-state model using relative survival in which suitable definitions of transition hazards and probabilities are introduced as well as their non-parametric estimators. In this work, we explore regression modelling in the extended model. We expand the Cox-type multiplicative model which is commonly used in multi-state models. An existing implementation for modelling the disease-specific hazards based on the EM algorithm [2] is upgraded for multi-state models where delayed entry has to be considered in the intermediate states. However, as issues arise in the case when the estimated overall mortality is smaller than the background mortality, we consider an alternative - the Aalen additive hazards model which can deal with this issue, thus we also extend it in the multi-state setting.

Results: A simulation study compares the two approaches in various data settings, especially when considering delicate factors such as delayed entry and small excess death rates. The usage of these models is illustrated using Dutch breast cancer data. The R code needed for applying the implemented work is shown using the packages mstate and relsurv.

Conclusion: As the total mortality in the multi-state model is split into population and excess components, questions arise on the covariate effects and long-term patient predictions that can be obtained from such a model. The two regression approaches address these questions and provide a further understanding of the studied disease.

References: *[1] Manevski D, Putter H, Pohar Perme M, Bonneville EF, Schetelig J, de Wreede LC. Integrating rela- tive survival in multi-state models-a non-parametric approach. Stat Methods Med Res. 2022;31(6):997- 1012. [2] Pohar Perme M, Henderson R, Stare J. An approach to estimation in relative survival regression. Biostatistics. 2009;10(1):136-146.*

ORAL CONTRIBUTED SESSION 10
Survival Analysis I

OC10-2 Use of pseudo-observations in the development and validation of prognostic models for competing risks

Molly Wells[1], Mark J. Rutherford[1], Paul C. Lambert[1,2]
1 Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom
2 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Background/Introduction: Pseudo-observations (PSOs) have increased in popularity in survival analysis due to the ability to analyse survival data using standard regression techniques. PSOs are calculated for each individual based on jackknifing and can then be used as an outcome variable. Using PSOs, we can estimate a parameter using the mean of value of the PSOs at a fixed point in time and use generalised linear models (GLM) to model the effect of covariates on the outcome.

To obtain PSOs, we need to specify a well-behaved estimator of the parameter of interest. If the parameter of interest is the survival function at a specific time, t, then the Kaplan-Meier estimator can be used. Similarly, if investigating competing risks or multistate models, the parameter of interest would be the cumulative incidence function (CIF) at time t, and the Aalen-Johansen estimator would be used.

Methods: For prognostic modes with competing risks, we extended the use of PSOs and compared them to standard survival models. PSOs are calculated for the probability of being in each disease state so that the PSOs sum to one. We propose a model that simultaneously models the PSOs for all disease states, rather than using separate models for each state. This approach is similar in principle to multinomial logistic regression. Our approach provides predicted probabilities for all disease states, which will always sum to one. We provide software to fit the model.

The use of PSOs also simplifies some standard measures of model discrimination and validation. For example, they are used to estimate the Brier score, expected to observed ratios, and time-dependent ROC curves to assess model discrimination. The approach extends easily to more general multistate models with potentially transient disease states.

We compared model predictions from the PSO approach to predictions from standard modelling approaches in survival analysis.

Results: For predictions at specific time points, models based on PSOs tend to perform better for both internal and external validation. This is particularly the case when there are non-proportional hazards that should (but are often not) be modelled. Using PSOs simplifies calculations of some standard models and validation tools.

Conclusions: If interest lies in only one or two points in time then, for prognostic models with competing risks, PSOs are easier to fit than complex survival models as we are not making strong modelling assumptions across the whole timescale.

ORAL CONTRIBUTED SESSION 10
Survival Analysis I

OC10-3 Semiparametric regression analysis of misclassified competing risk data

Theofanis Balanos, Constantin Yiannoutsos, Giorgos Bakoyannis
Indiana University, Indianapolis, United States

In the field of competing risk analysis, misclassification of the event types may lead to significantly biased estimates and invalid conclusions. In a recently published work, we addressed this issue for situations where the true event type is known for a subset of the study cohort (internal validation sample). The present research introduces a semiparametric regression analysis methodology for the case where no internal validation sample exists. Our approach leverages an estimated misclassification model (Mpofu et al., 2020) from an external population or a prior study, where a validation sub-sample is available. This model is used to estimate covariate-specific misclassification probabilities which are then incorporated in an appropriate sieve pseudo-likelihood function. This method, while similar in spirit with the approach of Ha and Tsodikov (2015), goes beyond their work in that misclassification of the event of interest is allowed to be time-dependent, and it approximates a full likelihood function instead of incorporating an ad-hoc adjustment for misclassification. The latter is shown to lead to substantially more efficient estimates through simulation experiments. We apply this methodology to real-world data obtained from large epidemiological cohorts participating in the East Africa IeDEA Consortium.

References: *[1] Ha, J., & Tsodikov, A. (2015). Semiparametric estimation in the proportional hazard model accounting for a misclassified cause of failure. Biometrics, 71(4), 941-949. https://doi.org/10.1111/biom.12338. [2] Mpofu, P. B., Bakoyannis, G., Yiannoutsos, C. T., Mwangi, A. W., & Mburu, M. (2020). A pseudo-likelihood method for estimating misclassification probabilities in competing-risks settings when true-event data are partially observed. Biometrical Journal, 62(7), 1747-1768. https://doi.org/10.1002/bimj.201900198*

OC10-4 Non-parametric estimation of net survival under dependence between death causes

Oskar Laverny[1], Nathalie Grafféo[1], Roch Giorgi[2]
1 Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, ISSPAM, Marseille, France
2 Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, ISSPAM, Hop Timone, BioSTIC, Biostatistique et Technologies de l'Information et de la Communication, Marseille, France

Introduction: Survival analysis produces valuable tools for prognosis of cancer patients. However, in population-based cancer studies, the cause of death – assumed binary, studied cancer or not – is usually unreliable or unavailable. Relative survival analysis takes this particularity into account to evaluate the excess mortality – due to cancer – with respect to population life tables. Non-parametric estimation of the net survival is possible through the Pohar Perme estimator (Pohar Perme & al (2012)), taking other causes of mortality into account. Derived similarly to Kaplan-Meier, it nevertheless relies on untestable independence assumptions between the time to death from cancer and the time to death from other causes.

Methods: We propose here to relax this untestable independence assumption using copulas (see Nelsen (2006)). If Adatorwovor & al (2023) provides parametric estimations in this generalized case, we derive here from first principles a non-parametric estimator that generalizes properly the Pohar Perme estimator which works for any dependence structure, by leveraging the underlying counting process and martingales.

Results: Our approach provides a new perspective on the Pohar Perme estimator and the acceptability of this assumption. We showcase the difference between the two estimators on population-based colorectal cancer registry data from Wolski & al (2020), and discuss potential extensions of the methodology.

Conclusion: The independence of the time to death from cancer and the time to death from other causes is a central assumption in relative survival literature, even if the epidemiological interpretation makes it hard to justify. This work, by allowing non-parametric estimation without this assumption, showcases on real data the influence of this assumption on the potential results and diagnostics.

References: *[1] Adatorwovor, R., Latouche, A., and Fine, J. P. (2023). A parametric approach to relaxing the independence assumption in relative survival analysis. In: The international journal of biostatistics, 18(2), 577-592. [2] Nelsen, R. B. (2006). An introduction to copulas. Springer. [3] Pohar Perme, M., Stare, J., and Estève, J. (2012). On estimation in relative survival. In: Biometrics, vol. 68, no 1, p. 113-120. [4] Wolski, A., Grafféo, N., Giorgi, R., and the CENSUR working survival group (2020). A permutation test based on the restricted mean survival time for comparison of net survival distributions in non-proportional excess hazard settings. In: Stat Methods Med Res, vol. 29, no 6, p. 1612-1623*

ORAL CONTRIBUTED SESSION 10
Survival Analysis I

OC10-5 Survival analysis under non-proportional hazards: Investigating non-inferiority or equivalence in time-to-event data

Kathrin Möllenhoff, Achim Tresch
Institute of Medical Statistics and Computational Biology, University of Cologne, Cologne, Germany

Background/Introduction: Time-to-event outcomes are frequently observed in medical research, for instance in the area of oncology or cardiovascular diseases. A commonly addressed issue is the comparison of a test to a reference treatment regarding survival. For this purpose, an analysis based on Kaplan-Meier curves, followed by a log-rank test, is still the most popular approach. In case of addressing non-inferiority or equivalence extensions of the log-rank test are used. Using one of these approaches, a direct interpretation is obtained by summarizing the treatment effect in one single parameter, given by the hazard ratio of the two treatments, assumed to be constant over time. However, in numerous trials hazards are non-proportional and these approaches suffer from a loss of power.

Methods: In this talk we propose a parametric framework to assess equivalence or non-inferiority for survival data. Assuming various time-to-event distributions, we first derive pointwise confidence bands for both, the hazard ratio and the difference of the survival curves. Second, we perform a test addressing non-inferiority and equivalence by directly comparing the survival functions at certain time points or over an entire time interval.

Results: We demonstrate the validity of the approach theoretically and by numerous simulations, varying the amount of censoring, the sample size and the underlying models, i.e. considering proportional and non-proportional hazard scenarios. We conclude that the method performs very well, even in settings where sample sizes are small.

Conclusion: Our suggested approach provides an alternative to extensions of the log-rank test commonly used to test for non-inferiority or equivalence. It allows for statements taking the whole observational period of a study into account, without requiring a constant hazard ratio.

ORAL CONTRIBUTED SESSION 11
Causal Inference I

OC11-1 A Bayesian multivariate factor analysis model for policy evaluation from time-series data

Pantelis Samartsidis, Shaun Seaman, Daniela De Angelis
University of Cambridge, United Kingdom

Background: The problem of estimating the effect of an intervention (or policy) from time-series observational data on multiple units arises frequently in many fields of applied research, including public health and epidemiology. Often, interest is in further studying structure in the estimated effects i.e., how these depend on a set of modifiers.

Methods: In this work, we propose a Bayesian causal factor analysis model for policy evaluation. The model includes a regression component to adjust for observed potential confounders, and its latent component can account for certain forms of unobserved confounding. Further, it can deal with outcomes of mixed type (continuous, binomial, count), increase efficiency in the estimates of the causal effects by jointly modelling multiple outcomes affected by the intervention and account for uncertainty in the number of latent factors through a hierarchical three-parameter beta prior on the loadings parameters. We extend our approach to model effect heterogeneity. Modelling effect heterogeneity is not straightforward in causal factor analysis due to weak identifiability of some of the model parameters. We circumvent this problem by using a modularization (cut posterior) approach that prevents post-intervention data from informing a subset of the model parameters. A Markov chain Monte Carlo algorithm for posterior inference is proposed, which exploits data augmentation techniques and geometric MCMC samplers to improve posterior mixing. We test our method on simulated data and use our it to evaluate the impact of local tracing partnerships (LTPs) on the various outcomes measuring the effectiveness of England's Test and Trace programme for COVID-19.

Results: We find that on average, LTPs improved case completion and timely case completion but did not influence the number of contacts elicited and contact completion. However, the effects exhibit great heterogeneity. Using our approach, we conclude that time since starting an LTP is an important driver of this heterogeneity.

Conclusion: Our method performs well on both simulated/real data, and is one of the few that can accommodate outcomes of mixed time and provide uncertainty quantification for all causal parameters of interest.

ORAL CONTRIBUTED SESSION 11
Causal Inference I

OC11-2 Quantifying the proportion of treatment effect based on joint models in tumour kinetics and survival

Georgios Kazantzidis[1], Francois Mercier[2], Ulrich Beyer[1], Virginie Rondeau[3]
1 Pharma Development, Data of Statistical Sciences, F. Hoffmann-La Roche AG, Basel, Switzerland
2 gRED, Clinical Pharmacology, Genentech, Basel, Switzerland
3 Bordeaux Population Health Research Center U1219, Inserm, Biostatistic Team, Univ. Bordeaux, Bordeaux, France

In the field of oncology clinical trials involving patients with solid tumors, the measurement of tumor size, specifically the sum of the longest diameters (SLD), plays a crucial role in assessing disease progression and treatment-related tumor shrinkage. To better understand the individual trajectory of SLD observations over time, nonlinear mixed effects models known as tumor growth inhibition (TGI) models are commonly employed. In our study, we extend the application of TGI models beyond longitudinal analysis by exploring their association with survival outcomes using joint modeling approaches. We specifically investigate the causal effects of a therapeutic intervention on survival, mediated by the biomarker SLD. Building upon the work of Le Coënt et al. [2022] and Zheng and Liu [2022], we adopt a causal inference framework to quantify the natural direct and indirect effects and estimate the proportion of treatment effects (PTE) conveyed by the biomarker. To achieve this, we implement a Bayesian approach and employ Markov chain Monte Carlo (MCMC) methods to estimate the posterior distribution of the joint model parameters. By leveraging the posterior samples, we predict both SLD values and survival outcomes. Our evaluation includes assessing the goodness of fit and predictive accuracy of the models, as well as exploring the sensitivity of our results to different prior specifications. Importantly, we extend our investigation beyond a single clinical trial and apply our TGI-OS joint model to two distinct studies involving patients with solid tumors who received immunotherapy. We analyze two randomized, open-label, international phase 3 studies including patients that had non-small-cell lung cancer and locally advanced and metastatic Colorectal Cancer. By comparing and interpreting the differences in the estimated PTE between these studies, we gain insights into the varying treatment effects mediated by the SLD biomarker in different patient populations. The implementation of the joint model is carried out using Stan, and we are committed to promoting reproducibility and transparency by sharing our code via GitHub.

References: *[1] Quentin Le Coënt, Catherine Legrand, and Virginie Rondeau. Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data. Biostatistics, 2022. [2] Cheng Zheng and Lei Liu. Quantifying direct and indirect effect for longitudinal mediator and survival outcome using joint modeling approach. Biometrics, 78(3):1233–1243, 2022.*

ORAL CONTRIBUTED SESSION 11
Causal Inference I

OC11-3 Harnessing survey data for causal inference: Generalising treatment effects in observational studies

Lizbeth Burgos-Ochoa[1], Felix J. Clouth[2]
1 Tilburg University, Tilburg, Netherlands
2 Erasmus Medical Centre Rotterdam, Rotterdam, Netherlands

Background: Survey data plays a vital role in public health and epidemiological research, despite historical challenges in deriving causal estimates from non-randomized data. However, the potential outcomes framework has provided conceptual clarity and appropriate methodological tools to address these challenges. Consequently, survey data is increasingly utilized for causal inference. While traditional survey methodology emphasizes corrections to enhance generalizability, popular causal inference approaches often overlook these adjustments. Frequently, survey weights are not included in analyses aimed at answering causal questions. These adjustments, become relevant when the intention is to generalize the treatment effect obtained from a sample to a well-defined target population. Previous research on generalizing results from randomized trials to target populations lacks extension to observational studies.1 In this project, we compare the use of various approaches to estimate a population average treatment effect (PATE) through a real-life example.

Methods: In the real-life example, our interest was in estimating the effect of a nonrandomized treatment, (residence in an urban environment) on obesity and mental health indicators. We used data from the National Longitudinal Survey of Youth 1997 (NLSY97). The NLSY97 is a representative sample of the American youth born 1980-84. Data are now available from Round 1 (1997-98) to Round 19 (2019-2020).2 We compare the use of fixed effects, g-computation, marginal structural models, and a doubly robust approach in terms of point estimates, efficiency, and flexibility for including the survey weights.

Results: We provide point estimates along with their corresponding 95% Confidence Intervals for all employed approaches. All analyses addressed potential confounding arising from non-randomized treatment assignment and accounted for the potential lack of external validity due to selection bias inherent to survey data. The results show a modest detrimental effect of living in urban environments on health outcomes.

Conclusion: Our study underscores the significance of incorporating adjustment weights in causal inference analyses using survey data when the aim is to generalise treatment effects to a well-defined population. Particularly marginal structural models and the doubly robust approach provided a simple framework to incorporate survey weights into the analysis.

References: *[1] Rudolph, K.E., Díaz, I., Rosenblum, M. and Stuart, E.A., 2014. Estimating population treatment effects from a survey subsample. American Journal of Epidemiology, 180(7), pp.737-748. [2] Moore, W., Pedlow, S., Krishnamurty, P., Wolter, K. and Chicago, I.L., 2000. National longitudinal survey of youth 1997 (NLSY97). National Opinion Research Center, Chicago, IL, 254.*

ORAL CONTRIBUTED SESSION 11
Causal Inference I

OC11-4 Comparison of causal discovery approaches to utilise temporal information for life-course Epidemiology

Claus Ekstrom
University of Copenhagen, Copenhagen, Denmark

Background/Introduction: Life-course data are common in epidemiology where individuals are followed over time, and variables have a known partial temporal ordering. The external temporal/tiered information improves the orientation of edges in causal discovery algorithms. We propose a temporal extension of the GES (Greedy Equivalence Search) algorithm for causal discovery and compare its performance, computing time and scalability in finite data settings to the TPC algorithm [1]. The methods are applied and evaluated on a dataset analysing the development of depression in Danish men. Method(s): We develop the temporal GES algorithm (TGES) for incorporating temporal information in score-based causal discovery. We compare the performance of TGES to the constraint-based TPC algorithm in two distinct ways: their ability to best recover the true Markov equivalence class of the data generating mechanism when no oracle property is available and which of the algorithms that best utilizes the temporal information to reduce the computing time.

Results: Analysis of 2928 Danish men born in 1953, followed from birth until age 65 years with information from several contacts throughout their lives. We apply the TGES and TPC algorithm to a total of 33 variables measured in 5 periods over their life course. Depression in early old age is found to be conditionally independent of all remaining variables in the data set given information about depression history in adulthood. Thus, there is no benefit in including any of the childhood information, or other variables measured in adulthood, to understand why a person develops depression in early old age. No causal effects of birth weight or birth length that span longer than youth are found. This is in contrast to myriad studies linking these factors to diabetes, death from ischemic heart disease, and mental health outcomes.

Conclusions: We developed the temporal GES algorithm to produce life-course models from observed data and compared its performance, advantages and disadvantages to the temporal PC algorithm in finite data settings. Both algorithms consider information from the whole life course jointly and allow for exploratory model building. This facilitates different approaches to building global models that can provide empirical evidence about presence or absence of causal links between exposures occurring in different periods.

References: *[1] Anne Helby Petersen, Merete Osler and Claus Thorn Ekstrøm. (2021). Data-Driven Model Building for Life-Course Epidemiology. American Journal of Epidemiology.*

OC11-5 Covariate-adjusted Robust Mixture Prior approach in clinical trials with historical controls

Andrea Callegaro
GSK, Belgium

Background/Introduction: Bayesian Dynamic Borrowing (BDB) designs to borrow historical data are being increasingly used in clinical drug development and accepted by regulatory agencies. These methods offer a robust approach to increase efficiency and strengthen evidence by integrating existing trial data into a new clinical trial.

Methods: In this work we extend a well-established approach for BDB of historical controls called Robust Mixture Prior (RMP) (Schmidli et al. 2004) by adjusting for covariates.

We adjust for covariates the historical control group and the current data, using methods from causal inference (inverse probability of treatment weighting, G-computation or double-robust estimation).

Results: Simulations demonstrate the operational characteristics of the proposed methods. Further adjusting for covariates reduces the drift between current and historical controls, with a beneficial effect on bias, control of type I error and power.

ORAL CONTRIBUTED SESSION 12
Prediction and Prognostic Models I

OC12-1 The Linearity Assumption: Unravelling its impact on prediction accuracy in multivariable models

Willi Sauerbrei, Edwin Kipruto, Jakob Moeller
Medical Center - University of Freiburg, Freiburg, Germany

Introduction: Selecting variables and identifying functional forms for continuous variables is crucial when building multivariable regression models. The intended purpose of a model is a key consideration when choosing a suitable analysis strategy. For descriptive models interpretability, transportability and general usability are important criteria. By contrast, models for prediction are often optimised for accuracy and may include more variables and allow complex functions.

The multivariable fractional polynomial (MFP) approach combines variable selection using backward elimination with a function selection procedure (FSP) for fractional polynomial functions. Its main goal is to identify influential variables and gain insight into their relationship with the outcome. With a small sample size, MFP may not be able to detect some non-linear functions, and selected models may differ substantially from the true underlying model. However, when the sample size is large, MFP can select functions and models that are similar to the underlying true model. Here, we will build prediction models using MFP approach and evaluate their performance on test data.

Methods: Using simulated data with six continuous (2 linear and 3 non-linear effects) and four categorical predictors, Sauerbrei et al. (2023) investigated the effects of sample size and model replicability on selected MFP models in eight sub-samples of a large data set. In total, 22 MFP models, all with non-linear functions, were selected. We used all the models and constructed additional ones by replacing non-linear functions with linear functions for all predictors or those with strong or weak effects. We compared the fitted values of these models using Bland-Altman plots. Additionally, with a large independent test dataset, we compared the prediction errors for all models, using the true model as a benchmark.

Results: MFP models derived from large sample sizes are similar to the true model, and the prediction errors from MFP are only slightly larger. Replacing weaker non-linear effects by linear functions does not substantially influence the prediction accuracy, but replacing stronger non-linear effects by a linear function worsens the fit and substantially increases the prediction errors.

Conclusions: Assuming linear functions for all continuous variables may not only worsen the model fit but can also lead to poor prediction accuracy on new data.

Reference: *[1] Sauerbrei, W., Kipruto, E., and Balmford, J. (2023). Effects of influential points and sample size on the selection and replicability of multivariable fractional polynomial models. Diagnostic and Prognostic, Research, 7(1), 7.*

ORAL CONTRIBUTED SESSION 12
Prediction and Prognostic Models I

OC12-2 The harms of class imbalance corrections for calibration in machine learning: A simulation study

Alex Carriero[1], Kim Luijken[1], Anne de Hond[1], Karel Moons[1], Ben Van Calster[2], Maarten Van Smeden[1]
1 University Medical Center Utrecht, Utrecht, Netherlands
2 KU Leuven, Leuven, Belgium

Introduction: Risk prediction models are increasingly used in healthcare to aid in clinical decision making. In clinical contexts, model calibration (i.e., assessing the reliability of risk estimates) is important. When a modeled event is rare, data available for model development exhibit class imbalance (i.e., individuals with vs. without the event of interest are not equally represented in the data). It is increasingly common for researchers to correct for class imbalance when building clinical prediction models, yet, the effect of such imbalance corrections on the calibration of machine learning models is largely unknown.

Methods: In our research, we studied the effect of imbalance corrections on model calibration for a variety of machine learning algorithms (logistic regression, support vector machine, random forest, XGBoost, RUSBoost, EasyEnsemble). Using extensive Monte Carlo simulations we compared the predictive performance of models trained with imbalance corrected data to models trained with data that were not corrected for imbalance across several different data-generating scenarios. Our findings were illustrated in a case study based on MIMIC-III data.

Results: In all simulation scenarios, prediction models developed without a correction for class imbalance consistently had equal or better calibration performance than prediction models developed with a correction for class imbalance. The miscalibration introduced by correcting for class imbalance was characterized by an overestimation of risk. Further, the miscalibration introduced was not always able to be corrected with recalibration.

Conclusion: For clinical prediction models which aim to produce reliable risk estimates, correcting for class imbalance may do more harm than good.

ORAL CONTRIBUTED SESSION 12
Prediction and Prognostic Models I

OC12-3 Prediction models at the crossroads of statistical inference and machine learning

Yuxuan Jin, Michael Kattan
Cleveland Clinic, Cleveland, United States

Introduction: With the abundance of electronic health records, marker, imaging and textual data, and the recent technological advancement in machine learning and AI methods, the biomedical research community has witnessed an increasing number of publications pertaining to prediction models that were developed for diagnostic and prognostic purposes. However, several methodologic issues arise when biomedical researchers employ different statistical and machine learning methods.

Methods: In this talk, we will focus on those commonly seen issues and propose easy-to-follow frameworks and feasible solutions. First, we will investigate the use of prediction models in analyzing prognostic factors and biomarkers without relying on p-values, the typical statistical inference framework. The limitations of p-values in this setting will be discussed, and a practical guide involving data visualization (such as boxplots, nonparametric smoothing line, and nomogram) and prediction performance measures (such as Area under the ROC curve, and Index of Predictive Accuracy) will be provided to better understand the role of prognostic factors and biomarkers [1]. Second, we will discuss eight common methodologic issues specific to prediction model development and evaluation, such as treating time-to-event endpoints as binary, the use of Shapley additive explanation (SHAP) values to assess variable importance and so on [2].

Results: These issues usually occur as a result of different approaches in analyzing, understanding, and interpreting prediction models from biomedical research, statistics and machine learning/AI communities. We will illustrate some of these issues using examples and showcase how those commonly used techniques, despite their popularity, may not be methodologically sound and viable. Suggestions will be given to solve those issues and references will be provided for future reading.

Conclusion: Overall, this talk underscores the importance of interdisciplinary collaboration and methodological rigor in advancing the field of biomedical prediction modeling. By bridging the gap between statistical inference and machine learning, we can harness the full potential of prediction modeling to improve diagnostic and prognostic outcomes in healthcare.

References: *[1] Jin Y, Kattan MW. Practical guide to the typical analysis of prognostic factors and biomarkers without the use of P-values. J Clin Epidemiol. 2023 Jun;158:179-184. doi: 10.1016/j.jclinepi.2023.03.025. Epub 2023 Apr 1. PMID: 37011769. 2. Jin Y, Kattan MW. Methodologic Issues Specific to Prediction Model Development and Evaluation. Chest. 2023 Jul 4:S0012-3692(23)00945-5. doi: 10.1016/j.chest.2023.06.038. Epub ahead of print. PMID: 37414333*

ORAL CONTRIBUTED SESSION 12
Prediction and Prognostic Models I

OC12-4 Assessing the impact on predictive performance of prognostic models when missing data mechanisms vary across model development phases: A simulation study

Aiden Smith[1], Paul Lambert[1,2], Mark Rutherford[1]
1 University of Leicester, Leicester, United Kingdom
2 Karolinska Institutet, Stockholm, Sweden

Background: Correct handling of missing data in prognostic models is vital to ensure their predictive accuracy when implemented in clinical practice. Current best practice dictates the use of multiple imputation or, in certain scenarios, regression imputation. These methods are effective when the missing data mechanism remains consistent across development, validation, and deployment phases, however in practice this is rarely the case.

When a prognostic model is deployed in practice, certain predictors will inherently be measured more frequently by healthcare professionals as the model dictates they are important for predictive accuracy. Therefore, both the proportion and pattern of missingness changes when compared to the development and validation data.

Failing to account for this mechanistic difference will impact the accuracy and stability of the prediction model.

Methods: We simulated data under various scenarios where missing mechanisms vary across the prognostic model development process, including zero missingness at implementation and differential missingness at implementation. We also emulate real-world scenarios where missing data mechanisms differ by location or population, as a proxy for external validation studies. We assume the missing mechanism is consistent across development and validation phases and then differs at implementation. Model performance measures (Brier score, calibration in the large, discrimination C-stat) are used to evaluate predictive performance.

Results: Initial results show that in scenarios where missing mechanisms are inconsistent across model development phases, predictive performance is reduced naive existing missing data handling methods. When no missingness is present at implementation, standard imputation methods are appropriate (assuming the imputation model is correctly specified and the data is MAR), however when differential missingness is introduced at implementation, more complex methodology is required.

Conclusion: When developing clinical prediction models to be utilized in clinical practice, significant care should be taken to evaluate the missing mechanisms across all phases of model development, and to consistently keep this under review. When data at implementation is subject to a different missing mechanism to that at development, predictive performance is reduced. Current missing data methods are not sophisticated enough to handle variation in missing mechanisms. Further methodological developments, such as model reformulation or recalibration should be investigated to better handle missingness in these scenarios.

ORAL CONTRIBUTED SESSION 12
Prediction and Prognostic Models I

OC12-5

Correcting for case-mix shift when developing clinical prediction models

StCA winner

Haya Elayan
University of Manchester, Manchester, United Kingdom

Background: When developing a clinical prediction model (CPM), a case-mix shift could occur in the development dataset where the distribution of individual predictors changes, potentially affecting the model performance. This study exploits the case-mix shift that is already observed in the development dataset to address the case-mix shift between the development and deployment phase of a CPM.

Methods: We propose two methods to correct for case-mix shift when developing CPMs by re-weighting the data samples before the case-mix shift (source set) to more closely match the samples after the case-mix shift (target set); assuming the target set reflects the target population, using a probabilistic similarity metric (membership-based method) and a distance similarity metric (distance-based method). We apply the proposed methods in a real-world dataset of myocardial infarction patients with out-of-hospital cardiac arrest within 90 days as the outcome. We design nine scenarios (including case-mix shift and no case-mix shift with a range of target/source sets sample sizes) to explore the impact on predictive performance of CPMs developed with the proposed methods in comparison to CPMs developed by either using all data samples but ignore the shift, or only using the most recent data. We report calibration and discrimination on development and 200 bootstrap samples.

Results: The proposed methods show promise in accounting for case-mix shift when developing a CPM, especially when the target set is not of sufficient sample size. In scenarios of a case-mix shift completely separating source and target sets with a target set of insufficient sample size; membership-based and distance-based models achieved optimism-adjusted calibration slope (cslope) of 0.76 and 0.69, respectively. While with a target set of sufficient sample size; membership-based and distance-based models achieved optimism-adjusted c-slope of 0.93 and 0.87, respectively. Moreover, in scenario of a case-mix shift partially separating source and target sets with a target set of insufficient sample size; membership-based and distance-based models achieved optimism-adjusted c-slope of 0.96 and 0.89, respectively.

ORAL CONTRIBUTED SESSION 13
Longitudinal Data Analysis I

OC13-1 Multivariate longitudinal modelling with non-normally distributed outcomes and endogenous covariates

Chiara Degan[1], Bart Mertens[1], Pietro Spitali[2], Erik Niks[2], Jelle Goeman[1], Roula Tsonaka[1]
1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands,
2 Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands

Introduction: Biomarkers play an essential role in clinical research, and various studies are being conducted to assess their validity. In longitudinal studies, biomarkers are endogenous time-varying variables when they depend not only on their past values but also on the history of the dependent variable. Therefore, when interested in measuring the relationship between longitudinal responses and endogenous covariates, standard Mixed Models fail and their incorrect use introduces biased estimates.

Methods: To address this issue, we propose an alternative use of two well-known multivariate models. One, which we refer to as Joint Mixed Model (Weiss, 2005, Fieuws and Verbeke, 2004-2006), induces the association via shared correlated random effects; the other, denominated as Joint Scaled Model (Rizopoulos, 2017), induces the association by copying and scaling the linear predictor of the endogenous covariate into the linear predictor of the outcome. A limitation in both cases is that the interpretation of the association is not straightforward and easy to communicate to scientists. Hence, we have analytically derived an association coefficient that measures the marginal relation between the outcome and the endogenous covariate. Fitting these models is challenging since their practical application is limited by their computational cost, which arises from high-dimensional integrations over the random effects. To fill this gap, a flexible Bayesian estimation approach, known as INLA, has been used.

Results: Both models were efficiently implemented in INLA, and the regression coefficient was successfully estimated and compared with standard mixed model marginal coefficients. We will present the results of a longitudinal study on patients with Duchenne Muscular Dystrophy, with a focus on evaluating the relationship between a bounded outcome and blood biomarkers. Based on the bayesian information criteria DIC, the joint models (JMM: -797.05, JSM: -829.94) outperformed the standard beta mixed model (-593.09).

Conclusion: These models have the advantage of being applicable in several settings, including very unbalanced datasets and with a low number of observations. They also handle missing values in both the outcome and the endogenous variable without requiring multiple imputation or deleting any information. Moreover, although we considered a continuous endogenous variable, the models are always applicable regardless of the outcome type (bounded, binary, etc.).

ORAL CONTRIBUTED SESSION 13
Longitudinal Data Analysis I

OC13-2 A Bayesian partition model for high-dimensional functional data

Marion Leila Claudine Kerioui[1*], Daniel Temko[1*], Shahin Tavakoli[2], Helene Ruffieux[1]
1 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
2 RISIS, GSEM, University of Geneva, Geneva, Switzerland
*Co-first authors

Longitudinal biomarkers are increasingly observed in health data problems and provide valuable insights into the evolution of the disease, immune system response or treatment effect. They are commonly modelled within a linear mixed-effects framework. Alternatively, Functional Principal Component Analysis (FPCA) approaches decompose the marker kinetics as a combination of latent smooth functions, possibly nonlinear in time [1]. FPCA has been extended to account for multivariate longitudinal biomarkers, where the individuals' scores corresponding to the latent functions are shared across biomarkers to borrow information [2]. Bayesian approaches have also been proposed, which circumvents costly estimations of covariance functions and allows simultaneous estimation of all parameters [3]. However, in applications involving many markers, assuming shared scores across all markers may be unreasonable, as only subsets of markers are expected to evolve in a coordinated fashion.

To describe such data of moderate to high dimensionality, we propose a Bayesian hierarchical framework that assumes a separate multivariate FPCA model on each subset of markers. The subsets are a priori unknown, and their detection is driven by the longitudinal data themselves via a mixture model. Hence, we restrict information borrowing among markers belonging to a same subset, while inferring their probability to belong to this subset. As the number of markers gets large, scalability of inference is crucial, which we achieve using a mean-field variational inference algorithm [4]. We build extensive simulations to assess our estimators' properties in various scenarios. We apply our new model to data from hospitalised SARS-CoV-2 patients [5].

The method could identify the correct subsets of variables evolving jointly, while accurately estimating the latent functions and the scores, within reasonable timescales. It also allowed reconstructing the markers kinetics in each individual subject. This model gave insights into the biological mechanisms driving disease progression and recovery, by identifying subsets of biomarkers covarying and likely contributing to common biological processes.

Our framework is applicable to any study involving many longitudinal biomarkers. In the future, it may help understand the latent mechanisms leading to patients' outcomes in other diseases.

References: *[1] J Ramsay, Encyclopedia of Statistics in Behavioral Science, 2005, 1st ed. Wiley. [2] C Happ et al, JASA, 2018, 113, no. 522: 649–59. [3] T Nolan et al, Bayesian Analysis, 2023. [4] D Blei et al, JASA, 2017, 112, no. 518: 859–77. [5] L Bergamaschi et al, Immunity, 2021; 54 (6): 1257-1275.e8*

ORAL CONTRIBUTED SESSION 13
Longitudinal Data Analysis I

OC13-3 Inference in longitudinal data analysis with terminating events

Roula Tsonaka[1], Nadine Ikelaar[2], Erik Niks[2], Menno van der Holst[3], Liesbeth De Waele[4], Marleen van den Hauwe[5], Lisa Vancampenhout[6], Liesbeth de Wreede[1]
1 Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
2 Neurology, Leiden University Medical Center, Leiden, Netherlands
3 Fysiotherapy, Leiden University Medical Center, Leiden, Netherlands
4 Development and Regeneration, Catholic University of Leuven, Leiden, Netherlands
5 Research Group for Neurorehabilitation, Catholic University of Leuven, Leuven, Belgium
6 Rehabilitation Sciences, Catholic University of Leuven, Leuven, Belgium

Introduction: In several studies, participants are followed up to monitor their functional abilities and relate them to factors such as gender, genetic markers, etc. For instance, in our motivating case study, patients diagnosed with Duchenne Muscular Dystrophy are monitored regularly on their motor ability using the North Star Ambulatory Assessment. This score takes values from 0 to 34, with higher scores indicating better motor performance. In practice, monitoring the participant's motor performance is terminated when the participant loses his ambulation.

Methods: Current approaches to model the progression of the functional score over time do not consider this complication. In particular, mixed effects models extrapolate the progression beyond the age at which the function was lost. This is characterized in the literature as the unconditional inference (Kurland and Heagerty, 2005) and describes the mean progression as if the terminating event had not occurred. In such settings, estimating the mean function for the dynamic cohort of ambulant patients is more relevant. This is characterized in the literature as partly conditional inference and explicitly considers the time when the function is lost, which might depend on the functional score. Rouanet et al, 2019 have proposed partly conditional inference for normally distributed outcome data and derived both marginal and conditional on the random effects mean progressions. However, their approximate method cannot be applied to bounded responses. Besides, deriving parameters with a valid marginal interpretation for the current ambulant patient is not possible. In this work, we propose partly conditional inference for bounded longitudinal data. In particular, we derive the mean progression given that the patient is still ambulant based on joint models of Beta mixed models and event time models. Besides, following the work of Hedeker et al, 2017 we derive both marginal and conditional on the random effects parameter estimates.

Results: We illustrate our proposal on data from our motivating study. As expected taking into account when the function is lost reflects the true variability in the data, which is not the case in the standard mixed model because data are imputed after the function is lost.

Conclusion: When the collection of longitudinal data is truncated by a terminating event, standard longitudinal analyses methods such as mixed models do not address the relevant target of inference. Our method can be used to estimate the marginal mean and marginal parameters for the dynamic cohort which has not experienced yet the terminating event.

ORAL CONTRIBUTED SESSION 13
Longitudinal Data Analysis I

OC13-4 Challenge your limits!

Davood Roshan Sangachin[1,2], **John Newell**[1,2]
1 School of Mathematical and Statistical Sciences, University of Galway, Galway, Ireland
2 CURAM, SFI Research Centre for Medical Devices, University of Galway, Galway, Ireland

Introduction: Interval estimation is a cornerstone of statistical inference. Regardless of whether an interval has been derived using a frequentist, Bayesian or computational approach the 'practical' interpretation is the same, a set of values 'likely' to contain the parameter of interest. Interval estimation in introductory statistics courses often involves frequentist approaches for one/two sample problems. A common misuse, particularly when comparing two treatments, is when a Confidence Interval is used instead of a Prediction Interval in estimating the likely improvement for a future individual receiving the treatment of interest. Prediction Intervals however are typically introduced later in the curriculum as a component of simple linear regression. Other solution to consider is to report Tolerance Intervals which aims to provide a range of values for the treatment effect for a certain percent of the population of interest. Despite their appeal, Tolerance Intervals are rarely included in introductory courses. They may feature in some medical statistics courses as population-based reference intervals to interpret a set of laboratory test results for a particular individual. In this presentation, a review and discussion of the use and misuse of confidence, prediction, and tolerance intervals and their place in the curriculum will be discussed. New approaches for generating personalised adaptive reference intervals for longitudinal monitoring will be also presented. Such adaptive reference intervals will adapt successively whenever a new measurement is recorded for an individual by accounting for both the between and within individual variability.

Methods: The methods to develop adaptive reference intervals include Bayesian framework and random intercept model. The Bayesian method estimates the intervals from posterior predictive distribution, while the random intercept model uses sufficient statistics from past data for computational efficiency.

Results: An application to real data from clinical trials and elite athlete monitoring programmes was carried out to compare the performance of different interval estimates in detecting abnormal measurements. The results showed adaptive methods are capable of triggering 'alerts' and can be used as an early warning system that warrant further attention and review.

Conclusion: Introductory courses in Statistics tend to emphasise the theoretical aspects and the computation of confidence intervals, often overlooking the practical applications and utility of other useful translatable summaries such as Prediction and Tolerance Intervals. Particularly, in the context of longitudinal monitoring, there's a need for reference intervals that are adaptive when more measurements are recorded from the same individuals over time.

ORAL CONTRIBUTED SESSION 13
Longitudinal Data Analysis I

OC13-5

StCA winner

Optimal sandwich variance estimator in penalised GEE for nearly separated longitudinal binary data with small samples

Awan Afiaz[1,2], **M. Shafiqur Rahman**[2]
1 Department of Biostatistics, University of Washington, Seattle, United States
2 Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka, Bangladesh

Background: Data separation arises in both independent and correlated binary data in biomedical studies and poses a substantial challenge that can lead to unreliable estimates and misleading inferences. This problem can occur due to a small sample size, a rare exposure or event, a very strong predictor or a linear combination of predictors, high within-subject correlation (ICC), or any combination of these issues. Penalized generalized estimating equations (GEE) have been shown to be the superior approach for handling separation in binary longitudinal data, along with bias-corrected sandwich variance estimators. Although the sandwich variance estimator is valid under misspecification of the working correlation structure in GEE, it is downward biased by design for small samples and requires large samples for the asymptotic advantages to take effect. This has led to the development of several modified robust variance estimators for GEE for small samples, which motivates finding the optimal sandwich estimator in the context of penalized GEE when there is near separation (sparsity) in the data.

Methods: The current study proposed a bias-corrected sandwich variance estimator for penalized GEE and compared its performance with ten extant sandwich estimators for nearly separated data using a simulation study. To motivate the need for an optimal sandwich estimator in penalized GEE, we demonstrated that the existing small-sample based estimators provided contradictory results when using dermatophyte-toe onychomycosis trial data. The proposed sandwich estimator does not require any additional assumptions beyond those already employed by the original sandwich estimator for GEE. We evaluated the proposed sandwich estimator by assessing the ratio of the average SEs and the empirical SD and by calculating the type-I error rates for Wald tests of the regression coefficients.

Results: Our simulation studies showed that the proposed estimator yielded nominal-level type-I error rates based on Wald tests of regression coefficients, regardless of whether the working correlation model was correctly specified. Furthermore, while existing approaches performed well when the number of subjects was high, the proposed estimator achieved nominal type-I error rates with sample sizes as low as 10, even in the most extreme scenarios.

Conclusion: Even though all existing sandwich estimators performed better as the number of subjects increased, exhibiting the usual asymptotic behavior of sandwich estimators, no other estimator uniformly achieved optimal performance faster (with respect to the number of subjects and ICC) than our proposed estimator.

ORAL CONTRIBUTED SESSION 14
Artificial Intelligence and Machine Learning

OC14-1 Continuous time Markov multistate models for health economics, with an application to artificial intelligence assisted pathology

Mark Clements, Xiaoyang Du, Mattias Rantalainen
Karolinska Institutet, Stockholm, Sweden

Introduction: Continuous time Markov multistate models (CTMMs) have found many uses in biostatistics, including models for disease prevalence and disease progression. However, most Markov models in health economics are based on discrete time, which has issues with an arbitrary ordering of processes and representing competing events and multiple events in a single time step. To address these issues, we propose using parametric CTMMs in health economics.

Methods: The proposed CTMMs are embedded within an ordinary differential equations (ODEs) framework. Kolmogorov's forward differential equations are used to model the transition probabilities. It is known that the gradients for the transition probabilities with respect to parameters can also be represented as an additional set of ODEs. As novel extensions, we show that discounted quality-adjusted life-years and costs, including both costs at transitions and costs per time in state, can be included as additional sets of ODEs. Net monetary benefit for a given cost-effectiveness threshold can then be calculated. Gradients for these quantities can be calculated using finite differences or by sets of additional ODEs. Variances for all of these quantities can be calculated using the delta method, incorporating uncertainty in the transition intensities, the health utilities and costs. These methods have been implemented for a range of parametric survival models in the rstpm2 package on the Comprehensive R Archive Network. We apply these methods to a cost-effectiveness analysis of artificial intelligence assisted breast pathology for treatment assignment. For the cost-effectiveness analysis, we assume a healthcare sector perspective, include costs for Sweden, discount at 3% per year for both quality-adjusted life-years and costs, and assume a pragmatic cost-effectiveness threshold of €100,000 per quality-adjusted life-year.

Results: We show that the treatment assignment based on artificial intelligence assisted breast pathology is cost-effective compared with standard care in Sweden. CTMMs neatly avoid many issues associated with discrete time Markov multistate models. The current implementation in R is flexible, but comparatively slow.

Conclusions: We provide a flexible framework for modelling cost-effectiveness using continuous time Markov multistate models. This framework could be extended to include (a) a cycle tree representation for chance nodes, (b) using phase-type distributions for time-in-state within a Markov framework, and (c) a re-implementation in a compiled language.

ORAL CONTRIBUTED SESSION 14
Artificial Intelligence and Machine Learning

OC14-2 UBEP Ethical Assessment Tool in clinical studies

Konstantina Thaleia Pilali, Honoria Ocagli, Assistant Corrado Lanera, Dario Gregori
University of Padova, Padova, Italy

Introduction: The integration of artificial intelligence (AI) technologies into clinical research represents a transformative shift in healthcare, offering unprecedented opportunities to enhance efficiency, accuracy, and innovation in medical practice. However, the proliferation of AI technologies raises profound ethical concerns regarding transparency, fairness, accountability, and privacy. Addressing these concerns requires systematic evaluation frameworks to assess the ethical implications of AI systems.

This work aims to propose an Ethical Assessment Tool (EAT) tailored to evaluate AI systems across key ethical dimensions.

Methods: The EAT was created as an open access tool available at UBEP Ethical Assessment Tool (unipd.it). The tool comprises distinct domains, including explainability, generalizability, open data, risk of bias, impact of wrong predictions, transparency, data privacy, reproducibility, fairness, accountability, stakeholder inclusiveness, and financial impact. The domains were derived by published guidance for statistics in clinical research. Each dimension is meticulously defined and operationalized to facilitate a thorough assessment.

Results: The tool provides according the option chosen between the various domain a radar chart with an overall score that goes from 0 to 1. The radar chart helps in condensing the information derived from each domain into a single polygon. Explainability assesses the transparency of the model's decision-making process, ranging from complete opacity to full transparency. Generalizability measures the model's performance on new data, from poor to excellent adaptation. Open data evaluates the accessibility and openness of the data used by the model, while the risk of bias assesses the fairness and impartiality of its predictions. The impact of wrong predictions gauges the consequences of model errors, from severe to negligible. Transparency evaluates the clarity and openness of the model's design and implementation. Data privacy assesses the protection of user data, from insecure to highly secure.

Reproducibility measures the ease of replicating the model's results, from challenging to straightforward. Fairness evaluates the model's impartiality across different groups, from consistently discriminatory to completely fair. Accountability assesses the mechanisms in place to ensure responsible model development and deployment. Stakeholder inclusiveness evaluates the extent to which stakeholders are involved in the model's development, from complete exclusion to full inclusiveness. Financial impact assesses the potential financial repercussions of model errors, from severe losses to negligible impact.

Conclusion: Ethical assessment is imperative to ensure the ethical integrity and societal acceptance of AI systems. The EAF offers a systematic tool to evaluate and address ethical considerations, fostering trust and ethical innovation in the AI ecosystem.

ORAL CONTRIBUTED SESSION 14
Artificial Intelligence and Machine Learning

OC14-3 Enhancing breast cancer prediction with statistical and deep learning approaches for longitudinal imaging data

Manel Rakez[1], Romain Pierlot[1], Brice Amadeo[2], Virginie Rondeau[1]
1 BIOSTAT team, Bordeaux Population Health Center, Bordeaux, France
2 EPICENE team, Bordeaux Population Health Center, Bordeaux, France

Introduction and Objective: Breast cancer is the leading cause of cancer death in women. Despite widespread mammography-based screening programs, irregular visit frequencies and intervals challenge current deep learning methodologies, originally designed for static datasets and inadequate for the dynamic nature of longitudinal imaging data. Conversely, joint models proficiently handle irregular screening visit times and intervals. This study addresses this disparity by introducing deep learning methods tailored to handle longitudinal mammograms obtained during breast cancer screening while accounting for irregular visits and time intervals. A comparative analysis is conducted against a joint model-based method.

Methods and Results: Using sequential mammography exams of 77,298 women involved in breast cancer screening in the United States, we employ a joint model to assess the longitudinal evolution of quantitative beast density metrics in relation to time-to-breast cancer occurrence, providing personalized risk estimation. Then, we explore the following deep learning methods. First, using only imaging data from women with exactly four screening visits, we use a combined convolutional and recurrent neural network model where the architecture mirrors the LRP-NET model [1], with the VGG16 replaced by a ResNet-18 to mitigate training. Second, in an effort to incorporate the entirety of available dataset information, we extend our approach in two ways. One involves allowing the model to incorporate missing visits by applying padding that introduces black images to fill the gaps. The second method utilizes a Tensorflow Wrapper – Time distributed wrapper, that duplicates convolutional layers based on the number of visits. On the other hand, a generative time-to-event model [2], integrating an Ordinary Differential Equation-based Recurrent Neural Network as an encoder, is adjusted to imaging data addressing irregular visits and spacing across women, providing a comprehensive evaluation of the temporal image features evolution and its impact on breast cancer risk prediction. Performance evaluations are conducted through a

woman-wise 10-fold cross-validation process, and all the methods, along with the joint model, are compared using the area under the operating characteristics curve.

Conclusions: This study introduces deep learning methods tailored for longitudinal imaging data with application to breast cancer screening, adeptly handling irregular visit frequencies and intervals. Through rigorous evaluations, we provide insights into the efficacy of these approaches compared to a joint model-based method and contribute to the optimization of breast cancer risk assessment in the context of accumulated imaging information.

References: *[1] S. Dadsetan et al., Pattern Recognition, 2022. [2] I. Moon et. al., Proc. Mach. Learn., 2022.*

ORAL CONTRIBUTED SESSION 14
Artificial Intelligence and Machine Learning

OC14-4 Real-time detection of Atrial fibrillation using meta-learning in the Amsterdam UMC Database

Mehran Moazeni[1], Sebastian Mildiner Moraga[1], Linda W. van Laake[2], Folkert W. Asselbergs[3,4,5], Emmeke Aarts[1]

1 Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands
2 Department of Cardiology, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands
3 Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centre, Amsterdam, Netherlands
4 Health Data Research UK, Institute of Health Informatics, University College London, London, United Kingdom
5 Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom

Introduction: Atrial Fibrillation (AF), a heart rhythm disorder that leads to an irregular and often rapid heart rate is a condition well-known to Intensive care unit (ICU) doctors, affecting up to 44% of ICU admitted patients. Given this, predicting AF early in the ICU can enable timely interventions, thereby decreasing negative hospital outcomes during a patient's stay such as prolonged length of stay (LOS) and increased mortality. Hence, we developed a continuous risk prediction model to detect AF onset using numerical ICU data.

Methods: Data from patients with a minimum duration of 36 hours were collected, including numerical values (such as heart rate and blood pressure), medications administered, and fluids injected, from the AmsterdamUMCdb, which is Europe's first open-access ICU database. A Long Short-Term Memory (LSTM) deep learning model was developed, and to ensure the model's parameters were generalizable across different patients, a Model-agnostic Meta-Learning (MAML) approach was utilized. The MAML model executed inner and outer loop updates using 2-hour data segments from each patient, covering a total continuum of 12 hours. To counteract the imbalance within the dataset, synthetic minority over-sampling techniques were used, and additional importance weights were applied to classes in the binary cross-entropy log loss function. During the testing phase, we further refined the model's performance for each test patient by applying a few inner updates. These updates were informed by identifying similar patients using embeddings in the training set, allowing us to adjust the model's weights for enhanced prediction and accuracy.

Results: The model achieved Area Under the Curve (AUC) scores of 0.73 and accuracy of 0.79. For understanding the model's decision-making process, we examined SHAP values. An analysis of the model's average risk scores 36 hours before the onset of AF showed a greater increase compared to non-AF patients. Examining the frequency of alarms per patient revealed that individuals in the AF group experienced a higher rate compared to non-AF.

Conclusion: The integration of the LSTM model within the MAML framework proved to be proficient at recognizing non-AF conditions, yet its performance in identifying AF instances was limited. The inclusion of Electrocardiogram data as supplementary input might enhance the model's capacity to detect AF with greater precision.

ORAL CONTRIBUTED SESSION 14
Artificial Intelligence and Machine Learning

OC14-5 An end-to-end modelling approach for capturing spatiotemporal patterns in two-photon imaging data

Fabian Kabus[1], Maren Hackenberg[1], Thibault Cholvin[2], Antje Kilias[2], Harald Binder[1], Marlene Bartos[2]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Germany, Freiburg im Breisgau, Germany
2 Institute for Physiology I, University of Freiburg, Medical Faculty, Freiburg, Germany, Freiburg im Brisgau, Germany

Background: In vivo two-photon calcium imaging (2PCI) is a widely used method to measure physiological activity in brain neurons. Data processing pipelines for 2PCI recordings typically rely on steps, registration, region of interest (ROI) segmentation, extraction of one temporal trace per neuron, followed by modeling of auto- and cross-correlation. Consequently, the performance of each step depends on the outcome of the previous steps. As a result manual tuning and repeated execution of each step are required, and information might get lost. Therefore an end-to-end approach that does not require segmentation and extraction might be attractive.

Methods: Instead of splitting the analysis into multiple independent steps, we develop an artificial neural network approach that operates at the level of the original temporal sequence of images. Specifically, our proposed neural network architecture comprises an outer convolutional autoencoder that learns a compressed representation of the input video. We then impose a statistical model on the resulting inner dimension-reduced representation. To obtain parameters that reflect auto- and cross-correlation, we use a vector autoregressive model (VAR) that is fit to the latent time series in each forward pass of stochastic gradient descent. We regularize the VAR parameters, based on prior knowledge, during training to encourage the emergence of structure in the latent space.

Results: We evaluate our approach on multiple 2PCI datasets with different modalities by visualizing the VAR parameters. Our proposed approach is seen to have pattern-finding capabilities across a diverse set of three different 2PCI datasets. The model captures spatiotemporal features, showcasing its adaptability to various modalities. Importantly, the integrated architecture minimizes the need for manual parameter adjustments and ROI curation, streamlining the analysis process.

Conclusion: We showcase an integrated method to characterize 2PCI recordings. By infusing knowledge-based constraints into an autoencoder with a latent VAR model, we overcome the limitations of traditional approaches. The demonstrated adaptability on diverse datasets underscores the generalizability of our approach. As we move towards a more automated methodology, our holistic approach stands out as a valuable tool for advancing in vivo two-photon calcium imaging studies.

ORAL CONTRIBUTED SESSION 15
Longitudinal Data Analysis

OC15-1 Borrowing from historical control data in a Bayesian time-to-event model with flexible baseline hazard function

Darren Scott[1], **Alex Lewin**[2]
1 Astrazeneca UK Ltd, Cambridge, United Kingdom
2 London School of Hygiene and Tropical Medicine, London, United Kingdom

There is currently a focus on statistical methods which use historical trial information to help accelerate the discovery, development, and delivery of medicine. Borrowing external information is appealing to practitioners and regulators for reasons of improved efficiency from smaller and faster trials, increased statistical power and fewer patients assigned to a less attractive treatment. Bayesian approaches naturally allow for the explicit integration of previous knowledge with new empirical data. Models can be constructed so that the borrowing is "dynamic" in the sense that the similarity of the data helps to determine how much information is used.

We are concerned with leveraging information from a historical control for time-to-event endpoints such as time-to-disease progression or time-to-death. These endpoints are the primary outcome in a variety of therapeutic areas, including oncology and cardiovascular diseases. The piecewise exponential model (PEM) is a popular model for a range of baseline hazards where the time points are fixed, and a borrowing structure is imposed for a historical dataset. Although proportional hazards are required, no parametric restriction is imposed on the baseline hazard.

Despite its convenient implementation, we demonstrate that this approach effects the borrowing capability of the model. We propose a Bayesian model with random time points, avoiding the need to specify time intervals where the baseline hazard changes, a key consideration in the PEM which impacts the variability of the baseline hazard and the amount of information borrowed from the auxiliary data. We impose a dependency between the baseline hazards through a Gaussian Markov random field prior which serves to smooth the posterior baseline hazard, improving both model estimation and borrowing characteristics as the posterior estimation is marginalised over the time intervals. By integrating over the random split points our approach allows borrowing of the baseline hazard without any constraint on the underlying shape.

We explore a variety of commensurate prior structures for the borrowing within our proposed model and assess their performance against established approaches. We evaluate the impact of the hyperparameter choice for controlling the smoothing, flexibility, and uncertainty of the baseline hazard. We demonstrate that our model leads to improved type I error in the presence of prior data conflict and increased power compared with standard PEM approaches. We have developed accompanying software which is freely available from R CRAN and enables easy implementation of the approach.

ORAL CONTRIBUTED SESSION 15
Longitudinal Data Analysis

OC15-2 Incorporating machine learning with network analysis for imbalanced multi-class classification

Nusrat Rabbee
National Institutes of Health, San Francisco, United States

Predicting modeling in imbalanced classification has garnered much attention; and many strategies have been proposed, particularly with sampling correction. However, how to integrate the elicitation of collaborative interrelationships between outcome and clinical variables and prediction of outcome using multi-level models fit for multi-class problems, using significantly imbalanced class distribution, remains an open research question. To resolve this, this talk describes a novel application of partial correlation network (PCN) to elicit multi-variable interrelationships and integrates it with a generalized additive model (GAM) for prediction of outcome. The PCN method was used to establish a network of relationships by which covariates were correlated with the multi-class outcome and to determine if the correlations were clinically sensible and consistent between different data sets. The GAM technique was then used to build a multi-level model for the multi-class problem with severely imbalanced class distribution and predict the outcome. To validate the PCN-GAM method, an independent test set was used to describe the various accuracy metrics of the classifier. With a severely imbalanced class data set – it is difficult to exceed the null classifier accuracy rate. However, the novel application of these methods (PCN to verify the biological relationships; and GAM for prediction) – the results show accuracy rates that were statistically superior to the null classifier.

Finally, this study provides insights into advancing the interdisciplinary research on biomedical data sets and prediction.

ORAL CONTRIBUTED SESSION 15
Longitudinal Data Analysis

OC15-3 Quantifying between-study heterogeneity in one-stage individual participant data meta-analyses of treatment-covariate interactions: A simulation study

Myra McGuinness[1,2], Joanne McKenzie[3], Andrew Forbes[3], Flora Hui[2], Keith Martin[2], Robert Casson[4], Amalia Karahalios[1]

1 Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia
2 Centre for Eye Research Australia, Melbourne, Australia
3 Monash University, Melbourne, Australia
4 University of Adelaide, Adelaide, Australia

Background/Introduction: It is recommended that measures of between-study effect heterogeneity be reported when conducting individual-participant data meta-analyses (IPD-MA). Methods exist to estimate $II2$ (the percentage of variation in the treatment effect due to between-study heterogeneity) when conducting two-stage IPD-MA, and when conducting one-stage IPD-MA with approximately equal numbers of treatment and control participants. However, IPD-MA may be conducted to investigate treatment-covariate interactions with unequal numbers of participants across subgroups and/or continuous covariates; here we extend formulae to estimate $II2$ in these seengs. We conduct a simulation study to assess the agreement in values of $II2$ between those derived from one-stage models via these new methods and those derived from equivalent two-stage models. An illustrative example estimating the difference in the treatment effect of an oral vitamin B3 supplement on the rate of visual field loss per year between glaucoma subtypes is provided.

Methods: Data were generated for trials with participants attending five visits over two years. The disease subtype ZZ was given to over a third of participants and active and control interventions (AA) were allocated 1:1. Scenarios differed by the magnitude of $\tau\tau 2$ (variance of the random between-study covariate-treatment effect parameter), number of trials, and residual error variance, and 1500 datasets were generated per scenario. Within each dataset a two-stage random-effects IPD-MA was fitted, with mixed-effects restricted maximum likelihood linear regression used to estimate the magnitude of the treatment-covariate interaction ($ttttttt \times AA \times ZZ$) within each trial. The analysis was then repeated using an equivalent one-stage random-effects model with residual error variances estimated separately for each trial. Inverse variance study weights were derived via these residual error variance terms, the number of observations, and variance of each of the interaction covariates within each trial. For each model, the value of $II2$ was calculated using an estimate of average within-study variance derived from inverse variance weights. The 95% limits of agreement (LOA) between these two estimates of $II2$ were calculated.

Results: Estimates of $II2$ ranged from 0.0-90.2% across scenarios. The difference in estimates of $II2$ between models was within the LOAs of -0.5 to 0.6 percentage points for 97.8% of replications, with an average difference of 0.04 percentage points. However, disparities were larger in datasets with smaller values of $II2$, with up to 5.3 percentage points difference between models.

Conclusion: The estimates of $II2$ derived from these extended methods can be interpreted similarly to those from existing formulae for two-stage models.

ORAL CONTRIBUTED SESSION 15
Longitudinal Data Analysis

OC15-4 Federated mixed effects logistic regression based on one-time shared summary statistics

Marie Analiz April Limpoco[1], **Christel Faes1**, **Niel Hens**[1,2]

1 Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute (DSI), Hasselt University, Hasselt, Belgium

2 Centre for Health Economic Research and Modelling Infectious Diseases (CHERMID), Vaccine \& Infectious Disease Institute, Antwerp University, Antwerp, Belgium

Introduction: Upholding data privacy especially in medical research has become tantamount to facing difficulties in accessing individual patient data. Federated learning has emerged as an option to preserve privacy of individual observations while still estimating a global model that can be interpreted on the individual level. It usually involves iterative communication between the data analyst and data providers. Model parameter estimates are updated either through manual transfer or through a computer network which does not require the analyst to have access to individual observations.

Methods: In this paper, we present a strategy to estimate a binary logistic regression with random intercept which requires data providers to share only summary statistics once. It involves generating pseudodata that matches the supplied summary statistics and using these into the model estimation process instead of the actual data. Our strategy is able to include multiple predictors which can be a combination of continuous and categorical variables.

Results: Through simulation, we show that our approach estimates the true model at least as good as the one which requires the pooled individual observations whenever the sample per data provider is sufficiently large (i.e. $n \geq 250$ for a single data provider ($m = 1$), $n \geq 30$ for $m = 5$, $n \geq 20$ for $m = 10$, $n \geq 10$ for $m = 15, 20$, $n \geq 5$ for $m = 50$, and $n \geq 2$ for $m = 100$). We demonstrate how it works on both publicly available and confidential data.

Conclusion: In conclusion, we are able to show that federated logistic regression with random intercept per data provider is possible with only summary statistics shared only once by the data providers to the data analyst. Unlike typical federated learning algorithms, our approach eliminates infrastructure requirements and security issues while being communication efficient and accounting for heterogeneity.

ORAL CONTRIBUTED SESSION 15
Longitudinal Data Analysis

OC15-5 Producing treatment hierarchies in network meta-analysis using probabilistic models and treatment choice criteria

Theodoros Evrenoglou[1,2], Adriani Nikolakopoulou[1], Guido Schwarzer[1], Gerta Rücker[1], Anna Chaimani[2]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg im Breisgau, Germany
2 Université Paris Cité, Center of Research in Epidemiology and Statistics, Inserm, Paris, France

Background: Network meta-analysis (NMA) allows synthesizing simultaneously the evidence on multiple treatments. A key output of NMA is the relative ranking of the treatments; nevertheless, it has attracted a lot of criticism. This is mainly due to the fact that ranking is a very influential output and, thus, prone to over-interpretations even when relative effects imply small differences between treatments. To date, common ranking methods rely on score metrics which are calculated from summary effects. Such metrics lack a straightforward interpretation, while it is still unclear how to measure their uncertainty.

Methods: We introduce a novel framework for estimating treatment hierarchies in NMA. First, we formulate a mathematical expression that defines a treatment choice criterion based on clinically important values. This criterion is applied to the relative treatment effects of included studies to produce paired data indicating treatment preferences or ties. Then, we synthesize the paired data across studies using an extension of the so-called 'Bradley-Terry model'. To parametrize the model, we assign to each treatment in the network a latent variable interpreted as the treatment's 'ability'. We estimate the ability parameters within a regression model using maximum likelihood estimation with standard errors based on the Fisher's information matrix. Higher ability estimates correspond to higher positions in the final ranking of the treatments. We further extend our model to account for study-level covariates that may affect treatment selection.

Results: We illustrate the proposed approach and compare it with common alternatives such as P-scores/SUCRAs, probability of producing the best value and PreTA ranking in two clinical datasets: a network comparing the effectiveness of 18 antidepressants for major depression and a network comparing 6 antihypertensives for the incidence of diabetes. In both examples, our approach provides a robust and easy to interpret treatment hierarchy which accounts for clinically important values and is presented alongside with uncertainty measures. In contrast, the alternative ranking metrics resulted into strict hierarchies with the tendency to elevate in the ranking list, or even place in the first position, treatments that belong to very sparse and nearly connected parts of the network.

Conclusion: The proposed framework offers a novel approach for ranking treatments in NMA based on concrete pre-defined criteria accounting for study-level covariates that may drive treatment selection. In contrast to existing approaches, it also preserves from over-interpretation of small and unimportant differences between treatments since the estimated abilities are accompanied with uncertainty measures.

ORAL CONTRIBUTED SESSION 16
Multiple Testing / Miscellaneous

OC16-1 A comparison of multivariate hypothesis tests for longitudinal outcomes

Robin Ristl, Bettina Kroyer
Medical University of Vienna, Vienna, Austria

In many randomized clinical trials, a metric outcome is measured repeatedly at subsequent pre-scheduled time-points and the aim of the study is to establish a difference in the longitudinal outcome. The trialist needs to decide whether to test a null hypothesis of no difference at a predefined single time-point or a multivariate null hypothesis of no difference at any time-point. The latter approach, when applied to clinically relevant time-points, may be preferred for reasons of robustness if the timing of a treatment effect is not well known in advance, and for reasons of efficiency if a treatment effect at more than one time-point is expected. A standard approach is to fit a mixed model for repeated measures and test a treatment by time interaction via an ANOVA-type F- or Chi-Squared test. However, depending on the correlation between repeated measures, these tests may exhibit a non-monotonic power function such that the power can decrease if the treatment effect at a specific time-point is increased. Alternative methods include a test for the maximum difference, a test for the maximum standardized difference and a test for the difference in areas under the curves.

Motivated from two studies in oncology and neurology with longitudinal outcomes, we systematically assessed the impact of different covariance structures, mean difference patterns and number of time-points on the operating characteristics of the beforementioned tests. Power calculations were performed via simulation and via numeric integration under an asymptotic normal approximation. We considered tests for a global null hypothesis of no treatment effect, as well as closed testing procedures based on the different hypothesis tests.

We further explore options to adapt the testing procedure based on interim analysis, in order to increase the overall power. A p-value combination approach is applied to control the family-wise type I error rate in presence of interim modifications.

Our results provide systematic guidance on the choice of an appropriate hypothesis test and on the respective power and sample size calculations.

ORAL CONTRIBUTED SESSION 16
Multiple Testing / Miscellaneous

OC16-2 Enhancing excellence in early phase dose-finding trials through SPIRIT and CONSORT Dose-finding Extensions (DEFINE) guidance

Christina Yap[1], on behalf of SPIRIT-DEFINE and CONSORT-DEFINE co-authors
The Institute of Cancer Research, London, United Kingdom

Background: The SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT (CONsolidated Standards Of Reporting Trials) Statements provide guidance for clinical trial protocols and reporting. Using the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) methodological framework for guideline development, SPIRIT 2013 and CONSORT 2010 have recently been extended to address the unique features of early phase dose-finding trials. These highly adaptive trials, which are typically conducted as phase I or seamless phase I/II trials, employ dose escalation/de-escalation strategies to recommend dosing regimens for subsequent trials while taking safety and parameters like pharmacokinetics, pharmacodynamics, biomarker activity, and clinical activity into account [1].

Methods and Results: Recognising the variability and incompleteness in the trial protocols and trial reports of the early phase dose-finding trials, the resulting consensus-driven SPIRIT-DEFINE and CONSORT-DEFINE guidelines recommend essential items be included in their trial protocols and trial reports [2]. The aim is to enhance greater transparency, reproducibility, and utility of results. SPIRIT-DEFINE introduced 17 new and modified 15 existing SPIRIT 2013 items [3], while CONSORT-DEFINE introduced 21 new items and modified 19 existing CONSORT items [4].

The importance of a widespread adoption of both SPIRIT and CONSORT extensions for early phase dose-finding trials is emphasized, focusing on new and modified items related to trial design, statistical methods, and analysis. The emphasis is placed on reporting details of methods for dose escalation/de-escalation strategies and decision-making criteria, as well as presenting key outcomes by dosing regimen. Exemplary cases will be discussed, and practical advice for clear reporting will be provided. In a distinctive aspect of the DEFINE development process, a toolkit for creating lay summaries of early phase dose-finding trial reports has been co-produced collaboration with patient and public partners, highlighting the essential involvement of patients' voices in reporting guidance development.

Conclusion: Collective commitment to the adoption and support of SPIRIT-DEFINE and CONSORT-DEFINE guidelines is crucial for enhancing completeness, interpretation, quality, and efficiency of peer review of early phase dose-finding trial protocols and findings. Ultimately, improvements in the design, conduct, and reporting of early phase clinical trials are expected to reduce research inefficiencies and inconsistencies, driving transformational advances in clinical care. This work is presented on behalf of the SPIRIT/CONSORT-DEFINE co-authors.

References: *[1] Solovyeva, O., et al., BMC Medicine, 2023. 21(1): p. 246. [2] Villacampa et al, eClinicalMedicine 2023, 60. [3] Yap, C., et al., BMJ, 2023. 383: p. e076386. [4] Yap, C., et al., BMJ, 2023. 383: p. e076387.*

ORAL CONTRIBUTED SESSION 16
Multiple Testing / Miscellaneous

OC16-3 A comparison of multivariate hypothesis tests for longitudinal outcomes

Lan Kelly[1], Nicole Pratt[1], Kara Cashman[2], Michelle Lorimer[2], John Farey[3], Ian Harris[4]
1 University of South Australia, Adelaide, Australia
2 South Australian Health and Medical Research Institute, Adelaide, Australia
3 University of Sydney, Sydney, Australia,
4 University of New South Wales, Sydney, Australia

Background: In cluster randomised crossover trials with two treatments, each cluster receives both treatments sequentially. The treatment sequence is determined by randomisation. Analysis must account for correlation both within clusters and between treatment periods. Cluster summary methods use the difference between cluster means as the unit of analysis, rather than person level data [1]. The absolute risk difference is estimated by regressing the cluster differences against the sequence indicator.

Methods: The advantages and disadvantages of cluster summary methods are outlined. Examples of subgroup and interim analyses which were not described in [1] are given from the CRISTAL trial [2]. Extension of the method for per protocol analysis and estimating relative risk are proposed. Relative risk estimates for a binary outcome are estimated by regressing the difference between the natural logarithm of the cluster means against sequence indicator. Analysis of the per protocol population is weighted with inverse probability weights for each cluster treatment. Probabilities are estimated at the cluster level from a hierarchical propensity model with per protocol as the outcome and person and cluster covariates.

Results: The method has been shown to produce unbiased estimates depending on the prevalence of the outcome and the intra-cluster and inter-period correlations [3]. One advantage of the analysis method is it does not explicitly require calculation of these correlation coefficients. In interim analyses, clusters will have unbalanced cluster sizes between treatments and there may be clusters which have not crossed over which requires modification of the method. Trials with rare outcomes may have cluster treatments with no outcomes, causing problems with relative risk estimates.

Conclusion: While cluster summary analysis is a simple and elegant method which does not require explicit consideration of intra-cluster and inter-period correlations, it may not be suited to trials which require interim analyses or estimating relative risks with rare outcomes. Using person level analysis to calculate inverse probability weights in per protocol analyses negates some of the benefits of the cluster summary method.

References: *[1] Turner RM et al. Statist. Med. 2007; 26:274–289. [2] Forbes AB et al. Clinical Trials 2015, 12(1):34–44. [3] Sidhu VS et al. JAMA. 2022;328(8):719-727.*

ORAL CONTRIBUTED SESSION 16
Multiple Testing / Miscellaneous

OC16-4 Multiple contrast tests under variance heteroscedasticity in general factorial designs

Matthias Becher, Frank Konietschke, Ludwig Hothorn
Charité – Universitätsmedizin Berlin, Germany

A common goal in clinical trials is to conduct tests on estimated treatment effects adjusted for covariates such as age or sex. Analysis of Covariance (ANCOVA) is often used in these scenarios to test the global null hypothesis of no treatment effect. However, in the several sample case, ANCOVA does not provide any information about individual null hypotheses and has strict assumptions such as variance homoscedasticity. If these assumptions are violated, validity of the results cannot be assured.

Konietschke et al. [1] propose an alternative to the general ANCOVA, which relaxes the assumptions and does not require variance homoscedasticity. We extend their method to a multiple contrast test procedure (MCTP), which allows us to test arbitrary linear hypotheses and provides information about the global hypothesis as well as the individual hypotheses. Further, we can calculate consistent simultaneous confidence intervals for the individual effects. To adapt the aforementioned method to MCTP we derive a small sample size approximation for the distribution of the test statistic via a multivariate t-distribution. This approximation is consistent even when group-wise variance heteroscedasticity is present. As an alternative we introduce as bootstrap based method to approximate the distribution of the test statistic.

To test the ability of our methods to control the nominal type-1 error rate we conduct an extensive simulation study and provide an example using data from a real clinical trial. The simulation results show that our methods control the nominal type-1 error rate well and consistently outperform a competitor proposed by Hothorn et al. [2]. For very small sample sizes or a higher number of groups the bootstrap method performs better than the non-resampling based methods. Still, all methods quickly converge towards the intended alpha level with increasing sample sizes.

In conclusion, conducting tests on covariate adjusted treatment effects is a common task in clinical trials. Our proposed methods offer a good alternative to classic ANCOVA without strict assumptions like variance homoscedasticity. Further, using the MCTP framework our methods provide information about both the global and the individual null hypotheses. The validity of our methods is shown in an extensive simulation study.

References: *[1] Konietschke, F., Cao, C., Gunawardana, A., & Zimmermann, G. (2021). Analysis of covariance under variance heteroscedasticity in general factorial designs. Statistics in Medicine, 40(21), 4732–4749. https://doi.org/10.1002/sim.9092. [2] Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models.. Biom J. 2008; 50(3): 346-363.*

ORAL CONTRIBUTED SESSION 16
Multiple Testing / Miscellaneous

OC16-5 Some approximations to the path formula for some non-linear models

Christiana Kartsonaki
Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

Introduction: We consider the decomposition of an association of an explanatory variable with an outcome in the presence of an intermediate variable into path-specific effects. These are sometimes referred to as direct and indirect effects. Path analysis dates back to Wright (1921), who introduced path diagrams and considered linear relationships between variables. Cochran (1938) gave a formula for such a decomposition for linear least squares regression. For models other than linear least squares regression the simple path formula does not in general hold exactly. This is related to noncollapsibility and the distinction between marginal and conditional parameters or estimands.

Methods: We derive analogous formulae for the situations where there is either a binary or a time-to-event response and either a binary or continuous intermediate variable. We describe decompositions into direct and indirect effects for logistic and proportional hazards models.

Results: For logistic regression the coefficients of the model omitting the intermediate variable are pushed closer to zero compared to the corresponding coefficients for linear regression, which are given by the original path formula. We give the term by which these coefficients are divided and describe when it is small and the formula approximates the original path formula. For Cox proportional hazards models omitting the intermediate variable gives time-dependent coefficients, and depending on the model for the association of the intermediate variable with the explanatory variable, there exist situations when the dependence on time becomes weak and the hazard takes a simple proportional form with the same modification of the regression coefficient as in a least squares analysis.

Conclusion: We have given some extensions to Cochran's formula for situations where logistic or proportional hazards models are involved. The approximations derived show how the relations involved differ from the standard path formula that holds for linear least squares regression. These approximations may be useful to assess whether in a particular setting using the simple linear regression path formula would be satisfactory. The extensions given here may be used in applications where it is of interest to understand to what extent an observed association between an explanatory variable and an outcome is explained by one or more intermediate variables, if the linear version is unlikely to be satisfactory.

Acknowledgment: I would like to thank D. R. Cox for his help with this work.

ORAL CONTRIBUTED SESSION 17
Causal Inference II

OC17-1 Sensitivity analysis of the average potential outcome with unobserved confounders

Jean-Baptiste Baitairian[1,3], Bernard Sebastien[1], Rana Jreich[1], Sandrine Katsahian[2,3], Agathe Guilloux[3]
1 Clinical Modeling and Evidence Integration, Sanofi, Gentilly, France
2 Centre d'Investigation Clinique 1418 (CIC1418) INSERM, Hôpital Européen Georges Pompidou, AP-HP, Paris, France
3 HeKA INRIA/INSERM/Université Paris Cité, PariSanté Campus, Paris, France

Estimating the effect of a continuous treatment or, more generally, a continuous intervention (e.g. concentration of a pollutant, drug concentration in plasma...) in observational data can be of great interest in many domains such as life sciences or economics. However, as opposed to binary treatments, continuous treatments suffer from less extensive research in the causal inference literature. Among others, [3] proposed ways to estimate the Average Potential Outcome (APO) (or dose-response function). Usually, the methodologies used rely on the unconfoundedness/ignorability assumption [4]. However, such an hypothesis can sometimes seem too unrealistic as observing all confounding variables is practically impossible. To overcome this issue, some recent works suggested using sensitivity models ([6], [1]) to bound the biased estimation of binary treatment effects and provide confidence intervals. [2] exploited this idea by extending the Marginal Sensitivity Model (MSM) [5] to continuous-valued interventions (CMSM) and proposed bounds for the APO in the case of high-dimensional and large sample data. Under the same sensitivity model, we derive new bounds for the APO that are tighter than [2]'s proposal by leveraging [1]'s Quantile Balancing constraint while preserving scalability. By freeing ourselves from their grid search step, we also reach lower computation times. We perform several experiments on simulated and two real datasets to validate our method in various situations and show our estimators outperform the existing ones.

References: *[1] Jacob Dorn and Kevin Guo. "Sharp sensitivity analysis for inverse propensity weighting via quantile balancing". In: Journal of the American Statistical Association (2022), pp. 1–13. [2] Andrew Jesson et al. "Scalable sensitivity and uncertainty analyses for causal-effect estimates of continuousvalued interventions". In: Advances in Neural Information Processing Systems 35 (2022), pp. 13892– 13907. [3] Nathan Kallus and Angela Zhou. "Policy evaluation and optimization with continuous treatments". In: International conference on artificial intelligence and statistics. PMLR. 2018, pp. 1243–1251. [4] Donald B Rubin. "Estimating causal effects of treatments in randomized and nonrandomized studies." In: Journal of educational Psychology 66.5 (1974), p. 688. [5] Zhiqiang Tan. "A distributional approach for causal inference using propensity scores". In: Journal of the American Statistical Association 101.476 (2006), pp. 1619–1637. [6] Qingyuan Zhao, Dylan S Small, and Bhaswar B Bhattacharya. "Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap". In: Journal of the Royal Statistical Society Series B: Statistical Methodology 81.4 (2019), pp. 735–761. 2*

ORAL CONTRIBUTED SESSION 17
Causal Inference II

OC17-2 Bridging the gap between multi-state modelling and causal inference to estimate the effect of treatment delay

Ilaria Prosepe, Saskia le Cessie, Hein Putter, Nan van Geloven
Leiden University Medical Center, Leiden, Netherlands

Background/Introduction: Physicians may opt to delay treatment initiation to first await natural recovery of the patient. The impact of specific treatment delay strategies, such as initiating treatment after waiting 1 month, is often unknown. Our goal is to estimate the marginal cumulative proportion of recovered patients under different treatment delay strategies, using observational data.

Methods: We formulate this problem as a three states multi-state model similar to the illness-death model. While multi-state models are effective for descriptive and predictive purposes, they are typically not used for causal questions. We formulate the delay strategy as setting the transition intensity for one of the transitions. We show that: (i) our quantity of interest is identifiable under the commonly used identifiability conditions of consistency, positivity, and conditional exchangeability; (ii) we propose an estimator that combines multi-state model with g-computation. We model the transitions via Cox proportional hazards models, with the covariates needed for confounding adjustment as covariates. We also provide an analytical formula for the variance for the proposed estimator.

Results: We run a simulation to numerically compare the proposed estimator to the alternative clone-censor-reweigh approach, where patients are artificially censored at the moment they deviate from the treatment strategy of interest and reweighted to correct for the subsequent dependent censoring. In the simulation we see that the proposed multi-state estimator makes more efficient use of data, as it borrows information across different treatment strategies. However, this estimator depends on correct specification of the outcome model. In situations where the time-to-treatment model is more safely assumed, clone-censor-reweight may be preferred. We showcase how the proposed methodology works on real data by estimating the effect of treatment delay on a cohort of 1896 couples with unexplained subfertility who seek artificial reproductive therapy.

Conclusions: We present formal conditions under which a multi-state model can be used to draw causal conclusions. Our proposed estimator presents many advantages: (i) it makes very efficient use of data; (ii) it can account for the continuous nature of the observed treatment delay variable in the observational data; (iii) it is easy to implement in R; (iv) it does not rely on the correct specification of the time-to-treatment model.

ORAL CONTRIBUTED SESSION 17
Causal Inference II

OC17-3 Surrogate endpoint evaluation based on causal inference and information theory in binary-continuous setting

Ariel Alonso[1], Fenny Ong[2], Florian Stijven[1], Wim Van der Elst[3], Geert Molenberghs[1,2], Ingrid Van Keilegom[1], Geert Verbeke[1], Andrea Callegaro[4]
1 KU Leuven, Belgium
2 Hasselt University, Belgium
3 Janssen Pharmaceutical, Belgium
4 GSK Vaccines, Belgium

Background: The development of methods to validate surrogate endpoints remains an active and intensely researched area due to its importance in expediting the process of clinical trials of a large number of new promising treatments. A general definition and quantification of surrogacy based on the concept of information theory has been introduced to provide a unified framework for such evaluation. In the setting where both surrogate and true endpoints are continuous or binary, a metric of surrogacy to assess the definition is developed. The current study aims to extend the evaluation to the setting in which the true and surrogate endpoints are binary and continuous, respectively.

Methods: Based on the concept of causal inference, a new model is proposed to describe the joint distribution of the potential outcomes associated with the putative surrogate and the true endpoint of interest. The identifiability issue inherent to this type of model is handled via sensitivity analysis. In the next step, using the information theory, a metric of surrogacy called individual causal association (ICA) is formulated. The methodology is evaluated via simulations and is further implemented using the data from a randomized clinical trial of an inactivated quadrivalent influenza vaccine.

Results: The simulation studies support the performance of the proposed method as a sensitivity analysis in the evaluation of surrogate endpoints. The analysis of the case study indicates some level of quantitative support in using the measurement of hemagglutination inhibition antibody titers as a surrogate for the protection effect of the influenza vaccine.

Conclusion: The proposed metric ICA has convenient theoretical properties and offers a simple yet intuitive interpretation to quantify the association between the individual causal treatment effect on the surrogate and the true endpoint.

ORAL CONTRIBUTED SESSION 17
Causal Inference II

OC17-4 The parametric G-formula for latent Markov models

Felix Clouth[1], Maarten Bijlsma[2], Steffen Pauws[1], Jeroen Vermunt[1]
1 Tilburg University, Netherlands
2 University of Groningen, Netherlands

Background: In the a situation where exposure status, covariates, and outcome variables vary over time, intermediate confounding can occur. For instance, exposure at baseline might affect some covariates at follow-up that confound the exposure-outcome relationship at follow-up. Furthermore, if outcome variables are assessed repeatedly, outcome variables might affect covariates and exposure status at follow-up confounding the exposure-outcome relationship at follow-up. To estimate the ATE of alternative exposure trajectories on the outcome, the parametric g-formula has been proposed. [1] However, the parametric g-formula cannot be directly applied when the outcome variable is unobservable, i.e., latent, and measured through several indicator variables. In this talk, I will outline how the parametric g-formula can be modified to accommodate latent outcomes.

Methods: We propose a step-wise approach for incorporating latent Markov models in the parametric g-formula. In the first step, a measurement model is estimated based on the indicator variables and individuals are classified at every timepoint. Next, structural models are estimated imposing a Markov constraint for the latent outcome. Separate models are estimated, one for the outcome, one for exposure status, and one each for every confounder. For these models, misclassifications occurring in the first step are corrected for. In the last step, micro simulations are performed as usual in the parametric g-formula.

Results: We illustrated our approach using LISS panel data to estimate the ATE of mental health status under two hypothetical scenarios, always employed and never employed during the study period. We estimate a small but significant ATE for the worst mental health state as identified by the measurement model. Furthermore, we show that the empirical distributions of all involved variables are appropriately approximated by our models under a natural course scenario.

Conclusion: We show that using a step-wise approach utilizing latent Markov models, the g-formula can be modified to accommodate latent outcomes.

References: *[1] Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. Mathematical Modelling, 7(9-12), 1393-1512.*

ORAL CONTRIBUTED SESSION 17
Causal Inference II

OC17-5 A new estimator and Stata command for estimating causal mediation effects with non-adherence in the presence of missing data

Anca Chis Ster, Sabine Landau, Richard Emsley
King's College London, United Kingdom

Introduction: and Objective(s): Many clinical trials report participant non-adherence. A standard intention-to-treat (ITT) analysis will estimate the causal effect of treatment offer without bias, though ignores the impact of non-adherence. To account for non-adherence, one can estimate a complier-average causal effect (CACE), the average causal effect of treatment receipt in the subgroup of participants who would comply with their randomisation. Evaluating how interventions lead to changes in the outcome (the mechanism) is key for the development of more effective interventions. A mediation analysis aims to decompose a total treatment effect into an indirect effect, one that operates via changing the mediator, and a direct effect. However, current methods for mediation analysis in clinical trials focus on decomposing the ITT effect, where the corresponding effects explain the mechanisms of treatment effectiveness. Finally, the reliability and interpretability of clinical trial results are affected by missing data in the mediator and outcome. It is crucial that missing data are addressed appropriately to avoid bias in study findings.

Method(s) and Results: Previous work has shown that the CACE can be decomposed into a direct effect, the Complier Average Natural Direct Effect (CANDE), and a mediated effect, the Complier Average Causal Mediated Effect (CACME). We propose a new estimator, based on Instrumental Variables (IVs) and Structural Equation Models (SEMs) for estimating the CACME and CANDE. A Monte Carlo simulation study following the ADEMP framework is conducted to evaluate the bias in CACE, CACME, and CANDE when there are missing data in the mediator and/or outcome. We construct three scenarios where the missing data are MCAR, six MAR scenarios, and six MNAR scenarios, to cover a range of realistic trial scenarios. We vary 8 parameters, including the trial size, the proportion of non-adherence, and the proportion of missing data. The simulation findings show that the linear IV-SEM estimator can obtain unbiased effects of the CACME and CANDE under MCAR and MAR missing data mechanisms. A new Stata command, compmed, is developed for practical implementation of the proposed estimator.

Conclusions: A new estimator, based on IVs and SEMs, is proposed for evaluating mechanisms of treatment efficacy. Simulations have demonstrated that the estimator can obtain unbiased estimates of the CACME, CANDE, and CACE when missing data are MCAR or MAR. A new Stata command, compmed, offers users a quick and convenient package for obtaining estimates of the CACME, CANDE, and CACE.

ORAL CONTRIBUTED SESSION 18
Methods for High Dimensional Data

OC18-1 Bayesian dynamic borrowing using the mixture prior: Impact of parameter choices

Vivienn Weru, Silvia Calderazzo, Manuel Wiesenfarth, Annette Kopp-Schneider
Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

Introduction: In clinical trials with small populations it is sometimes desired to borrow information from external sources to increase efficiency of the current study. For example, the current standard arm may already have been investigated in a previous trial with a similar setting or a pediatric trial may borrow information from an adult trial investigating the same condition.

Methods: Bayesian methods have seen widespread application in this context, due to the ease in incorporating external information via the specification of informative prior distributions. One such approach is the robust mixture prior (Schmidli et al, 2014) arising as a weighted mixture of an informative prior and a robust prior inducing dynamic borrowing that allows to borrow most when the current and external data are observed to be similar. The mixture prior requires the choice of three quantities: the mixture weight, and the mean and dispersion of the robust component. Some general guidance is available, but a case-by-case study of the impact of these quantities on specific operating characteristics seems lacking. For normal endpoints, a standard recommendation for the robust component is a unit-information prior, i.e. a prior with variance equal to that of one observation. With regards to the location of this component, we consider two scenarios: same location as the informative component or at a null hypothesis value in the case of testing. We assess impact of these parameters on frequentist operating characteristics (Type 1 error rate/power, MSE) as well as bimodality of the posterior distribution.

Results: The results show that all three quantities can strongly impact the operating characteristics. In particular, as already known, variance of the robust component is linked to robustness. Less known, however, is that location can have a strong impact on type I error rate which can even become unbounded. Further, the impact of the weight choice is strongly linked with the robust component's location and variance.

Conclusion: Parameters of the robustifying component in the mixture prior need careful thought. In addition to unwanted frequentist operating characteristics, some parameter choices could result in a bimodal posterior distribution which could lead to disjoint Highest Posterior Density Intervals thereby complicating decision making.

References: *[1] Schmidli, H., et al. (2014). Biometrics, 70(4), 1023-1032.*

ORAL CONTRIBUTED SESSION 18
Methods for High Dimensional Data

OC18-2 Prediction powered inference for clinical trials

Pierre-Emmanuel Poulet[1], Bruno Jedynak[2], Stanley Durrleman[1]
1 Institut du Cerveau - Paris Brain Institute - ICM, Inria, CNRS, Inserm, Sorbonne Université, AP-HP, Hôpitalde la Pitié Salpêtrière, Paris, France
2 Portland State University, Portland, United States

Introduction: The past decade has seen an increased interest in enhancing clinical trials by leveraging Machine Learning models trained on observational cohorts or on previous clinical trial's data. Currently the use of a system's predictions as a prognostic factor in an ANCOVA is one way to improve the power of a clinical trial which has the regulatory green light. We introduce a new method to estimate treatment effect based on the Prediction Powered Inference (PPI) and PPI++ framework [2, 1].

Method: We propose a new estimator for the average treatment effect (ATE) using predictions based on baseline covariates. We show that this new estimator is unbiased and has a smaller confidence interval than the classical ATE estimator. We derive sample size reduction for an equally powered clinical trial. Moreover, the PPI method is generic and allows us to derive formulas that improve on other estimators for various quantities of interest. We specifically showcase prediction-powered estimators for Kaplan-Meier curves and linear systems.

Results: We validate the consistency and robustness of this new estimator through simulations. Our results on synthetic data demonstrate that our estimator has equivalent performances as a prognostic adjusted analysis, while being more straightforward to derive. We show that under mild hypotheses we reach similar power as a standard two-arm randomized clinical trial with a (1 − R2) smaller sample size, where R2 is the coefficient of determination. We also provide simulations for survival data. Finally we apply our method to a clinical trial in Alzheimer's disease and obtain the same results with less participants.

Conclusion: Prediction Powered Inference thus allows us to build a new class of unbiased estimators, which can lead to a reduction of sample size in clinical trials.

References: *[1] Anastasios N. Angelopoulos, John C. Duchi, and Tijana Zrnic. PPI++: Efficient Prediction-Powered Inference. arXiv:2311.01453 [cs, stat]. Nov. 2023. [2] Anastasios N. Angelopoulos et al. "Prediction-powered inference". In: Science 382.6671 (Nov. 2023). Publisher: American Association for the Advancement of Science, pp. 669–674. 1*

ORAL CONTRIBUTED SESSION 18
Methods for High Dimensional Data

OC18-3 Targeted maximum likelihood estimation for clinical trials with survival outcome

Aris Perperoglou, Tom Taverner
GSK, London, United Kingdom

Background: Targeted Maximum Likelihood Estimation (TMLE) provides a comprehensive methodology for the estimation of causal parameters, offering a double robust approach and a flexible framework to combine machine learning with traditional statistical theory and inference. Mainly used in observational studies, TMLE is also increasingly gaining interest for its potential to enhance efficiency in clinical trials analysis. However, the application of TMLE in survival studies presents challenges, partly because attributing causal interpretation to survival estimands, like the hazard ratio, may be misleading. [1]

Methods: Rytgaard and van der Laan [2] demonstrated that TMLE can be adapted for survival data, with a focus on studies with extended follow-up periods of more than 5 years. Although clinical trials are generally shorter and may benefit less from that framework, there are examples of cohort trials where the study design might result in non-simultaneous randomisation and, consequently, possible confounding with cohort waves. For those, indirect treatment comparisons might be needed that rely on defining propensity scores and adjusting analysis accordingly. Here, we employ TMLE for indirect comparisons in such settings. A complexity that needs to discussed is the fact that TMLE targets parameters that might be more difficult to explain to a non-statistical audience. Our focus lies in demonstrating feasibility and efficacy of TMLE within these constraints.

Results: We present findings from simulations and a cohort trial to benchmark TMLE against other methodologies, with a special focus on interpretability within clinical trial settings. Our examples highlight the practical applications of TMLE in addressing the challenges posed by cohort randomised designs and the potential for indirect comparisons. The results underscore the adaptability of TMLE to complex survival data and its comparative effectiveness and interpretability alongside traditional statistical methods.

Discussion: While exploring novel statistical methods like TMLE in clinical settings holds promise, successful implementation requires overcoming crucial hurdles. Regulatory bodies need to adapt their guidelines to accommodate these methods, and clinicians must fully understand their principles and limitations. A major challenge lies in translating statistically robust estimands into metrics readily interpretable and meaningful for clinical decision-making. Our work addresses this challenge by showcasing how TMLE, within specific contexts, can deliver results comparable to existing methods, potentially easing the transition for clinical application.

References: *[1] Hernán, M. A. (2010). The hazards of hazard ratios. Epidemiology (Cambridge, Mass.), 21(1), 13.2. [2] Rytgaard, H. C., & van der Laan, M. J. (2024). Targeted maximum likelihood estimation for causal inference in survival and competing risks analysis. Lifetime Data Analysis, 30(1), 4-33*

ORAL CONTRIBUTED SESSION 18
Methods for High Dimensional Data

OC18-4 Sensitivity analysis for unmeasured confounding in indirect treatment comparisons

Shijie (Kate) Ren[1], Sarah Ren[1], Nicky Welton[2], Mark Strong[1]
1 University of Sheffield, United Kingdom
2 University of Bristol, United Kingdom

Background: Health care decision making requires reliable estimates of treatment effects and considers randomised controlled trials (RCTs) as the gold standard for providing direct evidence on the treatments of interest. However, RCTs comparing treatments of interest are not always available. In practice, indirect treatment comparisons are often required to compare evidence from different trials. Population-adjusted indirect comparison methods such as matching-adjusted indirect comparisons (MAIC) and simulated treatment comparisons (STC) are useful tools to correct trial population differences in indirect treatment comparisons. In the case of no common comparators, unanchored MAIC and STC assume that all effect modifiers and prognostic factors are accounted for, which is largely considered impossible to meet. The aim is to address the limitation of the current population adjustment methods where certain prognostic factor and/or effect modifiers are not reported in the comparator trial.

Methods: We propose to formally quantify the unmeasured confounding in indirect treatment comparisons using a sensitivity analysis approach based on STC. STC is an outcome regression approach, where a statistical model is fitted using the individual patient-level data, and then the fitted model is used to predict the outcomes that would have been observed in the comparator population. In our proposed approach, we treat the mean of the marginal distribution for the unmeasured confounders as sensitivity/bias parameters. We sample individual covariates using the NORTA algorithm given the observed covariates and assumed either fixed values or distributions for the sensitivity/bias parameters for the comparator population. The unanchored STC is then implemented using standardisation/marginalisation with sampled individual covariates to ensure the derivation of a suitable marginal treatment effect without aggregation bias for health technology assessment evaluations.

Results: The STC-based sensitivity approach explores the sensitivity of indirect treatment effect estimates to the presence of unmeasured confounder(s) by varying the values for the sensitivity/bias parameters. This approach can either be deterministic or probabilistic. It is easy to implement regardless of the dimension and data type of unmeasured confounders.

Conclusions: The proposed sensitivity approach provides a formal quantitative assessment of the bias associated with not adjusting for certain important covariates that are missing in the comparator's trial. This could help to address the major limitation of the population adjustments methods and provide a more reliable and more robust estimate of treatment effects for unanchored treatment comparisons.

ORAL CONTRIBUTED SESSION 18
Methods for High Dimensional Data

OC18-5 New clinical trial design borrowing information across patient subgroups based on fusion-penalized regression models

Marion Leila Claudine Kerioui[1,2], Alexia Iasonos[1], Mithat Gonen[1], Andrea Arfe[1]
1 Departement of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, United States
2 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom

In oncology, basket clinical trials aim to assess the efficacy of a drug in several subgroups of patients, called baskets, with different tumour types. In this context, information borrowing strategies may increase the probability to detect drug efficacy in active baskets, by shrinking together the estimates of the parameters characterizing the drug effect in baskets with similar drug activity. Existing designs typically rely on Bayesian hierarchical structures to perform information borrowing, either across all baskets (Bayesian Hierarchical Models, BHM) or groups of baskets (Bayesian Model Averaging, BMA).

Here we propose to use fusion-penalized regression model to borrow information in the design of phase 2 single-arm basket trials with a binary outcome. In these studies, fusion penalties enforce similar estimated probabilities of response to treatment in baskets with similar drug activity, by shrinking towards zero their pairwise differences. We carried out intensive simulations to assess the performances of the proposed design for the LASSO, Ridge and Elastic-Net penalties. The BHM and BMA were used as comparators, as well as two common approaches whereby all baskets are either pooled together or considered separately. We computed the probabilities to reject the null hypothesis, i.e. to declare drug activity, in active (marginal power) and inactive baskets (type I error). All designs were calibrated via simulations by minimizing the Mean Square Error under given constraints to keep the probability to conclude to drug activity in at least one basket below 10% when the drug was truly inactive in all baskets.

Overall, the designs relying on fusion-penalties performed similarly to the Bayesian designs, increasing the marginal power in active baskets compared to the design assuming baskets independence. As expected, all designs relying on information borrowing also lead to type I error inflation. Interestingly, when simulating trials with a realistic pattern of heterogeneous response across baskets, our proposed designs outperformed the Bayesian designs, with a controlled type I error (less than 4%) and the highest marginal power (up to 88%). They also performed well when active baskets were much smaller than the inactive ones (respectively 8 and 16 patients in average), with a controlled type I error and a 71% marginal power against 62 and 64% for the BHM and BMA designs respectively.

As the planning of a clinical trial requires careful sensitivity analysis through intensive simulation studies, our designs constitute an interesting alternative to Bayesian designs thanks to its easy implementation and fast computation.

ORAL CONTRIBUTED SESSION 19
Imperfect Data

OC19-1 Pseudo-observations for censored covariates

INVITED

Rebecca Betensky
New York University, NY, United States

Censored covariates are a common problem in regression analyses. These may include age of onset or a biomarker measurement, for example. We present the use of pseudo-observations to stand in for the censored covariate as an alternative to imputation or thresholding based approaches. The advantage of this approach is that observations are not deleted from the analysis. We consider two methods: one based on thresholding and one based on the restricted mean survival time. We explore the properties in simulation studies and in application to an Alzheimer's disease study with maternal age of onset as the covariate.

OC19-2 Estimators under informative covariate censoring: An application to Huntington's disease

INVITED

Jesus Vazquez[1], Marissa Ashner[2], Karen Mardner[3], Yanyuan Ma[4], Tanya Garcia[1]
1 University of North Carolina at Chapel Hill, Chapel Hill, United States
2 Duke University, Durham, United States
3 Columbia University, New York, United States
4 Penn State University, State College, United States

Background/Introduction: Cognitive dysfunction is a symptom of Huntington's disease (HD) and can serve as an early marker for evaluating treatments intended at delaying motor skill decline. Understanding cognitive dysfunction as a function of age at clinical diagnosis is a prerequisite for their use in clinical trials, but this remains challenging because patients drop out or the study ends, resulting in a censored value for age at clinical diagnosis. While right-censored time-to-event outcomes have been studied for decades, handling time-to-event covariates, also known as censored covariates, is now of growing interest. So far, the literature has assumed that the covariate censoring is noninformative, i.e., event time is independent of censoring time, and has yet to establish methods when covariate censoring is in fact informative.

Methods/Results: We present five methods capable of handling both informative and noninformative covariate censoring, namely a complete case, inverse probability weighting (IPW), augmented complete case, augmented IPW, and maximum likelihood estimators. We establish the robustness properties of these estimators in terms of when these estimators remain consistent despite incorrect distributional assumptions. We additionally present a hypothesis test that can be used in practice to check if the covariate censoring is truly informative or noninformative. All methods are applied to ENROLL-HD, a non-interventional observational study of HD. We compare the effect estimates and variances of each estimator.

Conclusion: We establish a framework of regression estimators with a covariate subject informative censoring with varying degrees of robustness and efficiency. A major feature of our work is establishing the conditions under which these estimators maintain consistency even if our assumptions regarding informative or non-informative covariate censoring are incorrect. We also establish asymptotic properties for all estimators and provide a straightforward hypothesis test to help determine whether covariate censoring is informative or not.

ORAL CONTRIBUTED SESSION 19
Imperfect Data

OC19-3

Regression and goodness-of-fit with interval-censored covariates

INVITED

Guadalupe Gómez-Melis, Klaus Langohr, Andrea Toloba López-Egea
Universitat Politècnica de Catalunya, Barcelona, Spain

Introduction: Interval-censored observations of a response variable are a prevalent feature in medical studies, often occurring when the response entails the duration until an event, which is periodically monitored. Another scenario arises when the explanatory variables are interval-censored while the response variable remains fully observed. We encounter this situation when analysing the association between plasma carotenoid concentrations and anthropometric, clinical, and biochemical parameters indicative of cardiovascular risk. The quantification of plasma carotenoid concentrations involves chromatography coupled with mass spectrometry and is subject to detection and quantification limits due to their minimal presence in the body. The sum of these compounds, of particular interest, is an interval-censored variable with a lower and upper limit given by the sum of the respective lower and upper limits of all the components.

Methods: In this work we extend the GEL (Gómez, Espinal and Lagakos) method in three key ways: first, by eliminating the discrete assumption regarding the support of the interval-censored variable; second, by employing a Generalized Linear Model accommodating responses from the exponential family to address non-normal distributions; and third, by introducing novel residuals to assess the model's validity.

Results: The PREDIMED-plus trial examines the association between plasma carotenoid concentrations and anthropometric, clinical, and biochemical parameters linked to cardiovascular risk. Specifically, we investigate the association between total carotenoid concentration and glucose levels, as well as obesity, while considering the interval-censored nature of the covariates. We utilise gamma regression and logistic models to analyse these associations. The sample data shows that total carotenoid concentration is inversely related both to glucose levels and the risk of obesity.

Conclusion: The increasing prevalence of biomedical issues requiring statistical approaches for handling censored covariates (specifically right- or interval-censored) highlights a growing need. Despite being in its early stages, our proposed general method offers a dependable solution for regression models with interval-censored covariates.

References: *[1] Gómez, Espinal and Lagakos (2003). Inference for a linear regression model with an interval-censored covariate. Statistics in Medicine 22(3):409-25. [2] Langohr and Gómez Melis (2014). Estimation and residual analysis with R for a linear regression model with an interval-censored covariate. Biometrical Journal 56(5):867-85. [3] Gómez Melis, Marhuenda-Muñoz and Langohr (2022). Regression Analysis with Interval-Censored Covariates. Application to Liquid Chromatography in Emerging Topics in Modeling Interval-Censored Survival Data, Pages 271-294. Ed: Sun and Chen. Springer. [4] Code https://github.com/klongear/ICbook*

ORAL CONTRIBUTED SESSION 19
Imperfect Data

OC19-4 Variable selection for survival data with multiply imputed covariates: A comparative simulation study

Vasiliki (Valia) Baralou, Giota Touloumi
Department of Hygiene, Epidemiology & Medical Statistics, Medical School, National & Kapodistrian University of Athens, Athens, Greece

Background/Introduction: Big data in healthcare pose new statistical challenges such as variable selection in the presence of missing values which are now ubiquitous in large databases. We aimed to compare through simulations the performance of different methods for variable selection on survival data with multiply imputed covariates.

Methods: Simulations were based on real data from people who inject drugs provided by the Greek Organisation Against Drugs to estimate individual risk of HIV, HBV or HCV infection. Baseline covariates included demographic, socio-economic, behavioral and laboratory data. Six scenarios were generated assuming different number of covariates (15 or 40; only 7 were informative), associations between the covariates and the outcome (linear, non-linear and interactions) and proportions of missing values per variable (0.1 or 0.2). All scenarios assumed 50% censoring, decreasing hazard, sample size N=2000 and missing completely at random mechanism. We handled missingness with multiple imputation by chained equations. Variable selection on survival data with fully observed covariates was performed using variable importance (VIMP) estimated by Random Survival Forest (RSF), backward elimination and least absolute shrinkage and selection operator (lasso) in Cox regression. To conduct variable selection with multiply imputed data, we applied: i) the prementioned methods on each imputed dataset and defined the final set of selected predictors as those selected in at least half of datasets and ii) the adjusted VIMP by RSF that accounts for both the between- and within-imputation variance in the VIMP scores. Performance was assessed by sensitivity, specificity, false positive rate and precision. C-index was used as evaluation metric on real data.

Results: When applied on complete data, backward elimination with the Cox model gave the best balance among all metrics in scenarios including 15 covariates whereas in those with 40 covariates, RSF ranked first. When applied on multiply imputed data, RSF performed at least similarly with regression-based approaches in all settings. In comparison with the approach based on selection frequency by VIMP, the adjusted VIMP increased sensitivity by approximately 25% when assuming linearity and including 40 covariates but did not improve VIMP's performance in the remaining settings. Methods performed similarly across settings with different proportions of missing data. In all scenarios, lasso gave the lowest sensitivity but perfect specificity values. When applied on real data, methods' prognostic performance was similar.

Conclusion: Regression-based approaches remained robust when deviating from linearity and including lower number of covariates. RSF seems preferable for variable selection as data dimensionality increases.

ORAL CONTRIBUTED SESSION 19
Imperfect Data

OC19-5 Incomplete reporting of incomplete data: Findings and recommendations from a scoping review

Rheanna Mainzer[1,2], A Margarita Moreno-Betancur[1,2], Cattram Nguyen[1,2], Julie Simpson[2,3], John Carlin[1,2], Katherine Lee[1,2]

1 Murdoch Children's Research Institute, Parkville, Australia
2 The University of Melbourne, Parkville, Australia
3 University of Oxford, Oxford, United Kingdom

Background: Missing data are common in observational studies and often occur in multiple variables required for estimation of a causal effect, such as the exposure, outcome and/or the variables that are used to control for confounding. Analyses involving multiple incompletely observed variables are not as straightforward as analyses with just one incomplete variable. For example, the standard missing data assumptions ("missing completely at random", "missing at random" [MAR] or "missing not at random") are difficult to interpret and assess in the context of multivariable missingness. It is not clear how the complexities that arise due to multivariable missingness are being addressed in practice. Therefore, the aim of this study was to review how missing data are handled in observational studies that use multiple imputation (MI) for causal effect estimation, with a particular focus on four key areas for reporting: missing data summaries, missing data assumptions, MI implementation, and sensitivity analyses to quantify uncertainty due to missing data assumptions.

Methods We searched five top general epidemiology journals for observational studies that aimed to answer at least one causal research question using MI, published between January 2019 and December 2021. Article screening and data extraction were performed systematically.

Results: Of the 130 studies included in this review, 108 (83%) derived an analysis sample by excluding individuals with missing data and 114 (88%) had multivariable missingness within the analysis sample. Forty-four (34%) studies provided a statement about missing data assumptions, 34 of which assumed that data were MAR, and just 11/44 (25%) studies provided a justification for these assumptions. MI was used for the primary analysis in 102/130 (78%) studies. The number of imputations, MI method and MI software were reported for 71%, 75% and 88% of studies, respectively, while details about the variables that were included in the imputation model(s) were reported 55-58% of the time. There were 67/130 (52%) studies that conducted a secondary analysis using an approach that handled the missing data in a different way from the primary analysis, with 56/130 (43%) studies using both MI and a complete case analysis for estimation. Clear justification for the secondary analysis was lacking in 99% of studies, and no study conducted a sensitivity analysis that included external information or additional assumptions about how the distributions of the missing and observed data differed.

Conclusion: Effort is needed to improve the rigour and reporting of causal analyses of observational data with multivariable missingness.

ORAL CONTRIBUTED SESSION 20
Clinical Trials III

OC20-1 Bootstrap confidence intervals in two-stage adaptive enrichment designs

Enyu Li[1], Nigel Stallard[1], Ekkehard Glimm[2], Dominic Magirr[2], Peter Kimani[1]
1 University of Warwick, Coventry, United Kingdom
2 Novartis Pharma AG, Basel, Switzerland

Background: With the advances in personalized medicine, two-stage adaptive enrichment designs for clinical trials assessing targeted therapies have received increased attention. In this design, the subpopulation that is likely to benefit from the treatment is identified based on the data from the first stage. In the second stage, the recruitment is restricted to the selected subpopulation. Data from both stages are then used to perform the confirmatory analysis for the treatment effect in the selected population. This "select and test" structure of adaptive enrichment designs leads to statistical challenges in hypothesis testing, point estimation and confidence interval calculation. Several methods concerning hypothesis testing and point estimation have been proposed. However, the question of constructing confidence intervals (CIs) has drawn less attention. In this work, we develop CIs that adjust for trial adaptation.

Method: We consider the adaptive enrichment design proposed by Jenkins et al. [1], which is realistic in practice, and the case when the outcome is normally distributed. Our approach combines the uniformly minimum variance conditional unbiased estimator (UMVCUE) and bootstrap sampling procedures for diagnostic tests proposed by Pepe et al. [2], which simulate the adaptivity of the trial, to construct bias-adjusted confidence intervals for the treatment effect in the selected population.

Result: We demonstrate that confidence intervals generated by our bootstrap algorithm achieve coverage probability very close to nominal level and are approximately symmetric with probability above and below true effect approximately equal. Using these metrics, it outperforms existing methods, including double bootstrap confidence intervals and duality confidence intervals.

Conclusion: The new interval estimation method shows a superior capability to correct selection bias and achieve nominal coverage probability over existing methods. It can be extended to other adaptive designs with binary or time-to-event endpoints when the distribution of the effect size estimate is asymptotically normal.

References: *[1] Jenkins, M., Stone, A. & Jennison, C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical statistics 10, 347–356 (2011) [2] Pepe, M. S., Feng, Z., Longton, G. & Koopmeiners, J. Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. Statistics in Medicine 28, 762–779 (2009)*

ORAL CONTRIBUTED SESSION 20
Clinical Trials III

OC20-2 Incomplete reporting of incomplete data: Findings and recommendations from a scoping review

Babak Choodari-Oskooei, Daniel Bratton, Mahesh KB Parmar
UCL, London, United Kingdom

Background: Multi-arm multi-stage (MAMS) randomised trial designs are an efficient, adaptive approach for testing many treatments simultaneously within one protocol. Royston et al. (2011, Trials 12:81) developed a framework for the MAMS trials with time-to-event outcomes where an intermediate outcome can be used at the interim stages for lack-of-benefit analysis.

In this talk, we address the problem of how to find efficient MAMS trials with particular pairwise or familywise operating characteristics. We also introduce two Stata commands, nstagebin and nstagebinopt, that can be used to facilitate the design of efficient MAMS trials with binary outcomes.

Methods: The MAMS design is constructed by specifying the stagewise one-sided significance level and power for each pairwise comparisons, along with the minimum target treatment effect for the outcome of interest in that stage and the allocation ratio for the trial. Given these design parameters, the sample size required for each analysis are calculated.

However, this approach is problematic for two main reasons. First, it may not result in a design with the desired overall operating characteristics (feasible design). Second, there are likely to be many feasible designs for any pair of overall operating characteristics, some requiring smaller sample sizes than others. Therefore, the chosen design may not be the most efficient, or optimal, one for a particular true treatment effect. We introduce admissible designs which minimise a weighted sum of the expected sample size under the global null and alternative hypotheses. Using an ongoing 8-arm 3-stage MAMS trial in surgery, ROSSINI-2, we describe how to find efficient admissible designs based on an optimality criterion, using an efficient algorithm and a systematic search procedure.

Results: The proposed approach markedly facilitates the search for feasible (that is, designs which have a prespecified overall type I error rate and power) and admissible MAMS designs. Admissible designs are likely to be the most efficient choice in practice, suited for a particular range of treatment effects. The final choice of design will depend on prior beliefs about the effectiveness of the treatment under study, the relative importance of the maximum and expected sample sizes to the investigators or both.

Conclusion: Our proposed approach and the new Stata commands facilitate the design (and uptake) of MAMS trials with binary outcomes where more than one research question can be addressed under one protocol.

References: *[1] Choodari-Oskooei B, Bratton D, Parmar MKB, Stata Journal, 23(3), 774-798. https://doiorg/10.1177/1536867X231196295*

ORAL CONTRIBUTED SESSION 20
Clinical Trials III

OC20-3 Modelling time-treatment interactions to increase power in multi-arm multi-stage, and platform trials

Ziyan Wang, Dave Woods
Statistical Sciences Research Institute (S3RI) at University of Southampton, Southampton, United Kingdom

Background: Long-running trials can lead to dynamic treatment effects, for example with the expected difference between treatment and control varying with time. Such effects may occur due to changes in the population from which the trial Is recruited or from changes in treatment efficacy, e.g., due to increased clinician experience with a new surgical technique [1]. Ignoring such time trends can lead to inflation in the type I error rate, lower power and potential bias in the estimator of the treatment effects. Recent work has proposed methods for the design and modelling of trials that include treatment-independent time effects [2]. However, the impact of interactions between time and treatments has not been extensively studied.

Methods: In this work we propose and study new definitions of dynamic treatment effects for models including a time-treatment interaction that are applicable for continuous or binary responses. We use simulation studies to assess the effectiveness, in terms of error rates, of these approaches for Bayesian multi-arm multi-stage trials including early stopping for efficacy and futility, and response adaptive randomisation (RAR).

Results: Our approach is shown to increase power and control bias in treatment effect estimation compared to adoption of the standard definition of treatment effects, whilst maintaining type I error rates. Further, we show how RAR ratios is robust to our new definitions of dynamic treatment effects. We also demonstrate how the definition can extend to a nonlinear time trend via a random effects model. All simulation results are generated using our "BayesianPlatformDesignTimeTrend" R package, available on CRAN.

Conclusions: Our approach provides a practical method and some new insights into the problem of addressing different patterns of unequal time trends in complex multi-stage trials.

References: *[1] K. M. Lee, L. C. Brown, T. Jaki, N. Stallard, and J. Wason. Statistical consideration when adding new arms to ongoing clinical trials: the potentials and the caveats. Trials, 22:1–10, 2021. [2] Roig, M. B., Krotka, P., Burman, C.-F., Glimm, E., Gold, S. M., Hees, K., Jacko, P., Koenig, F., Magirr, D., Mesenbrink, P., et al. (2022). On model-based time trend adjustments in platform trials with non-concurrent controls. BMC medical research methodology. 22.1, pp. 1–16*

ORAL CONTRIBUTED SESSION 20
Clinical Trials III

OC20-4 A general and computationally tractable approach for exact statistical analysis in response-adaptive designs

Stef Baas[1,2], **Sofía Villar**[1], **Peter Jacko**[3,4]
1 University of Cambridge, Cambridge, United Kingdom
2 University of Twente, Enschede, Netherlands
3 Lancaster University, Lancaster, United Kingdom
4 Berry Consultants, Abingdon, United Kingdom

Clinical trials using a response-adaptive (RA) design adjust the allocation of participants to treatments sequentially based on the observed (response) data so far. This is done with the aim of reaching a certain objective such as increasing expected patient outcomes or statistical power. Proposals for RA designs of clinical trials face greater scrutiny in reviews by regulatory agencies, in part due to concerns surrounding type I error inflation, which can occur for standard tests under currently used RA designs. Furthermore, commonly used test approaches incorporating the RA design have many limitations: they either only work for specific designs and in specific scenarios, have a risk of model misspecification or Monte Carlo error, are conservative, or computationally intractable. This paper focuses on type I error control for statistical tests on binary outcomes collected from a trial with a control and treatment group using any RA procedure, either randomized or deterministic. We develop a general approach to construct conditional exact tests for RA designs, extending Fisher's exact test, and an unconditional exact test for RA designs, generalizing Barnard's test. An efficient implementation of forward recursion is used to compute the critical value as well as operating characteristics for clinical trials with large trial sizes (up to around 1,000 participants on a standard computer). We compare several tests on data collected using the randomized dynamic programming RA procedure, where a conditional exact test almost uniformly shows highest power. Furthermore, two real-life applications are considered. The first considers a case study based on a trial using a form of deterministic RA procedure based on a modification of the play-the-winner rule, where the exact tests, while controlling type I error, are shown to have higher power than the proposed test. The second application involves a second trial that used a blocked Bayesian group sequential RA design, where under the assumed optional stopping threshold there is substantial type I error inflation under misspecification of the success probability under the null hypothesis, while the proposed tests exactly control type I error. In conclusion, our method provides a general and computationally tractable method to ensure type I error control in complex and real-life RA designs. An important observation is that while the unconditional exact test is often the exact test resulting in the highest power for non-response-adaptive designs with equal allocation, our results show that in RA designs a conditional exact test can provide a more powerful alternative.

ORAL CONTRIBUTED SESSION 20
Clinical Trials III

OC20-5 An extended Bayesian semi-mechanistic dose-finding design for phase I oncology trials using pharmacokinetic and pharmacodynamic information

Chao Yang[1], Yisheng Li[2]
1 Eli Lily, Indianapolis, United States
2 The University of Texas MD Anderson Cancer Center, Houston, United States

Background: Su et al. (Pharmaceutical Statistics, 2022) proposed a semi-mechanistic dose-finding (SDF) design for phase I oncology trials using joint pharmacokinetic (PK) / pharmacodynamic (PD) modeling of the dose-toxicity relationship. They assume a latent, generic PD effect, and a 1-compartment intravenous (IV) bolus PK model. The improvement margin of the SDF design on the percent of correct selection of maximum tolerated dose (MTD) over the continual reassessment method (CRM) in a motivating phase I trial of gamma-secretase inhibitor for solid tumors is small. Besides, little is known on the SDF design's performance for treatments with more complex PK mechanisms.

Methods: Motivated by a phase Ib/II clinical trial of anti-CD20/CD3 T cell therapy for non-Hodgkin lymphomas, we extend the SDF model framework by incorporating measurements of a PD biomarker relevant to the primary dose-limiting toxicity (DLT). We propose joint Bayesian modeling of the PK, PD, and DLT outcomes. We conduct extensive simulation studies to evaluate the operating characteristics of the extended SDF design, for example, in terms of the percentage of correct selection of MTD and average number of patients allocated to MTD. We compare the proposed design with the existing SDF design (SDF-woPD), and other common phase I trial designs including the modified toxicity probability interval (mTPI), Bayesian optimal interval (BOIN) designs, and CRM.

Results: Our simulation studies show, on average, that the proposed design outperforms the mTPI, BOIN, CRM, and SDF-woPD designs, under a variety of dose-toxicity relationship scenarios. When the working PK model and the class of link function between the cumulative PD effect and DLT probability is correctly specified, the proposed design also yields better estimated dose-toxicity curves than CRM and SDF-woPD. Our sensitivity analyses suggest that the design's performance is reasonably robust to prior misspecification for the parameter in the link function, as well as misspecification of the PK model and class of the link function.

Conclusion: The extended SDF design is promising for improving efficiency of phase I oncology trials where measurements of toxicity PD marker are available to facilitate evaluation of the dose-toxicity curve.

ORAL CONTRIBUTED SESSION 21
Joint Models III

OC21-1

StCA winner

Joint modelling of longitudinal data and informative visiting process to predict subject-specific probabilities of future gaps in care

Achilleas Stamoulopoulos, Christos Thomadakis, Giota Touloumi
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Greece

Introduction: In HIV cohort studies, longitudinal markers, such as CD4 counts, are measured at patients' visiting times. When visiting times are driven by underlying health conditions, standard analyses of these measurements yield biased results. Moreover, longer period s without a visit usually suggest a gap in care. In this work, we propose to model the evolution of a longitudinal marker jointly with the visiting process and, based on this model, to derive personalized predictions of the probabilities of the next visit being longer than prespecified cut off time intervals.

Methods: For marker's evolution, a standard linear mixed model is assumed. For the visiting process, a proportional hazards model for the gap times between the marker's observation times is utilized. The two models are linked by the inclusion of correlated, normally distributed, random effects, and by including the last observed marker value as a time varying covariate in the visiting process submodel. Properties of the multivariate normal distribution are used to turn the multidimensional integral needed to derive log likelihood contribution for each subject to a unidimensional one This integration dimension reduction integrated in the estimation procedure allows straightforward extension to multiple normally distributed markers, even for categorical ones through the latent normal/probit parameterization. For model parameters' estimation, a pseudo adaptive Gauss Hermite rule, using only information of the visiting process was applied Given a patient's longitudinal history and visit times, we derive a formula for the conditional probability of his/her next gap time to exceed a relevant threshold A confidence interval for this quantity can be calculated using a Monte Carlo simulation method.

Results: A simulation study mirroring our motivating example was performed using 200 datasets of 500 patients. The proposed model, being correctly specified, yielded approximately unbiased results (% 0.09 3.85) and good coverage rates (93.3%-96.9%). Application on AMACS data reveals a significant, yet small effect of the observed CD4 counts on the visiting probabilities, and relatively weak correlations 0.07 to 0.31) between the random effects of the two 4% of MSM ha d a high probability (> 0.5) of having a gap in care ≥1.5 years compared to 18% of PWID.

Conclusions: The proposed model is a useful alternative for analyzing multiple markers under an informative visiting process as it substantially reduces the computation burden, needing only one dimensional integration. In practice, personalized predictions for the next visit time from a well calibrated model could be used to identify individuals more prone to disengage from care.

ORAL CONTRIBUTED SESSION 21
Joint Models III

OC21-2 Joint modelling of the temporal relationships between multivariate longitudinal markers of Alzheimer's disease progression and clinical endpoints

Anaïs Rouanet[1], Karine Pérès[2], Viviane Philipps[1], Cécile Proust-Lima[2]
1 University of Bordeaux, Bordeaux, France
2 Bordeaux Population Health, Inserm U1219, Bordeaux, France

Background and Introduction: Alzheimer's disease and related dementias (ADRD) are a major public health issue with substantial economic and social costs. Yet, their underlying mechanisms are still poorly understood. The wealth of biomarker data now available offers an unprecedented opportunity to better understand the complex pathological pathways of ADRD (encompassing neurodegeneration, cognitive decline, functional autonomy), essential to improve prevention strategies and individual healthcare. Jack et al. (2013) proposed a theoretical model of the Alzheimer's pathological cascade; however, it has not been statistically validated yet. Indeed, current statistical tools are inefficient at describing jointly multiple ADRD biomarkers and capturing their temporal inter-relationships rigorously. The objective was to develop an approach to model multivariate longitudinal outcomes and understand how each repeated outcome influences the change in others, while accounting for attrition by dementia and death.

Methods: We propose an original joint modelling framework to describe the complex dynamics of the different dimensions involved in ADRD progression and apprehend their causal temporal relationships. Taddé et al. (2020) proposed a dynamic causal approach that models multiple continuous repeated outcomes by a network of latent processes in a multivariate linear mixed model. Using difference equations, they model their temporal inter-relationships by describing the change over time of each latent process according to different characteristics of the others. We extended this approach to handle ordinal repeated outcomes and competing risks. The associations with ADRD diagnosis and death are taken into account via a shared random effect joint modelling approach, to handle potential informative missingness mechanisms. The parameters are obtained through Maximum likelihood estimation using a Newton-Raphsonlike algorithm. The estimation procedure is available on https://github.com/anarouanet/CInLPN2.

Results: The methodology is applied on a French observational cohort on cognitive ageing, the Paquid Study (Letenneur et al., 1994). The objective is to disentangle the temporal relationships between cognitive impairment, depressive symptomatology and functional autonomy, which are major drivers of ADRD natural history. The results show a reciprocal temporal relationship between the functional and cognitive dimensions. Moreover, depressive symptomatology seems to be a confounding factor between cogni1ve decline and dementia.

Conclusion: The results would help understand the complex interplay between key pathological dimensions of ADRD, especially in the presence of reverse causation, and identify relevant target domains which may slow down the evolution of the diseases. While initially designed for dementia research, this methodology offers the opportunity to address critical questions in other chronic diseases characterized by multiple longitudinal markers.

References: *[1] Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology (2013), 12, 207–216. [2] Letenneur, Luc, Daniel Commenges, Jean-François Dartigues, et Pascale Barberger-Gateau. « Incidence of dementia and Alzheimer's disease in elderly community residents of southwestern France. » International journal of epidemiology (1994) 23(6): 1256-61. [3] Taddé, Bachirou O., Hélène Jacqmin-Gadda, Jean-François Dartigues, Daniel Commenges, et Cécile Proust-Lima. « Dynamic modeling of multivariate dimensions and their temporal relationships using latent processes: Application to Alzheimer's disease ». Biometrics (2020) 76(3): 886-99*

ORAL CONTRIBUTED SESSION 21
Joint Models III

OC21-3 Simultaneous modelling of longitudinal measurements and multi-state data with a cured fraction using a joint model

Behnaz Alafchi, Elaheh Talebi-Ghane
Department of Biostatistics, School of Public Health, Modeling Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

In many biomedical cohort studies, the progression of disease was modeled instead of single event times. In this context, the progression of disease is considered as a multi-state process and the transition between different health states are modeled in order to simultaneously examine the impact of the longitudinal repeated measurements on disease progression [1]. Moreover, there are some situations that some patients may not experience any recurrent events nor death during a long follow-up period; these patients can be denoted as the cured group. Although there are some few studies that proposed joint model in the presence of cure fraction [2,3], we could not find any study that consider cured fraction for joint modeling of longitudinal and multi-state data. The purpose of this study was to proposed a joint model for longitudinal and multi-state data in the presence of cure fraction. This suggested model consists of two sub-models: a linear mixed-effects sub-model for longitudinal data, and a multi-state sub-model with transition-specific proportional hazards for times of transitions between different health states in the presence of cure fraction. These sub-models are linked by shared random effects which is assumed to follow a multivariate normal distribution. Moreover, the probability of cured fraction is described by the logistic model and include four different states that can occur for patients: (1) who experienced the recurrent event and death; (2) who survived after experiencing relapse; (3) who deceased without recurrent event; (4) who may experience neither recurrence nor death, which we referred to idiomatically as cured patients. The model parameters are estimated using the maximum likelihood estimation based on numerical integration for a parametric and piecewise baseline hazard function and their standard errors were obtained by the sample covariance matrix across the bootstrap datasets. The simulation study was applied to evaluate the performance of proposed model. The results demonstrated that in the presence of cure fraction, a considerable advantage was observed in the new model regarding the regression coefficients, their standard errors and the probability of coverage and AIC mean which consequently led to a better performance in comparison with reduced model. The proposed joint model enabled us not only to analyze the times of transitions between different health states and estimate the effect of covariates on them and simultaneously on the cure rate, but also, estimate the effect of covariates on the longitudinal responses, while accounting for the relationship among all responses.

References: *[1] Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust-Lima C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. Statistics in medicine. 2016;35(22):3933-48.2. [2] Liu L, Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. Biometrics. 2016 Mar;72(1):204-14. [3] Kim YJ. Joint model for recurrent event data with a cured fraction and a terminal event. Biometrical Journal. 2020 Jan;62(1):24-33.*

ORAL CONTRIBUTED SESSION 21
Joint Models III

OC21-4 Analysis of disease trajectories and treatment effects by combining variational autoencoders with random effects models

Clemens Schaechter[1], Maren Hackenberg[1], Max Behrens[1], Thorsten Schmidt[2], Felix Ndonfack[2], Astrid Pechmann[3], Janbernd Kirschner[3], Harald Binder[1]

1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

2 Department of Mathematical Stochastics, Faculty of Mathematics and Physics, University of Freiburg,Freiburg, Germany

3 Department of Neuropediatrics and Muscle Disorders, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Background: In longitudinal clinical registries for rare diseases, patients may receive a variety of treatments, including transitions between treatments as new therapeutic options emerge. As an additional challenge, different measurement instruments might have been used for different follow-up visits to characterize patient trajectories tailored to the patient's age or capabilities. Therefore, an approach is needed that models the effect of treatment and treatment switches on patient trajectories, while providing a joint representation across measurement instruments.

Methods: We model patient trajectories within a dimension-reduced representation. Specifically, we utilize variational autoencoders to map measurement instruments at the item level to a latent space. We use separate encoder and decoder networks for every measurement instrument and subsequently align their latent representations. Latent trajectories and effect s of treatment and treatment switches are modeled through a random effect s linear regression model that is estimated simultaneously with the autoencoder parameters. Baseline information is incorporated through fixed effects, ensuring that the model accurately captures their effect on each patient's trajectory.

Results: We illustrate the proposed approach with data from the SMArtCARE registry on patients with spinal muscular atrophy (SMA). There, our approach is seen to handle multiple measurement instruments with various temporal observation patterns and irregularly timed visits. The latent representation is seen to be improved by imposing a random effect s model. The later also provides treatment effect estimates to achieve a nuanced understanding of each patient's individual disease progress and the impact of various treatment options.

Conclusion: Our approach can model disease trajectories and treatment effect s in rare diseases with various measurement instruments, as shown in the SMArtCARE registry for spinal muscular atrophy. By imposing a random effect s model onto the dimension-reduced latent representation, we combine the benefits of flexible representation learning and classical statistical modeling.

ORAL CONTRIBUTED SESSION 21
Joint Models III

OC21-5 Joint modelling of time-dependent biomarker variability and time-to-event outcomes

Felix Boakye Oppong[1,2], Dimitris Rizopoulos[1,3], Thierry Gorlia[2], Nicole S. Erler[1,3]
1 Department of Biostatistics, Erasmus Medical Center Rotterdam, Rotterdam, Netherlands
2 EORTC Headquarters, Brussels, Belgium
3 Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, Netherlands

Background: Assessing the typically complex relationship between repeatedly measured biomarkers and the risk of experiencing a particular clinical outcome is essential in many medical applications. Joint models for longitudinal and survival data have become the method of choice for this task since they allow us to include (combinations of) different characteristics of the longitudinal trajectories as predictors of the event outcome.

In oncology and cardiology, instability in specific disease markers may predict a patient's risk of experiencing a particular event. This is the case in our motivating study, a randomized, multicentre, phase 3 trial in patients with newly diagnosed glioblastoma where the variability in haematological measures is likely associated with the risk of death.

However, the commonly used association structures, such as the expected value or slope of the longitudinal trajectory or the area underneath it, are based on the estimated trajectory and ignore the variability in the marker measurements. Incorporating marker variability enhances understanding of the markers' relationship with clinical outcomes, potentially leading to better risk stratification and treatment personalization.

Only a few studies have investigated how to consider a longitudinal marker's variability in a joint model, and existing approaches cannot be fitted using standard joint modelling software.

Method: We propose a two-step approach to capture the variability of a time-dependent biomarker and link it to a time-to-event outcome in a joint model.

Specifically, residuals are extracted from a mixed model and transformed to obtain a patient- and time-dependent measure of the marker's variability (i.e. the variance or standard deviation). A mixed model for the variability is then included in the joint model.

Our approach facilitates flexible modelling of the marker's variability over time, e.g., using splines, as well as using different functional forms for the effect of the marker's variability on the time-to-event outcome.

Using our motivating data of glioblastoma patients, we compare different transformations and specifications of the mixed model for the residuals regarding their implications for interpretation and computational aspects.

Results: Using linear mixed models of log- or log2-transformed standard deviation or variance gave consistent results for value, slope and cumulative effect associations of white blood cell variability with mortality risk. The use of a Gamma model resulted in slightly different conclusions.

Conclusion: Our proposed approach provides a convenient extension of the existing joint modelling methodology to incorporate the variability of longitudinal markers. It can be fitted using standard joint model software.

ORAL CONTRIBUTED SESSION 22
Clinical Trials IV

OC22-1 Joint modelling of survival and longitudinal data for sequential decision-making in clinical trials

Zhujie Gu[1], **Haiyan Zheng**[2], **Jessica Barrett**[1]
1 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
2 University of Bath, Bath, United Kingdom

Background: Joint modelling of time-to-event (TTE) data and longitudinal biomarkers associated with the TTE endpoint can improve efficiency of survival trials and provide insight into the causal pathway of the treatment on the TTE endpoint. To date, research on leveraging joint modelling in group sequential designs (GSD) for survival trials is limited. A major challenge is to summarise the potentially time-varying direct and indirect treatment effect mediated by the biomarkers into a single clinically interpretable test statistic, based on which a group sequential trial can be designed.

Methods: In this study, we propose a flexible parametric joint modelling framework that facilitates GSD based on extrapolated survival time. The joint modelling framework consists of a linear mixed effect sub-model for modelling the trajectories of biomarkers, and a full-parametric accelerated failure time (AFT) sub-model for modelling the relationship of the TTE outcome with the biomarker trajectories and the treatment. The parameters of the AFT sub-model provide clear interpretation regarding the direct and indirect treatment effects in terms of prolonged or shortened survival time, and are known to be robust in case of omitted variables. An expectation–maximisation algorithm is developed to obtain the maximum likelihood estimator (MLE) based on the accumulating trial data. Test statistics constructed based on the MLE follow the canonical joint distribution, so we derive group sequential boundaries for efficacy/futility.

Results: The methodology is evaluated using extensive simulations. Simulation results suggest improved efficiency over traditional GSD without leveraging the longitudinal biomarkers.

Conclusion: Our proposed methodology can improve efficiency of group sequential designs by joint modelling of time-to-event outcome and its associated longitudinal biomarkers, and provide clear interpretations.

ORAL CONTRIBUTED SESSION 22
Clinical Trials IV

OC22-2 Automated, efficient and model-free inference for randomised clinical trials via data-driven covariate adjustment

Kelly Van Lancker[1], Iván Díaz[2], Stijn Vansteelandt[1]
1 Ghent University, Ghent, Belgium
2 New York University, New York, United States

Background: In 2023, the U.S. Food and Drug Administration (FDA) released guidance for industry on "Adjustment for Covariates in Randomized Clinical Trials for Drugs and Biological Products". Covariate adjustment is a statistical analysis method for improving precision and power in clinical trials by adjusting for pre-specified, prognostic baseline variables. Though covariate adjustment is recommended by the FDA and the European Medicines Agency, many trials do not exploit available information in baseline variables or only make use of the baseline measurement of the outcome. Specifically, its practical implementation has been hindered by the regulatory mandate to pre-specify the optimal baseline covariates that will be adjusted for along with the functional form of the outcome model.

Methods: We address this by using automated data-adaptive methods, which `adapt' to the data by making use of automated variable selection (e.g., stepwise AIC-based selection, Lasso) or machine learning methods (e.g., random forest regression, neural networks). Nonetheless, their application also introduces potential concerns as they are generally optimized for predictive accuracy, and may be sub-optimal for treatment effect estimation. Inspired by the debiased learning literature, we expand the covariate-adjusted estimator suggested in FDA's guidance to facilitate its compatibility with the aforementioned data-adaptive approaches. Specifically, the independence between treatment and baseline covariates renders these covariate-adjusted estimators immune to model misspecification as well as variable selection uncertainty. It moreover guarantees a valid inference that is insulated against model misspecification. This is a strong result, given that the debiased learning usually requires correctly specified models. We provide a detailed methodology overview and empirical study results.

Results: The findings offer a promising avenue for improving the statistical power of trial analyses through data-adaptive covariate adjustment. We show that, surprisingly, a) one of the proposed estimators achieves exact unbiasedness -even when using machine learning predictions that are themselves biased, and b) valid inference can be obtained in the non-trivial case where the propensity score may become a complicated expression of the estimated parameters, even when the outcome model is misspecified.

Conclusion: The proposed approaches enable the use of data-adaptive methods without distorting the interpretation or validity of the treatment effect estimate and its standard error. They simplify the difficulty of pre-specifying the statistical analysis to pre-specification of the full covariate set, along with the algorithm for variable selection/model building that will be employed. This form of pre-specification is what we should strive for in the future.

ORAL CONTRIBUTED SESSION 22
Clinical Trials IV

OC22-3 An overview of prognostic scoring adjustments: Applied to neurodegenerative diseases

Harry Parr, Doug Thompson, Aris Perperoglou
Statistics and Data Science Innovation Hub (SDS-IH), GSK, United Kingdom

Introduction: Adjusting for covariates in RCTs can improve statistical efficiency for estimating and testing treatment effects [1]. Extensions of this use prognostic scores (PS), composed by a single composite 'oracle' variable which contains the counterfactual prediction of what would have happened in the case of the status quo (standard-of-care/control), i.e. the true PS. To estimate the oracle, a prognostic model (or AI/ML ensemble) is trained on (observational) historical controls data to construct subjects' estimated PSs. These PSs are adjusted for in the primary analysis (linear) model, improving precision of the estimated treatment effect. This is advantageous as it does not adversely compromise type-I error, unlike alternative historical borrowing methods. Recent studies focus on methods to reduce the original sample size (SS) calculation, whilst maintaining their power; there has been less focus on power improvement [2]. In settings where it is not desirable to reduce SS (e.g. rare diseases), the focus is to maximise power and hence probability-of-success (PoS). We demonstrate and apply a methodological framework to quantify power and PoS improvement.

Methods: First, we review the current literature and then present a workflow to train an ensemble of ML models to construct PSs and then quantify the power improvement. Synthetic data is generated, motivated from real Alzheimer's disease studies using mixed-effect model of repeated measures (MMRM). The endpoint is a change of cognitive decline from baseline. We perform simulations for two methods reducing the asymptotic variance, and the effective SS inflation approaches. We stress test our assumptions under various scenarios, including non/linearity between PS-outcome, heterogeneous treatment effects, and population shift, to quantify the impact of change. We vary necessary parameters to give a range of plausible values and summarise rules-of-thumb of SS reductions, or power improvements.

Results: We demonstrate the use of PSs in MMRM, and quantify power improvement. We see empirically and confirmed with simulations that incorporating PSs is robust and does not worsen bias, and often improves our precision of our treatment effect, improving power by an additional >5-10% from intended 80%.

Conclusions: Leveraging observational data and machine learning techniques can effectively capture the nonlinear and temporal dynamics between the PS and outcome, enhancing trial efficiency and maximising PoS. We surmise the expected improvement given plausible assumptions that could be incorporated into a statistical analysis plan. These methods ought to be applied as standard where possible.

References: *[1] FDA-2019-D-0934 (2023). [2] S. Siegfried et al., Biom. J. 65 (2023) 2100349.*

ORAL CONTRIBUTED SESSION 22
Clinical Trials IV

OC22-4 Principal treatment effects: A novel approach to non-linear dose-response modelling in complex interventions

Mollie Payne, Ben Carter, Amy Hardy, Richard Emsley
King's College London, United Kingdom

Background: Talking therapies are well-recognised mental health interventions, and yet it is not clear how much therapy (the 'dose') is required for a patient to achieve a good outcome. Despite a number of studies estimating the optimal number of sessions needed, the methods that are commonly used ignore selection bias due to non-adherence, are unable to estimate non-linear effects of dose or lack causal interpretation [1]. Recent causal inference methods, despite being causally valid, require an assumption of linearity, which is unlikely to be met.

Our aim was to develop a method that allow us to ascertain causal effects of dose from complex interventions, primarily without making linearity assumptions and accounting for selection bias.

Methods: Our proposed method combines ideas from structural mean modelling and principal stratification, eliciting a two step procedure to estimate the causal effect of treatment for those who would attend S=s sessions, had they been offered treatment. We predict the number of sessions that would be attended by the sample and then estimate the effect of treatment on outcome within each strata of sessions. This can be grouped if a large number of sessions are planned.

We evaluated this method using a Monte Carlo simulation study in the ADEMP framework [2], where we assessed the performance over a variety of trial scenarios. We varied the strength of predictors of dose, unmeasured confounding and the dose-response function. We evaluated performance based on measures of bias, coverage and empirical standard error. We apply the method to a trial of psychological therapy in paranoia.

Results: We demonstrated the ability of our proposed method to capture non-linear effects of dose as well as address the selection effects that led to different levels of dose received. This method can easily be applied to real complex intervention data when evaluating the effect of dose, within the strata of patients who would have attended S=s sessions.

The performance of the proposed method depends on the strength of predictors of dose, estimates of bias may be inflated if we have weak predictors and are unable to predict the correct range of therapy sessions.

Conclusion: There is a clear need for dose-response modelling methods that are applicable to complex interventions and allow for non-linear modelling of dose. Our method is able to model non-linear trends, even when the underlying dose function is unknown. When we are able to predict dose, we obtain unbiased estimates for the principal treatment effect.

References: *[1] Payne, M., Stringer, D., Carter, B. & Emsley, R. A systematic review of statistical methods in dose response modelling for complex interventions. [submitted]. [2] Morris, T. P., White, I. R. & Crowther, M. J. Using simulation studies to evaluate statistical methods. Statistics in Medicine, 38, 2074-2102*

ORAL CONTRIBUTED SESSION 22
Clinical Trials IV

OC22-5 Estimating the Complier Average Causal Effect (CACE) for randomised therapy trials – Is a binary compliance definition appropriate?

Teresa Lee, Baptiste Leurent, Julie Barber, Rumana Omar
University College London, London, United Kingdom

Background/Introduction: Non-compliance happens in randomised controlled trials (RCTs) when participants only receive part of the intended treatment or intervention. While intention-to-treat (ITT) analysis is recommended to estimate the average benefit, it is also commonly of interest to estimate the causal effect in those that complied with the intervention. This is not straightforward, as compliance is usually associated with outcomes, and naïve analysis such as per-protocol are known to be biased. A more popular approach is to estimate complier average causal effect (CACE) which uses randomisation as an instrument to estimate causal effect in compliers. CACE was initially developed for binary compliance (receiving the treatment or not at all), and it relies on several assumptions such as exclusion restriction. In RCTs of therapy with multiple sessions, it is common to define compliance using a threshold, with compliers being those attending at least n sessions, and others assumed to be non-compliers. This dichotomisation is likely to violate the exclusion restriction assumption to some extent, but it is not clear how much this impacts the CACE estimates.

The aim of this study is to examine the performance of different methods to estimate the causal effect of treatment in compliers when the amount of intervention received is non-binary, such as in therapy trials.

Methods: We used simulations to examine CACE analysis assuming binary compliance threshold and CACE assuming a linear effect proportional to the number of sessions attended in terms of bias in the estimate of the treatment effect. A comparison was also made with the ITT estimate. We simulated a therapy trial with different compliance levels, different types of linear and non-linear associations between compliance and outcomes, and different dose (number of sessions)-response associations.

Results: When the true dose-response association was a jump function (i.e. no treatment effect below a specified cut off) the binary CACE analysis tended to estimate with minimal bias, provided the chosen compliance cut off was close to the true value. Under most other scenarios, the biases were smallest when using a linear CACE analysis.

Conclusion: In RCTs of therapies with multiple sessions the true dose-response association is rarely known, and unlikely to follow a jump function as assumed for a standard binary CACE analysis. The continuous CACE method which allows linear treatment effects, can be interpreted as the average causal effect by session attended and performed generally well under different dose-response assumptions.

ORAL CONTRIBUTED SESSION 23
Competing Risks I

OC23-1 Non-Markov non-parametric estimation of complex multistate outcomes after hematopoietic stem cell transplantation

Judith Vilsmeier[1], Sandra Schmeller[1], Daniel Fürst[2,3], Jan Beyersmann[1]
1 Institute of Statistics, Ulm University, Ulm, Germany
2 Institute of Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service, Baden Wuerttemberg – Hessen and University Clinic Ulm, Ulm, Germany
3 Institute of Transfusion Medicine, Ulm University, Ulm, Germany

Background: In many studies, probabilities of non-standard time-to-event endpoints are of interest which are more complex than overall survival. One such probability is chronic GvHD- and relapse-free survival, the probability of being alive and in remission after stem cell transplantation and not suffering from chronic graft-versus-host disease (GvHD), with chronic GvHD being a recurrent event. Because the probabilities for such complex, non-standard endpoints with recurrent events may not fall monotonically but may also rise again, one should not use a simple Kaplan-Meier estimator for the estimation of these probabilities, but the Aalen-Johansen estimator. In some multistate models, it is also possible to estimate probabilities for complex, non-standard endpoints using linear combinations of Kaplan-Meier estimators.

Methods: For these linear Kaplan-Meier combinations, we propose a wild bootstrap procedure for inference and to obtain confidence bands with the aim to compare the results with confidence bands obtained using the wild bootstrap for the Aalen-Johansen estimator in non-Markov scenarios. In the proposed wild bootstrap procedure for the linear combinations of Kaplan-Meier estimators, the limiting distribution of the Nelson-Aalen estimator is approximated using the wild bootstrap and transformed via the functional delta method. This approach is easily adaptable to different multistate models.

Results: Using real data, confidence bands are generated using the wild bootstrap approach for the chronic GvHD- and relapse-free survival. The coverage probabilities of confidence intervals and confidence bands generated by Efron's bootstrap and by the wild bootstrap are examined with simulations. The results show that the confidence bands and confidence intervals of the linear combination of Kaplan-Meier estimators obtained with the wild bootstrap are wider than those of the Aalen-Johansen estimator and that their coverage probabilities are greater than 95%.

Conclusion: A violation of the Markov assumption does not affect the Aalen-Johansen estimator nor the wild bootstrap as long as state occupation probabilities are estimated and censoring is random. This, together with the fact that the linear combination of Kaplan-Meier estimators can estimate probabilities to be out of bounds, points to the Aalen-Johansen estimator as a good candidate for a meaningful estimator for complex outcomes even in the non-Markov case.

ORAL CONTRIBUTED SESSION 23
Competing Risks I

OC23-2 Semiparametric regression analysis of competing risks data with missing not at random cause of failure

Giorgos Bakoyannis[1], Aristofanis Rontogiannis[2], Ying Zhang[3], Wanzhu Tu[1], Constantin Yiannoutsos[1]
1 Indiana University Indianapolis, Indianapolis, United States
2 Athens University of Economics and Business, Athens, Greece
3 University of Nebraska Medical Center, Omaha, United States

Background: The analysis of competing risks data with incompletely observed cause of failure typically relies on the missing at random (MAR) assumption. However, this assumption is not realistic in many real-world settings. This may lead to biased estimates and invalid conclusions.

Methods: We propose a rigorous sensitivity analysis methodology for the semiparametric proportional cause-specific hazards model under a class of missing not at random (MNAR) mechanisms, which contains the MAR mechanism as a special case. Our goal is to estimate a set of population regression coefficients under the different MNAR mechanisms and compute a confidence interval for this set. This confidence interval can be used to assess the robustness of qualitative conclusions about statistical associations against not-at-random missingness of the cause of failure.

Results: The methodology is shown to be asymptotically valid using empirical process theory. Extensive simulation experiments show that the methodology performs well even with small sample sizes and large missingness rates. The simulation experiments also show that methods that rely on the MAR assumption lead to bias and invalid conclusions. The proposed methodology is applied to competing risks data from a large multicenter HIV study in sub-Saharan Africa where a substantial portion of causes of failure is missing not at random.

Conclusions: The proposed methodology provides a rigorous way to evaluate the robustness of the concussions from a competing risks analysis in situations with potential MNAR causes of failure.

OC23-3 Variable selection for penalised multi-state models incorporating molecular data

Kaya Miah[1], Jelle J. Goeman[2], Hein Putter[2], Annette Kopp-Schneider[1], Axel Benner[1]
1 Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany
2 Department of Biomedical Data Sciences, Leiden University Medical Center (LUMC), Leiden, Netherlands

Background/Introduction: In the era of precision medicine with increasing molecular information, the use of a multi-state model is required to capture the individual disease pathway along with underlying etiologies with greater precision. Especially the availability of big data with numerous covariates induces several statistical challenges for model building. For multi-state models based on high-dimensional data, effective modeling strategies are crucial to determine an optimal, ideally parsimonious model.

Methods and Results: Standard methods integrate regularization into the fitting procedure to conduct variable selection. In the multi-state framework, linking covariate effects across transitions is needed to conduct joint variable selection. A useful technique to reduce model complexity is to address homogeneous covariate effects for distinct transitions based on a reparametrized model formulation. We integrate this approach to data-driven variable selection by extended regularization methods within multi-state model building. We propose the sparse-group fused lasso (SGFL) penalized Cox-type regression in the framework of multi-state models combining the penalization concepts of pairwise differences of covariate effects along with transition grouping. For optimization, we adapt the alternating direction method of multipliers (ADMM) algorithm to transition-specific hazards regression in the multi-state setting.

In a simulation study and application to acute myeloid leukemia data, we evaluate the SGFL penalization's ability to select a sparse model distinguishing between relevant transition-specific and similar cross-transition effects. We investigate settings in which the combined penalty is beneficial compared to only fused or grouped regularization.

Conclusion: Thus, effective model selection strategies in multi-state survival analysis are required for enhancing comprehension and interpretation of individual disease pathways, distinct oncological entities and tailored precision therapies, leading to improved personalized prognoses.

ORAL CONTRIBUTED SESSION 23
Competing Risks I

OC23-4 Bridging multi-state models and network theory: An application to multimorbidity and aging research

Caterina Gregorio[1], Francesca Ieva[2,3], Amaia Calderón-Larrañaga[1], Davide Liborio Vetrano[1]
1 Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Solna, Sweden
2 MOX - Modelling and Scientific Computing Laboratory, Department of Mathematics, Politecnico di Milano, Milan, Italy
3 HDS, Health Data Science center, Human Technopole, Milan, Italy

Introduction: Multi-state models represent state-of-the-art methods to model transitions across health states over time and individuals' paths. This is of great interest in the context of many clinical conditions such as multimorbidity (i.e., the co-occurrence of multiple diseases in the same individual) where there is the need to characterize natural history-related patterns and stratify individuals according to their risk of progressing to more severe morbidity stages or dying prematurely. In this context, we aim to propose innovative approaches based on coupling multistate models and network theory.

Methods: One challenge in general multi-state models is that as the number of transitions increases, the estimation of the quantities to describe the trajectories of individuals becomes analytically unfeasible. On the other hand, they can be obtained using microsimulation after estimating the transition hazards from the data. We propose to use such an approach coupled with well-established network methods (i.e., Dijkstra´s algorithm) to identify clinically relevant features of a multi-state model such as the most common paths, the quickest paths to death, and a measure of the importance of each transition in terms of the impact of its "interruption" if we were hypothetically able to prevent people from transitioning. Transition hazards are estimated using the Royston-Parmar model to ensure the flexibility of the models in terms of the shape of the baseline hazards and the inclusion of time-dependent effects of the covariates as appropriate.

Results: Data from an ongoing 12-year population-based study (the Swedish National Study on Aging and Care in Kungsholmen) was used. We included 1,157 adults aged ≥60 years with mild multimorbidity at baseline. The states considered for describing the natural history of multimorbidity were "mild", "complex: psychiatric/respiratory", "complex: neuropsychiatric", "complex: sensory impairment", "complex cardiometabolic, and "death". As covariates, relevant socio-economic and cardiometabolic risk factors were included. Through the proposed approach, we obtained i) a temporal characterization of multimorbidity evolution, ii) related risk profiles, and iii) a list of transitions to be targeted to reduce the flow of individuals through the highest-burden paths.

Conclusions: Multi-state models can be seen as networks. We showed how this perspective opens the possibility to infer clinically relevant trajectory-related characteristics that go beyond classical measures such as life expectancy and time spent in different health states. Moreover, the proposed approach provided novel insights regarding the fingerprint of trajectories of older people affected by multimorbidity.

ORAL CONTRIBUTED SESSION 23
Competing Risks I

OC23-5 A joint spatiotemporal multivariate model with competing risks

Juliette Ortholand, Stanley Durrleman, Sophie Tezenas Du Montcel
Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpitalde la Pitié Salpêtrière, Paris, France

Respiratory failure is the leading cause of death in Amyotrophic Lateral Sclerosis (ALS). The use of Non-Invasive Ventilation (NIV) as a means of survival is symptomatic of an advanced stage of the disease. A better understanding of its timing could help improve patient support planning. Despite the vast databases available, reconstructing a complete disease course remains complex due to the limited number of patients followed from symptom onset to death. The main challenge is, therefore, to realign patients on a common disease timeline (temporal aspect) and on a common disease presentation (spatial aspect). In this context, we developed a joint spatiotemporal multivariate model with competing risks, extending a joint univariate model [1]. We developed a joint model using an existing longitudinal spatiotemporal submodel with a latent disease age and random effects, named sources, controlling disease presentation. For the survival submodel, we used a competing cause-specific model, estimating a Weibull distribution from latent disease age with a proportional impact of the sources on the hazard. Estimation was carried out with an open-source package, Leaspy, using an MCMC-SAEM algorithm, and was validated on simulated data. 2,177 ALS patients were extracted from the PRO-ACT database, associated with 16,400 visits, 570 NIVs and 245 deaths or tracheostomies. We built the model using three subscores of the revised ALS Functional Rating Scale, and two competing events: NIV and death or tracheostomy. The model described two spatial dimensions of the disease presentation. One corresponding to higher risk of NIV (HR: 1.20[1.11,1.31]), higher risk of death or tracheostomy (HR: 1.36[1.21,1.56]), later bulbar progression (delay: 7.92[6.34,9.92] months), later fine motor progression (delay: 1.95[1.05,2.59] months) and earlier gross motor progression (delay: -8.50[-9.01,-7.94] months). A second to lower risk of NIV (HR: 0.93[0.86,0.99]), lower risk of death or tracheostomy (HR: 0.82[0.73,0.92]), later bulbar progression (delay: 16.07[14.66,17.29] months), earlier gross motor progression (delay: -6.59[-7.00,-6.21] months) and earlier fine motor progression (delay: -3.24[-4.16,- 1.98] months). We characterized heterogeneity of event risk using random effects: after correction for speed and onset, men had a significantly lower NIV risk factor than women (HR: 0.92[0.90,0.93]) and a significantly lower death or tracheostomy risk factor (HR: 0.85[0.82,0.87]), whatever the onset site. We have developed a joint spatiotemporal model with competing risks to analyze the timing of NIV occurrence. This work opens perspectives to design predictive and personalized therapeutic strategies for ALS and many other diseases.

References: *[1] Ortholand et al., preprint, 2024, https://arxiv.org/abs/2401.17249*

ORAL CONTRIBUTED SESSION 24
Survival Analysis II

OC24-1 Flexible hazard regression for generalised interval-censored time-to-event data and the gicsurv Rshiny app

Jordache Ramjith[1,2], Marianne Jonker[2], Teun Bousema[1]
1 Department of Medical Microbiology, Radboudumc, Nijmegen, Netherlands
2 IQ Health Science department, Radboudumc, Nijmegen, Netherlands

Background: Time-to-event analysis is a fundamental tool in epidemiology and clinical research. We present a model for generalised interval-censored (GIC) time-to-event data where the baseline hazard is flexibly estimated through penalised splines. Here GIC data refer to data that may have exactly-known event-times/right-censored event-times/interval-censored start-times/interval-censored event-times and any combination thereof. The existing literature for double-interval-censored time-to-event or interval-censored start-times is restricted to stringent parametric assumption which may result in biased survival estimates and covariate effects and there are no available software. A flexible baseline hazard provides information on the underlying risk of an event over time without restrictive parametric assumptions, which may be essential to draw meaningful conclusions.

Methodology: To model a flexible baseline hazard, a B-splines specification of the baseline hazard is used, and for simplicity proportional hazards of covariate effects is assumed. To avoid overfitting, the B-splines basis coefficients are penalised. The penalty factor dictating the amount of smoothness is estimated simultaneously with the B-spline parameters and the covariate effects in a two-step maximum likelihood estimation approach. The first step utilises the link between spline parameters and random effects to estimate all model parameters. In the second step, using restricted maximum likelihood and a Laplace approximation, the random effects variance is estimated where the inverse of the random effects variance is equivalent to the penalty factor. The steps are repeated until the criteria for convergence is reached. Through a simulation study, bias and other performance measures for the estimation procedure are evaluated. A real-world application highlights the usefulness of the methodology and the Rshiny app gicsurv that compliments the methodology.

Results: Analysis of simulated data show unbiased estimated regression coefficients and consistent standard errors. A comparison between the gold-standard Cox model estimates using the underlying exactly-known event-times and estimates from our approach applied to exact/interval-censored/double-interval-censored times all show nearly perfect correlation for regression coefficients and standard errors. In a real-world application on a Ugandan cohort based on 104 incident asymptomatic malaria infections, we looked at two time-to-event outcomes. Sickle-cell trait and higher baseline parasite density were associated with a higher hazard of gametocyte appearance over time while higher baseline parasite densities were also associated with a lower hazard of malaria clearance over time.

Conclusion: The model and methodology highlighted in this paper along with the Rshiny app, provide an efficient tool for estimating flexible hazard and survival curves, and/or covariate effects for GIC time-to-event data.

ORAL CONTRIBUTED SESSION 24
Survival Analysis II

OC24-2 Group sequential methods based on supremum logrank statistics under proportional and non-proportional hazards

Jean Marie Boher[1,2], Thomas Filleron[3], Pierre Bunouf[4], Richard J Cook[5]
1 Institut Paoli Calmettes, Marseille, France
2 Aix Marseille Univ, INSERM, IRD, SESSTIM, Marseille, France
3 Biostatistics Unit, Institut Claudius Regaud-IUCT-O, Toulouse, France
4 Laboratoires Pierre Fabre, Toulouse, France
5 Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada

Background: In recent years there has been an increased appreciation that in many settings the proportional hazards assumption may fail to hold. The use of combination tests based on multiple weighted logrank test statistics are recommended for use in oncology trials because of their sensitivity to a wide range of alternatives, including detecting early, intermediate or late differences. We propose new tests based on the combination of multiple supremum logrank statistics as an alternative method for providing robust power under proportional and non-proportional risk alternatives, and show how to conduct repeated significant tests for detecting early, intermediate and late differences in randomized clinical trials.

Methods: We consider the setting of a clinical trial in which n independent individuals are randomized to receive either an experimental treatment or standard care, with a balanced or unbalanced random allocation scheme between arms, and followed up on study. We consider maximal combination of supremum logrank tests created by successively analyzing and excluding early events, to detect different treatment effects scenarios. A Monte Carlo method is introduced to derive the critical boundaries and derive p-values. The empirical performance were evaluated using extensive simulations studies. These include assessment of the agreement between the nominal and empirical type 1 error rate and an assessment of the empirical power for different types of alternatives, including proportional and non proportional hazards situations of interest.

Results: Versatile test procedures based on maximal combination of supremum logrank statistics provided robust power under proportional hazards and non proportional hazards alternatives. In particular, the look at significant differences at interim stages using the early-emphasis supremum logrank statistics created by the successive analysis of prior events and the combination of early and late-emphasis supremum statistics created by successively removing the earlier events at the later planned analysis times tightly controlled the cumulative type 1 error rates. This sequential testing strategy led to stagewise rejection probabilities comparable to those reported by the reference Maxcombo test under PH, while outperforming its performance in settings of early treatment differences or crossing survivals.

Conclusion: As it is not easy to anticipate the nature of differences in survival beforehand, versatile test procedures based on multiple test statistics are recommended in immuno-oncology trials, where early deleterious effects and delayed therapeutic benefits can occur. As the Maxcombo test, the method combining multiple supremum logrank statistics offer a valid and robust method for detecting treatment differences in survival endpoints.

ORAL CONTRIBUTED SESSION 24
Survival Analysis II

OC24-3 Penalised flexible model for the mortality ratio of patients with a chronic disease compared to the general population

Mathieu Fauvernier[1,2], Emmanuelle Dantony[2,1], Fabien Rollot[3,4,5,6], Emmanuelle Leray[7], Zoé Uhry[8,2], Laurent Remontet[2,1]

1 LBBE UMR CNRS 5558, Université Lyon 1, Villeurbanne, France
2 Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France
3 Université de Lyon, Université Claude Bernard Lyon 1, Lyon, France
4 Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, Hôpital Neurologique Pierre Wertheimer, Hospices Civils de Lyon, Bron, France
5 Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, Lyon, France
6 EUGENE DEVIC EDMUS Foundation against multiple sclerosis, state-approved foundation, Bron, France
7 Univ Rennes, EHESP, CNRS, Inserm, ARENES UMR 6051, RSMS U 1309, Rennes, France
8 Direction des Maladies non Transmissible et des Traumatismes, Santé Publique France, Saint-Maurice, France

Background / Introduction: People with a chronic disease have a higher mortality than the general population and there are two classics frameworks for quantifying this mortality gap. The first assumes an additive decomposition of the overall mortality rate ?O as the sum of the general population rate ?P plus an excess mortality rate (EMR) due to the disease. The second assumes a multiplicative decomposition of ?O as the product of ?P multiplied by a factor called relative mortality (RM). Recently, flexible penalized models were developed for EMR, based on multi-dimensional penalized splines, while only semi-parametric simple models are available for RM. The aim of this work is then to i) propose a flexible penalized model for RM and ii) contrast the two frameworks and show they are consistent provided the models used are flexible enough.

Methods: Building on the penalized parametric framework developed for EMR, we modelled the logarithm of the RM as a function of time and covariates, using penalized natural cubic splines as marginal bases with quadratic penalties to achieve smoothness. Non-linear and time-dependent effects and general interaction can be modelled by forming the tensor product of the marginal bases of time and the covariates. The degree of penalization is controlled by smoothing parameters estimated by either optimizing the Laplace approximate marginal likelihood (LAML) or likelihood cross-validation criterion (LCV). The performance of the proposed flexible RM model is assessed in a simulation study and illustrated using real data.

Results: The penalized RM model offers great flexibility while maintaining good inference properties. The fully parametric form of this model enables us to easily estimate the shape of RM according to time and covariates, providing essential epidemiological information. Because both RM and EMR models are flexible and make no strong assumptions (such as linearity, proportionality, absence of interaction), we show that it is possible to fit either one of them and to retrieve the other one's predictions, which proves that both approaches are consistent.

Conclusion: Thanks to the proposed approach, the multiplicative model benefits from the advantages of the penalized framework, which allows flexibility while avoiding over-fitting. While additive and multiplicative models are often opposed in the literature, we show here that their flexible versions are complementary and describe the same reality, seen from different angles, concerning the mortality of people with chronic disease.

ORAL CONTRIBUTED SESSION 24
Survival Analysis II

OC24-4 A general mixture cure model taking into account uncensored immune individuals

Georgios Karakatsoulis
Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

Background: The Mixture Cure (MC) models constitute an appropriate, attractive and easily interpretable method when studying a time-to-event variable in a population comprised of both susceptible and immune individuals. In the literature, those models usually assume that the immunes are unobservable. However, there are cases in which an immune may be observable. For example, when studying the distant metastasis during the lifetime or the miscarriage during pregnancy, individuals that have died without a metastasis or have given birth are certainly immunes. Common MC models ignore this information and consider them all censored, thus yielding in risk of assigning low immune probabilities to cured individuals.

Methods: In this study, we consider a MC model in which both the susceptible and the immune population comprise of individuals that may be censored or not. That is, apart from the individuals that have experienced the event (susceptible), others may not have experienced it but be considered as uncensored immunes. The objective is to construct a mathematical formulation that manages to examine the potential risk factors associated with both the probability of the event occurrence (incidence component) and the time until that happens (latency component). To this end, we build a general optimization problem based on the likelihood function and the "E-step" of the EM algorithm, by modifying the weights used in the hitherto MC models, the solution of which leads to the maximum likelihood estimators (MLE).

Results: Theoretical results ensure that the proposed correctly assigns zero weights to the uncensored immune individuals. Also, the resulting MLEs tend to the ones obtained through (1) already studied MC models if all immunes are censored, (2) the logit and probit models if all cases are uncensored, and (3) already studied survival models if all cases are susceptible. Hence, the above-mentioned models constitute special cases of the proposed optimization problem. Furthermore, through both theoretical and simulated results the association of the MLEs with the censorship and the rate of the uncensored immunes in the sample are examined.

Conclusions: The proposed model offers an alternative approach for MC models with known uncensored immune individuals. In contrast to existing approaches, this model gives the opportunity to assign zero immune probabilities in individuals that are known to be cured, thus showing improved performance, especially when either the rate of the immunes is high or those immunes display low censoring times.

ORAL CONTRIBUTED SESSION 24
Survival Analysis II

OC24-5 Hazards: Key quantities for analysis, interpretation and understanding of time-to-event data

Jan Beyersmann[1], Claudia Schmoor[2], Martin Schumacher[2]
1 Ulm University, Germany
2 Medical Center Freiburg, Germany

Background/Introduction: Censoring makes time-to-event data special and requires customized statistical techniques. Survival and event history analysis therefore builds on hazards which are the identifiable quantities in the presence of rather general censoring schemes. Because they are conditional quantities, given previous survival, cumulative hazard increments may be estimated based on the current risk set - those still alive and under observation. But it is precisely their conditional nature that has made hazards subject of critique from a causal perspective: A beneficial treatment will help patients survive longer than had they remained untreated. Hence, in a randomized trial, randomization is broken in later risk sets, which, however, are the basis for statistical inference.

Methods: The aim of this talk is to briefly survey this dilemma - after all, mapping analyses of hazards onto probabilities in randomized trials is generally viewed as still having a causal interpretation. We will argue that a causal interpretation is possible taking a functional point of view. We illustrate matters with examples from benefit-risk assessment: Prolonged survival may lead to more adverse events, but this need not mean that the novel treatment has a worse safety profile.

Results: These and other examples illustrate that the situation at hand is conveniently parametrized using hazards, that the need to use survival techniques is not always fully appreciated and that censoring not necessarily leads to the question of "what, if no censoring"? Hence, our concern about hazards is not primarily about causality, but still about the subtleties of interpreting hazards correctly.

Conclusion: We conclude that the discussion should not be about whether hazard contrasts may have a causal interpretation, but how to correctly interpret such a contrast and the realization that the analyses of hazards should routinely be translated onto probabilities.

ORAL CONTRIBUTED SESSION 25
Longitudinal Data Analysis II

OC25-1 Dynamic prediction of individual survival time based on RMST using longitudinal covariates

Chengfeng Zhang[1], Zhaojin Li[1], Zijing Yang[1,2], Baoyi Huang[1], Yawen Hou[3], Zheng Chen[1]
1 Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China
2 Stomatological Hospital, Southern Medical University and Guangdong Provincial Stomatological Hospital, Guangzhou, China
3 Department of Statistics and Data Science, School of Economics, Jinan University, Guangzhou, China

Background: In the field of clinical chronic diseases, common prediction results (such as survival rate) and effect size hazard ratio (HR) are relative indicators, resulting in more abstract information. However, clinicians and patients are more interested in simple and intuitive concepts of (survival) time, such as how long a patient may live or how much longer a patient in a treatment group will live1-4. In addition, due to the long follow-up time, resulting in generation of longitudinal time-dependent covariate information, patients are interested in how long they will survive at each follow-up visit.

Method: In this study, based on a time scale indicator—restricted mean survival time (RMST)—we proposed a dynamic RMST prediction model by considering longitudinal time-dependent covariates and utilizing joint model techniques. The model can describe the change trajectory of longitudinal time-dependent covariates and predict the average survival times of patients at different time points (such as follow-up visits).

Results: Simulation studies through Monte Carlo cross-validation showed that the dynamic RMST prediction model was superior to the static RMST model. In addition, the dynamic RMST prediction model was applied to a primary biliary cirrhosis (PBC) population to dynamically predict the average survival times of the patients, and the average C-index of the internal validation of the model reached 0.81, which was better than that of the static RMST regression.

Conclusion: The proposed dynamic RMST prediction model has better performance in prediction and can provide a scientific basis for clinicians and patients to make clinical decisions.

References: *[1] Blagoev, K. B., Wilkerson, J. & Fojo, T. Hazard ratios in cancer clinical trials—a primer. Nat. Rev. Clin. Oncol. 9, 178-183 (2012). [2] Royston, P. & Parmar, M. K. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat. Med. 30, 2409-2421 (2011). [3] Hasegawa, T. et al. Restricted mean survival time as a summary measure of time-to-event outcome. Pharm Stat. 19, 436-453 (2020). [4] Conner, S. C. et al. Adjusted restricted mean survival times in observational studies. Stat. Med. 38, 3832-3860 (2019).*

ORAL CONTRIBUTED SESSION 25
Longitudinal Data Analysis II

OC25-2 Dynamic random survival forests using functional principal component analysis for the prediction of survival outcomes from time-varying predictors

Corentin Ségalas, Hugues De Courson, Robin Genuer, Cécile Proust-Lima
Univ. Bordeaux, France

Introduction: Individual dynamic prediction of survival events from medical history is statistically challenging, notably because the history includes time-dependent predictors measured irregularly and with error. Regression calibration models, landmark models and joint models have been proposed to this purpose. However the landmark approach does not make use of all available data, the joint models estimation becomes intractable when the number of longitudinal predictors increases, and two-stage regression calibration methods ignore the informative truncation of the longitudinal process due to the event. In this work, we propose to combine functional data analysis tools and random survival forests to predict a survival outcome from longitudinal predictors.

Methods: Random forests are a prediction method which can capture complex and nonlinear relations between predictors and the outcome. It has been extended to handle survival outcome and competitive risks under the framework of random survival forests (RSF). However, they cannot include time-dependent predictors. To this end, an extension of RSF has recently been proposed. For each tree and at each split, individual trajectories of selected longitudinal predictors are summarized by the predicted random effects from mixed models. One advantage is that informative truncation of predictors by the event is taken into account as the nodes become homogeneous. However, this approach is parametric and implies to specify trajectories and random effects distributions. Recently, functional principal component analysis (FPCA) has been extended and implemented to handle sparse and irregular functional data. This non-parametric method projects functional data into a finite functional space whose coordinates along the functional components, the FPCA scores, are a good summary of the individual temporal trajectories? In this work, we propose to summarize longitudinal trajectories using FPCA scores and to split at each node according to these.

Results: A simulation study illustrates the good performances of the methodology implemented in the R package DynForest. The method is also applied to predict the risk of severe complication in patients hourly monitored in neuro-intensive care unit after a subarachnoid hemorrhage.

Conclusion: We developed a fully non-parametric approach to predict clinical events from many irregularly measured predictors. Moreover, because this method is based on random forests, it allows for variable importance assessment and variable selection, opening the machine learning black-box.

ORAL CONTRIBUTED SESSION 25
Longitudinal Data Analysis II

OC25-3 Weight trimming for flexible inverse probability of treatment and intensity weighting

Grace Tompkins, Michael Wallace, Joel Dubin
University of Waterloo, Waterloo, Canada

Background: Many observational studies are subjected to irregular longitudinal data where the observation times vary across individuals. Further, the observation times may be related to the longitudinal outcome. In this setting, we refer to the observation process as informative. Failing to account for the informative observation process may result in biased estimates of the causal effect of a treatment on the longitudinal outcome. Observational data may also not have randomized treatment assignments. In this setting, failing to account for the treatment assignment mechanism may bias any causal estimates. A recently proposed flexible weighting method, which combines inverse probability of treatment weighting (IPTW) and inverse intensity weighting (IIW) through multiplication, can account for both informative observation and non-randomized treatment assignment processes. We refer to this weighting method as the flexible inverse probability of treatment and intensity weighting (FIPTIW) method. However, extreme weights may occur from either weighting method, which may become even more extreme when multiplied together as part of the FIPTIW approach. Currently, there are no studies investigating the impact of weight trimming on such multiplicative weights.

Methods: Through a numerical study, we investigate the impact of extreme weights on FIPTIW. Varying degrees of informativeness are simulated for each of the treatment assignment and observation processes. We then consider two methods of weight trimming where the weights are trimmed either individually prior to multiplication, or trimmed after multiplication. We then apply the methodology to a Malaria data set to show the impact of weight trimming on real data.

Results: Through simulation, we show that weight trimming is beneficial when extreme IPTW weights are produced based on the underlying treatment assignment process. However, when extreme IIW weights were produced because of the informativeness in the observation process, weight trimming did not improve (or hinder) the performance of the FIPTIW method. We recommend trimming FIPTIW weights to the 95th percentile when extreme weights are present. The results also showed that at this percentile, the difference in estimation when trimming FIPTIW weights prior to multiplying compared to trimming after multiplying are negligible. In the real data analysis, failing to trim weights underestimated the average treatment effect of interest.

Conclusions: We recommend weight trimming when extreme weights are present when employing the FIPTIW method for irregular longitudinal data. One can choose to trim weights either before or after multiplying the IIW and IPTW weights together when employing FIPTIW.

ORAL CONTRIBUTED SESSION 25
Longitudinal Data Analysis II

OC25-4 Analysing longitudinal semi-continuous outcomes with excess zeros

Sara Baart[1,2], **Dimitris Rizopoulos**[1,2]
1 Department of Epidemiology, Erasmus MC, Rotterdam, Netherlands
2 Department of Biostatistics, Erasmus MC, Rotterdam, Netherlands

Background/Introduction: Longitudinal Patient Reported Outcome Measures (PROMs) are of great interest in medical research. They allow us to monitor patients' well-being or progress after an illness. Linear mixed effects models (LMM) are often used for these data, however, they are not always suitable. PROMs are often limited to a range, such as 0-10 or 0-100, meaning that the outcome is bounded. Additionally, measurements on the boundary can occur frequently: patients stop experiencing symptoms of the illness, resulting in excess zeros. These data will violate the assumptions of the LMM.

Methods: The motivating data come from the HOMECOMIN study, where 106 patients followed a home monitoring program after a COVID-19 hospitalization. Frequent measures of symptom lists were filled in for six months. We focus on the outcome coughing (0-10). In 42% of the measurements, the outcome was 0 (no coughing symptoms). We contrast the ordinary mixed model with hurdle, Tobit and Tweedie mixed models. Hurdle models are two-part models, in which the zeros are modeled separately from the non-zeros. This two-part model results in two sets of coefficients. Marginal coefficients are introduced, combining the coefficients to aid in interpretation. Tobit models are developed for censored data. Finally, a Tweedie model is estimated on the data. The Tweedie distribution is a special case of an exponential distribution that can have a cluster at zeros. We use the concept of scaled simulated residuals to compare the different models and focus on the interpretability of the results.

Results: The LMM and Tobit model showed clear deviations from the assumptions of the model. Based on the residuals, the Tweedie and hurdle models are suitable but are more complex to interpret. For the hurdle model, the marginal coefficients allow us to conclude that patients start with an average coughing score of 1.6. During follow-up the average score did not change (0.02 increase per week, 95% CI: -0.003-0.05). Additionally, the number of patients with no symptoms did not change during follow-up (OR per week 0.99, 95% CI: 0.92-1.06). For patients with symptoms, the average score also did not change (-0.001 per week on the log scale, 95% CI: -0.017-0.016).

Conclusions: Repeatedly measured PROMs often have bounded outcomes and excess zeros, violating the assumptions of the LMM. Other models are more suitable. In hurdle models, the excess zeros are modeled explicitly. Marginal coefficients overcome the difficulty in the interpretation of the coefficients.

ORAL CONTRIBUTED SESSION 25
Longitudinal Data Analysis II

OC25-5 A variance-component score test for the comparison of gene-set transcriptomic profiles of vaccines

Arthur Hughes[1,2], Denis Agniel[4,5], Rodolphe Thiébaut[1,2,3], Boris Hejblum[1,2]
1 University of Bordeaux, INSERM Bordeaux Population Health Research Center, INRIA SISTM, Bordeaux, France
2 Vaccine Research Institute, Créteil, France
3 CHU de Bordeaux, France
4 Department of Medical Informatics, Harvard Medical School, Boston, United States
5 RAND Corporation, Santa Monica, United States

Introduction: Gene expression is a dynamic process which describes how genetic information, stored in DNA, is transformed into proteins which perform functions in the body. Vaccination induces changes in gene expression which can be measured in the blood mere hours after administration using RNA-sequencing technologies. Comparing these so-called gene expression profiles between vaccines has the potential to reveal insight into their mechanisms of action, providing rationale for future vaccine design. However, the high-dimensionality of RNA-seq data, which typically describes the expression of tens of thousands of genes, leads to issues with both the modelling and interpretation of the data. Therefore, between-vaccine comparisons can be performed at the gene set level, where the units of investigation are defined as groups of genes performing the same biological function. These gene-set approaches have the advantage of combining small signals from each gene, yielding greater statistical power to detect differences, as well as providing results with natural biological interpretations.

Methods: We present a variance-component score test for the comparison of vaccine gene-set profiles. The mathematical framework is based on a linear mixed-effects model, which allows for the encoding of complex experimental designs and the investigation of countless biological hypotheses. The test results in both quantitative (p-values) and qualitative (visualisation) outputs for 1 comparative analyses.

Results: We show desirable properties of our test with simulated data, including a high statistical power and strict control of the false positive rate. The test is demonstrated on public systems vaccinology datasets (Diray-Arce et al. [2022]) in a comparison of longitudinal gene-set profiles of yellow fever (YF17D) and influenza (TIV) vaccines. We show biologically interpretable differences, such as an upregulation in interferon activity after vaccination with YF17D compared to TIV, and distinct dynamics of the inflammation response.

Conclusion: Our statistical test utilises gene sets for formal quantitative and qualitative comparisons to be made between molecular profiles.

References: *[1] Joann Diray-Arce, Helen E. Miller, Evan Henrich, Bram Gerritsen, Matthew P. Mul`e, Slim Fourati, Jeremy Gygi, Thomas Hagan, Lewis Tomalin, Dmitry Rychkov, and et al. The immune signatures data resource, a compendium of systems vaccinology datasets. Scientific Data, 9(1), 2022. doi: 10.1038/s41597-022-01714-7.*

ORAL CONTRIBUTED SESSION 26
Prediction and Prognostic Models II

OC26-1 Externally validating clinical prediction models for recurrent events

Alexandra Hunt[1], Thomas Spain[1], Laura Bonnett[1], Tony Marson[1], Hein Putter[2]
1 University of Liverpool, Liverpool, United Kingdom
2 Leiden University, Leiden, Netherlands

Background/Introduction: Prediction models for repeated (recurrent) medical episodes such as seizures and asthma attacks are increasingly being developed. Validation, including in an independent dataset, is necessary to evaluate the performance of any prediction model prior to use in clinical practise. Performance is evaluated via calibration and discrimination.

A submitted systematic review of recurrent event prediction models identified that of the 301 included studies, only 75 (24.9%) internally validated their model(s) and three (1.0%) validated their model(s) in an external dataset.

This may be due to difficulties adapting existing calibration and discrimination methodology to the recurrent event setting. Calibration of recurrent event models can be evaluated using predicted and observes counts, rather than probabilities. However, due to the dynamic nature of recurrent models, the discriminative ability of the model is likely to vary over time and most current discrimination methods, do not take this into account.

We therefore propose a method of evaluating discrimination of a recurrent event model and demonstrate its use, alongside evaluation of calibration, in recurrent event models predicting an individual's epilepsy seizure count.

Method: Data from the SANAD study was used to build two prediction models to estimate future seizure counts in a proposed timeframe. Models were built using the Andersen and Gill (AG) and Prentice, Williams & Petersons (PWP) variants of Cox's proportional hazards model. To validate the models, data from SANAD II, containing 1510 patients with newly diagnosed epilepsy, will be used. SANAD and SANAD II have the same design but considered mostly different treatments. The models will be evaluated using calibration methods including modified calibration plots (event counts rather than probabilities) and observed over expected count ratios. Our proposed discrimination method will evaluate the concordance of pairs of individuals in the dataset according to their observed and predicted event counts.

Results: Initial work shows better model calibration in SANAD II from the PWP SANAD model than the AG SANAD model. Discrimination evaluation is ongoing at the time of submission but full results will be presented at the conference. We believe that our proposed method of discrimination can be developed into statistical software to make external validation of recurrent event models easier to conduct.

Conclusion: Our proposed discrimination method and our suggested modification to existing calibration models ensure prediction models for recurrent events can be externally validated. This is a vital step towards improving care for people with conditions typified by repeated episodes.

ORAL CONTRIBUTED SESSION 26
Prediction and Prognostic Models II

OC26-2 Understanding algorithmic fairness for clinical prediction in terms of net benefit and health equity

Jose Benitez-Aurioles[1], Alice Joules[2], Irene Brusini[2], Niels Peek[3], Matthew Sperrin[1]
1 University of Manchester, Manchester, United Kingdom
2 IQVIA, London, United Kingdom
3 University of Cambridge, Cambridge, United Kingdom

Background: There are increasing concerns about the fairness of clinical prediction models across sensitive attributes, such as ethnicity or gender. Net benefit (NB) is a measure of the clinical utility of a model through a weighted sum of its true and false positives. We propose assessing the fairness of a prediction model by expanding NB to quantify and compare the clinical impact of a model in each subgroup.

Methods: We extend the formula for NB to allow for subgroup comparison, through an additional inequality term, and account for the value of well-calibrated models through a weighted area under a rescaled decision curve. We name this metric subgroup net benefit (sNB) and measure it in units of true negatives (TNs) per 10,000 persons. The overall and minimum sNB across all subgroups can be maximised together without trade-off, as long as there are no resource constraints.

We showcase the metric through an example predicting type 2 diabetes across ethnicities (Asian, Black, Other and White). We compare the sNB of three models, fitted to a prospective cohort of UK Biobank participants: a logistic regression model in which ethnicity is excluded as a predictor (NoSA), one in which it is included (SingleSA), and an ensemble of propensity-weighted subgroup-specific regressions (MultiSA) [1]. These show three levels of 'subgroup personalisation', which should in theory progressively improve the sNB in each ethnicity by reducing statistical bias due to subgroup heterogeneity. The models are compared to the default policy of identifying no patients (NoOne).

Results: The sNB was highest in patients of White ethnicity (for NoOne, NoSA, SingleSA and MultiSA: 9807, 9820, 9821 and 9821 TNs per 10,000 persons) and lowest for patients of Asian ethnicity (9347, 9389, 9443 and 9442). The three models reduced the gap in sNB between the two populations compared to the default policy, suggesting that their introduction could reduce health inequalities by 'levelling up' the most disadvantaged subgroups. SingleSA and MultiSA performed equally well overall, except for better performance of the latter in patients of Black ethnicity (9585 and 9600).

Conclusion: We propose an approach to quantify the fairness of clinical prediction models by extending the definition of net benefit. This approach considers clinical context to better understand the role of predictive models in upholding health equity.

References: *[1] Steingrimsson, J.A. et al. (2023). American Journal of Epidemiology, 192(2), pp.296–304.*

ORAL CONTRIBUTED SESSION 26
Prediction and Prognostic Models II

OC26-3 Conceptualisation and visualisation of multivariable models for the analysis of biomarkers for patient selection in clinical trials

Nathalie Weidinger
Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

Background/Introduction: In clinical trials, the use of prognostic and predictive biomarkers is crucial for enhancing patient selection and advancing precision medicine. Prognostic biomarkers capture disease progression, while predictive biomarkers identify patients who will benefit most from a treatment. For clinical drug development it is essential to discern the prognostic and/or predictive value of a biomarker.

Current methods for biomarker identification are available as separate implementations with distinct requirements and differently structured results, complicating their comparison and combination. So far, no comprehensive investigation regarding assessing of biomarker methods in specified scenarios exist.

Methods: We developed an R-based framework that integrates several multivariable methods to assess the prognostic and predictive potential of biomarkers and to facilitate a harmonised visualisation of biomarker data across methods. We conducted a comprehensive investigation of promising methods under various simulation scenarios including different complexities and characteristics. Based on scores and/or a selection of biomarkers the methods return, the evaluation across methods could be performed for the simulated datasets, which are generated with pre-specified prognostic and predictive biomarkers.

In total, seven methods for prognostic and nine for predictive biomarker identification have been implemented and evaluated, focusing on binary and continuous biomarkers and responses, respectively. Investigated methods are Knockoffs, VirtualTwins, SIDES, predMOB, PPLasso, INFO+, VSURF, the machine learning methods XGBoost (with Shapley values, variable importances, Shapley interaction values, Friedman's H-Statistic) and GLMNet (with Shapley values, variable importances).

Results: The results of the selected simulations reveal that, contingent on the scale of the biomarkers and response (binary/continuous), only certain methods provide satisfactory results regarding correct biomarker identification (e.g. predMOB, VSURF and machine learning methods). Generally, methods to identify prognostic biomarkers yielded more reliable results, while methods investigating predictive biomarkers did not consistently exhibit expected results, especially in the case of a binary response or lower sample sizes.

Conclusion: The developed framework is the first step in systematically rating biomarker methods in different scenarios. With the conducted simulations guidance is provided regarding the selection of methods when analysing biomarker data of clinical trials.

Our work provides a perfect basis for further research, simulations and evaluations for different sample sizes, other effect types and parameter settings of the methods.

With specific characteristics and assumptions of a planned trial, simulated data can be evaluated with the presented framework to find the methods providing most accurate results for this use case.

ORAL CONTRIBUTED SESSION 26
Prediction and Prognostic Models II

OC26-4 Comparing modern machine learning methods for predicted individual treatment effect

Pamela Solano[1], Thomas Jaki[1,2]
1 Faculty for Informatics and Data Science, FIDS, Universität Regensburg, Regensburg, Germany
2 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom

In precision medicine dealing with biomarker uncertainty is a prerequisite to accurately capture an individual's disease pathway. To achieve this, predictive models should address population heterogeneity, complex structures and high-dimensionality to estimate the Predictive Individual Treatment Effects, PITE. PITE is a simple and intuitive way to estimate individual treatment effects and consists of the difference between treatment and control prediction for each individual [1,2]. This construction implies that any approach to prediction can be applied, but performance comparison is not straightforward since the true individual treatment effect is not observed. In this work we compare state-of-the-art Statistical models and Machine Learning algorithms in their capacities to estimate the PITE. We compare many different methods classified into regularization approaches, such as Lasso, Ridge, Elastic Net and Neural Networks, Non-linear/complex structure such as GAM, Random Forest or BART and others and investigate these methods in ideal conditions and explore their limitations. Different performance metrics are presented for all methods based on a series of simulated data encompassing multiple profile, different sample size, and independent and correlated biomarkers. The risk (Expected Squared Error) and sensitivity [3] to detect if the methods can correctly estimate whether a treatment is beneficial or not to a patient, regardless of the magnitude were calculated. We find that for independent biomarkers and reasonable sample size, all methods achieved small risk (< 0.05) and more than 0.95 sensitivity. However, for small sample size (< 100), only Bayesian regularization achieved small risk. For more complex biomarker relationships on the other hand risk is increased (0.05-0.15) while sensitivity remained high. In these situations procedures such as artMachine, blackboost, and spikeslab performed particularly well. Although many methods are compared in this work, others can be considered as well. The results can be used as insights for designing clinical trials with reproducible biomarkers and improve drug labeling in precision medicine. BartMachine, blackboost and Bayesian regularized Neural Networks algorithms showed good performance in higher order interactions since by nature they are prepared to work in context where non-linearity, high-dimensional are present in the data. In addition to comparing methods, the results can be used to detect outlying individuals, in the sense that they are hard to predict by all methods. Machine learning algorithms bring about accurate and fully interpretable PITEs can effectively support clinicians in the personal diagnosis (treatment guided) decision process transforming the patient care.

References: *[1] A. Lamont, M. D. Lyons, T. Jaki, E. Stuart, D. J. Feaster, K. Tharmaratnam, D. Oberski, H. Ishwaran, D. K. Wilson, and M. L. V. Horn, "Identification of predicted individual treatment effects in randomized clinical trials," Statistical Methods in Medical Research, vol. 27, pp. 142–157, Mar. 2016. [2] T. Jaki, C. Chang, A. Kuhlemeier, and M. L. Van Horn, "Predicting individual treatment effects: Challenges and opportunities for machine learning and artificial intelligence," KI - K¨unstliche Intelligenz, Jan. 2024. [3] N. M. Ballarini, G. K. Rosenkranz, T. Jaki, F. König, and M. Posch, "Subgroup identification in clinical trials via the predicted individual treatment effect," PLOS ONE, vol. 13, p. e0205971, Oct. 2018. 1*

ORAL CONTRIBUTED SESSION 26
Prediction and Prognostic Models II

OC26-5 Improving prediction models by incorporating external data with weights based on similarity

Max Behrens[1,5], Maryam Farhadizadeh[1,5], Angelika Rohde[2], Alexander Rühle[3,4], Nils Nicolay[4], Harald Binder[1,5]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany
2 Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany
3 Department of Radiation Oncology, University of Freiburg–Medical Center, Freiburg, Germany
4 Department of Radiation Oncology, University of Leipzig, Leipzig, Germany
5 Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany

Introduction: In medical settings, constructing prediction models from limited observational data is a common challenge. This scenario often arises in multi-center studies, where data from a particular medical center may represent only a subset of the larger data set. Given the potential large differences between centers, there is a need to develop prediction models tailored to the data from a specific target center in question. To mitigate the constraints posed by a small sample size, it is beneficial to leverage information from external data sets, namely, those available from other centers.

Methods: There exist methods that either allocate weights to each external data set as a whole or to individual observations within those data sets. To integrate both information more comprehensively regarding similarities among data sets and individuals, we present a strategy that merges the two aspects into weights. These weights are subsequently utilized within a likelihood for the estimation of regression models. Individual weights are based on the probability of belonging to the target center given by a logistic regression model. To account for similarity on the data set level, we adjust individual weights by the inverse probability of the logistic regression models to distinguish a randomly drawn observation of the target center from a randomly drawn observation from an external center. This approach is essentially characterized by a form of inverse probability weighting.

Results: To set the stage, we outline various scenarios in which the relationship between covariates and the outcomes vary across data sets. These scenarios then guide the design of a simulation study based on structural equation models with latent variables. Within the simulation, we also highlight the utility of the effective sample size in comparing different approaches to subgroup modeling. Our method is further evaluated in a real-world scenario, focusing on the prediction of radiotherapy doses for elderly patients with head and neck squamous cell carcinoma. As with the simulation study, our proposed method demonstrates superior predictive accuracy in instances where the external data sets exhibit similarities. The degree of similarity between data sets is assessed through the dataset-level component of our weighting scheme.

Conclusion: To sum up, we introduce a method for measuring the similarity between the target data set and other data sets. This measurement of similarity is then utilized to incorporate observations from these additional data sets, enhancing the accuracy in prediction modeling tasks focused on a specific subset with limited data.

ORAL CONTRIBUTED SESSION 27
Competing Risks II

OC27-1 Mixture cure cause-specific hazard Cox models for competing risks data with partly interval censoring

Serigne Lo[1,2], Joseph Descallar[3,4,5], Houying Zhu[3], Jun Ma[3]
1 The University of Sydney, Sydney, Australia
2 Melanoma Institute Australia, Sydney, Australia
3 School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia
4 Ingham Institute for Applied Medical Research, Liverpool, Australia
5 South West Sydney Clinical Campuses, UNSW, Liverpool, Australia

In medical studies, competing risks survival data are commonly estimated using partial likelihood methods, with the assumption that any right-censored patient will experience one of the competing event beyond the conclusion of the study period. In some cases, however, a patient may be considered "cured" from the risks of interest, meaning that none of the risks will occur, resulting in a cured fraction. Furthermore, if disease-progression is the event of interest, the exact event times are unknown and often subject to interval censoring. When dealing with such data, employing the standard analysis approach based on cause-specific hazards models while ignoring the cured fraction, could lead to biased parameter estimates. This article introduces a novel approach that accommodates for interval censoring and a cured fraction within cause-specific Cox models when analsing competing risks data. More specifically, we propose a new maximum penalized likelihood approach to simultaneously estimate logistic regression parameters for the cured fraction, cause-specific Cox models regression coefficients, and their non-parametric baseline hazards. Asymptotic properties are developed, and simulation studies show reduced bias and improved coverage probability compared with the partial likelihood approach with a mid-point imputation. The new method was applied on a phase III two-arm randomised clinical trial data for advanced melanoma patients. The primary outcome is defined as time to intracranial failure competing with death (without disease progression).

ORAL CONTRIBUTED SESSION 27
Competing Risks II

OC27-2 Joint frailty modelling of multiple time-to-events and longitudinal measures

Jagathnath Krishna Kumarapillai Mohanan Nair, Divya Dennis, Anand Hari
Regional Cancer Centre, Thiruvananthapuram, Thiruvananthapuram, India

Background: In caner survival studies, disease progression can be assessed with longitudinal study designs where the patients are observed over time and the covariates information (biomarkers (carcinoembryonic antigen -CEA)) are measured repeatedly during the follow up period. In these studies, the primary interest is on time-to-event, which might be explicit, eg. death, recurrence, metastasis, etc. Modeling of longitudinal data on survival outcomes can be inefficient due to its correlated structure. In such situation multi-state models, which can account the correlated structure of the multiple events is useful. Apart from this correlated structure, there may exist random heterogeneity between these events. Hence in the present study we derived joint multi-state frailty models (JMSFM) for estimating the risk for multiple events simultaneously and to access the heterogeneity between the events.

Methodology: For accounting heterogeneity, longitudinal outcome and multiple time-to-events, we derived a JMSFM. The longitudinal sub-model was modeled using linear mixed model and the survival sub-model using multi-state model (MSM) and multi-state frailty model (MSFM). We assumed frailty parameters to follow either gamma or log normal distributions. The latent variable was used to link the longitudinal and time-to-event sub-models. The parameters were estimated using maximum likelihood estimation. The derived models were illustrated using colon cancer patient data. The covariate considered for risk prediction were, composite stage, lymph node involvement, T4, age, sex, PNI, LVE and comorbidity; and CEA as longitudinal outcome. The information criteria, AIC, BIC and AICc were used for model identification.

Results: A joint multi-state models and two joint frailty models (joint multi-state gamma frailty model and joint multi-state log normal frailty model) with four states and six transitions were derived. Based on information criterion, JMSFM were found to be better predictive than JMSM. The present study identified that PNI (transition from diagnosed as disease to death), Composite Stage (transition from recurrence to death; transition from metastasis to death) lymph-node involvement and T4 along with the longitudinally measured CEA value as significant prognostic factors for predicting the multiple time-to-events based on the proposed JMSFM and also found that for each transition state, the longitudinal observation (CEA) has strong association with corresponding survival events (η ranges from 1.25 to 1.44).

Conclusion: Thus we conclude that joint multi-state frailty model as a better model for simultaneous prediction of multiple events in presence of random heterogeneity in a longitudinal study design.

ORAL CONTRIBUTED SESSION 27
Competing Risks II

OC27-3 Non-parametric frailty model for the natural history of prostate cancer; using data from a screening trial

Ilse Cuevas Andrade[1], Ardo van den Hout[1], Nora Pashayan[2]
1 University College London, London, United Kingdom
2 University of Cambridge, Cambridge, United Kingdom

Background: Prostate cancer is the second most diagnosed cancer in men. The prostate-specific antigen (PSA) test measures the PSA levels in the blood. A PSA level above 4.0 ng/mL can be considered abnormal. However, different factors can cause the PSA level to fluctuate. The PSA levels can be used to assess if further clinical investigation is required, such as imaging or biopsy studies.

In this study, we consider panel data from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. The PLCO trial included 76,685 men aged 55 to 74 years, who were randomised in the screening and control group. The screening group received an annual PSA test for 6 years.

Methods: We aim to extend the four-state model developed by Bhatt et al (2021) to account for unobserved heterogeneity. Following a multi-state framework, a mixed-effect model for survival can be fitted, known as the frailty model. The frailty term, as a random effect, can be specified for an individual or can be cluster-specific. One way to specify this type of model is by assuming a parametric frailty distribution, e.g. log-normal. However, it can be computationally expensive and can become infeasible in a high-dimension setting.

Conversely, a non-parametric frailty approach removes the assumption of normality for the random effects, and it can be less demanding computationally. The non-parametric frailty model can then be used to inform screening frequency before transitioning to a clinically diagnosed state.

Results. We consider unobserved heterogeneity by fitting a non-parametric frailty model with a class-specific intercept for the transition from being healthy to screen-detectable. The non-parametric frailty model seems to capture unobserved heterogeneity across the patients.

Conclusion: Further extension of this model will be explored accounting for time-dependent hazards and by linking the latent class to the longitudinal biomarker. The model will account for interval, right-censored, and left-truncated data.

References: *[1] Bhatt R, van den Hout A, Pashayan N. (2021). A multi-state survival model of the natural history of cancer using data from screened and un-screened population. Statistics in Medicine, 40, 3791 – 3807.*

Acknowledgements: *Research reported in this publication was supported by the National Cancer Institute (NCI) of the National Institutes of Health under Award Number U01CA253915. The authors thank the National Cancer Institute for access to NCI's data collected by the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.*

ORAL CONTRIBUTED SESSION 27
Competing Risks II

OC27-4 Dynamics of excess mortality during the Covid-19 pandemic in the Netherlands – Impact of vaccination and infection analysed through a multi-state model incorporating relative survival

Liesbeth de Wreede[1], Marije H. Sluiskes[1], Damjan Manevski[2], Hein Putter[1], Eva A.S. Koster[1]
1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
2 Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Background/Introduction: The Covid-19 pandemic has led to excess mortality. Notably, the reported numbers of excess deaths are higher than the numbers of deaths from Covid-19. Evaluating pandemic-related mortality should therefore not only be based on cause-of-death data. Most studies show the highest excess mortality in persons aged 65 years or older. The protective effect of anti-Covid-19 vaccination against death from Covid-19 was lower for older persons and also lasted shorter than for younger persons. We investigated the impact of infection with and vaccination against Covid-19 on excess mortality during 2020-2021 in persons aged 65 years or older in the Netherlands.

Methods: We incorporated relative survival modelling into an extensive multi-state model considering vaccination, the acute and post-acute phase of (re)infection, and death based on the recent methodology by Manevski et al (2022). The key assumption of relative survival is that the observed hazard of death is the sum of the (unobserved) background and excess hazards where the former is derived from population tables or other historical data. The multi-state model is a time-inhomogeneous Markov model, where time is calendar time, stratified by sex and age. In this model, transition probabilities can be estimated by the Aalen-Johansen estimator. We applied the model on real-world, nationwide and unselected data from Statistics Netherlands (CBS).

Results: At the end of 2021, starting from 1 January 2020, the total cumulative probability of excess mortality was 0.3%, 4% of all mortality. This percentage was markedly higher for older persons, especially men, both in absolute terms and as a percentage of total mortality, e.g., for men above 90 years excess mortality was 4.8%, 10% of all mortality. Almost all excess mortality took place after an infection but its probability was much lower for persons who had received a vaccination. Vaccinated persons who did not become infected experienced a negative excess hazard, implying their survival was better than during the pre-Covid-19 years.

Conclusions: The novel multi-state model incorporating relative survival enables to split all mortality in background and excess mortality with and without intermediate events. The current application shows the value outside the traditional context of relative survival where the population component of the mortality of cancer patients is modelled. Further extensions incorporating negative hazards and regression modelling of background and excess hazard are under development.

References: *[1] Manevski D, Putter H, …, de Wreede LC. Integrating relative survival in multi-state models-a non-parametric approach. Stat Methods Med Res. 2022;31(6):997-1012.*

ORAL CONTRIBUTED SESSION 27
Competing Risks II

OC27-5 Assessing treatment effects with adjusted restricted mean time lost in observational competing risks data

Haoning Shen[1], Chengfeng Zhang[1], Yu Song[1], Zhiheng Huang[1], Yanjie Wang[1], Yawen Hou[2], Zheng Chen[1]
1 Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China
2 Department of Statistics and Data Science, School of Economics, Jinan University, Guangzhou, China

Introduction: According to long-term follow-up data of malignant tumor patients, assessing treatment effects requires careful consideration of competing risks. The commonly used cause-specific hazard ratio (CHR) and sub-distribution hazard ratio (SHR) measure the relative values of the instantaneous risk rate between groups and may exist challenges in terms of proportional hazards assumption and clinical interpretation. Recently, the restricted mean time lost (RMTL) has been recommended as a supplementary measure for better clinical interpretation. Moreover, for observational study data in epidemiological and clinical settings, due to the influence of confounding factors, covariate adjustment is crucial for determining the causal effect of treatment.

Methods: In this study, for competing risks data, we used the inverse probability weighting method to develop point and interval estimates of weighted RMTL difference, allowing for covariate adjustments. Furthermore, we further consider the changes in treatment effects over time by constructing a dynamic RMTL difference curve and simultaneous confidence bands.

Results: In simulations, the adjusted RMTL difference is approximately unbiased, and the interval estimate is robust. Then this method was applied to real-world cervical cancer patient data, revealing improvements in the prognosis of patients with small cell carcinoma of the cervix. The results showed that the protective effect of surgery was significant only in the first 20 months, but the long-term effect was not obvious. Radiotherapy significantly improved patient outcomes during the follow-up period from 17 to 57 months, while radiotherapy combined with chemotherapy significantly improved patient outcomes throughout the entire period.

Conclusion: In conclusion, we propose this approach that is easy to interpret and implement for assessing treatment effects in observational competing risk data.

References: *[1] McCaw ZR, Claggett BL, Tian L, et al. Practical Recommendations on Quantifying and Interpreting Treatment Effects in the Presence of Terminal Competing Risks: A Review. JAMA Cardiol. 2022;7(4):450-456. [2] Wu H, Zhang C, Hou Y, Chen Z. Communicating and understanding statistical measures when quantifying the between-group difference in competing risks. Int J Epidemiol. 2023;52(6):1975-1983*

ORAL CONTRIBUTED SESSION 28
Electronic Health Data

OC28-1 Predictive modelling of longitudinal statin adherence trajectories with functional data analysis and machine learning in Finnish nationwide cohort

Andrea Corbetta[1,2,3], Matteo Ferro[3], Mattia Cordioli[3], Markus Perola[4], Francesca Ieva[1,2], Andrea Ganna[3]
1 MOX - Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Milan, Italy
2 HDSC - Health Data Science Center, Human Technopole, Milan, Italy
3 FIMM - Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
4 THL - Finnish Institute for Health and Welfare, Helsinki, Finland

Introduction: Poor adherence to medication poses a significant barrier to reaping the advantages of pharmacotherapies. Thus, it is crucial to pinpoint factors linked to and predictive of inadequate adherence to prescribed treatment. In the Finregistry project [2], we examined five-year-long trajectories of adherence to statins in 507,183 individuals in Finland. We propose Functional Data Analysis methods to detect groups of decreasing trajectories. With socioeconomic and medical registries from Finregistry, we leverage ML methods to predict individuals with negative trends based on baseline information, offering insights for tailoring treatment and implementing monitoring.

Methods: We defined trajectories as consequent measures of the Medication Possession Ratio (proportion of time where the medication supply is available, defined between 0 and 1) from the first recorder purchase of statin therapy. We smoothed the trajectories with 5-knotted penalized cubic splines, computed Functional Principal Components and performed K-Means clustering on the FPC scores. We trained seven prediction models (Logistic Regression, Lasso Regression, Nearest Neighbour, Naïve Bayes, Random Forest, XG-boost RF, Single-Layer-Perceptron) to distinguish between good adherence and declining adherence with 41 socioeconomic, demographic and geographic predictors.

Results: We identified six trajectories' groups: one showed a declining trend (A; n= 31,885), one an increasing trend, two of them an oscillating trend within the 0.8 and 1 band, one a steady trend around 0.5, and one a steady trend around one (B; n=271,552).

The models tend to predict most observations as belonging to group B; in terms of Matthews Correlation Coefficient, the best classifier is the SLP (MCC=0.038), while the one that classifies best the actual declining adherence individuals is the Naïve Bayes classifier (TPR=0.32).

Nonetheless, the prediction models offer compelling results on which factors impact the prediction of a declining adherence pattern. We have a different mother tongue than Finnish/Swedish (OR=1.85[1.58-2.18]), being divorced (OR=1.30[1.23-1.37]), living in rural areas (OR=1.14[1.04-1.24]), the number of children (OR=1.09[1.07-1.1]), and receiving income support (OR=1.15[1.04-1.26]) which harm adherence while worse health (Charlson comorbidity score OR=0.97[0.95-0.98), being married (OR=0.94[0.91-0.98], increased age (OR=0.73[0.69-0.77]), time in the long term care (OR=0.77[0.73-0.81]) and the treatment

being for secondary prevention (OR=0.78[0.75-0.82]) are associated with the steady trend. Statistical significance was tested with FDR correction.

Conclusions: Detecting patients who decline statin use over time was proven not to be trivial (median AUC=0.57), even for reasonably complex models and accounting for low prevalence. Nevertheless, investigating socioeconomic and demographic factors associated with declining adherence gives insights into which individuals are most likely to need treatment monitoring.

ORAL CONTRIBUTED SESSION 28
Electronic Health Data

OC28-2 Emulation of target cluster trials of complex interventions: Estimands, methods and application

Clemence Leyrat[1], **Camille Maringe**[2]
1 Department of Medical Statistics, Inequalities in Cancer Outcomes Network, London School of Hygieneand Tropical Medicine, London, United Kingdom
2 Department of Health Services Research and Policy, Inequalities in Cancer Outcomes Network, London School of Hygiene and Tropical Medicine, London, United Kingdom

Background: The emulation of target trials from observational data has become a popular framework for the estimation of causal treatment effects in the absence of – or in complement to – randomised trials (RCTs). Target trial emulation allows a better transparency in the analysis of observational studies via a thorough description of the population, eligibility criteria, estimands and statistical methods. Although most of the published methodological work on emulated trials and their applications describe the emulation of individually RCTs, many complex interventions in public health are implemented at cluster-level (e.g. at hospital or regional level). We propose an extension of the framework to the emulation of cluster-RCTs, describe statistical methods for their analysis and compare them empirically for the estimation of causal cluster-specific and population-average effects.

Methods: Linking the Census of the Cancer Workforce in England, conducted by MacMillan in 2017, to cancer registry data, we emulated a target cluster trial evaluating the benefits of the provision of cancer nurse specialists (CNS) at Trust level on access to curative treatment for patients diagnosed with early-stage lung cancer. Mixed models were used to estimate cluster-specific effects and population-average after marginalisation on the distribution of the covariates and random effects. These estimates were compared to those obtained from weighted and unweighted GEEs, as well as from weighted and unweighted cluster-level analyses. Heterogeneity by subgroup was also investigated.

Results: Overall, among the 130 included Trusts, 99 had a provision of CNS below target (1 CNS/80 lung cancer patients; control arm) and 31 above target (intervention arm). There was no effect of the provision of cancer nurse specialists on the primary outcome (OR= 0.94 [0.69;1.28]), although the results suggested the presence of informative cluster size, with a larger effect of the intervention in smaller hospitals (OR= 1.24 [0.81;1.89] vs 0.68 [0.45;1.03] in larger hospitals). This informative cluster size had implications for the choice of the analysis

Method: GEE with an independence working correlation structure or cluster-level methods are more robust in such settings. Furthermore, the investigation of interactions between intervention and cluster-level factors enabled the identification of inequalities at cluster-level, also characterised by a strong estimated intracluster correlation of 0.11.

Conclusion: Cluster-RCT emulation is a promising approach for the estimation of the causal effect of complex cluster-level interventions from observational data. We provide practical recommendations to researchers for its implementation and analysis, illustrate and suggest solutions to common challenges.

ORAL CONTRIBUTED SESSION 28
Electronic Health Data

OC28-3 Linking real world data to assess post-acute mortality following COVID-19 infection compared to matched test-negative people in England: Findings from the ECHOES national dataset

Hester Allen[1,2], Katie Hassell[2], Francesca Bell[2], Marti Catala-Sabate[1], Annika Jodicke[1], Albert Prats-Uribe[1], Gavin Dabrera[2], Daniel Prieto-Alhambra[1], Ines Campos-Matos[2]

1 Health Data Sciences, Centre for Statistics in Medicine, University of Oxford, United Kingdom
2 COVID-19 Vaccines and Epidemiology Division, UK Health Security Agency, United Kingdom

Background: Mortality risk in the acute phase of SARS-CoV-2 infection has been well characterised, however longer-term risks associated with COVID-19 remain poorly understood. Evaluation of Post-Acute COVID-19 Health Outcomes (ECHOES) dataset, a nationwide cohort of individuals testing for SARS-CoV-2 comprising multiple linked level real-world data is built to investigate post-acute COVID-19 outcomes in England, and to use these data to estimate risk of post-acute death >28 days to 1 year after SARS-CoV-2 infection, relative to those testing negative.

Methods: National COVID-19 test data was linked to vaccine registry, Hospital Episode statistics data, and death registration data. Positive and negative SARS-CoV-2 tests from May 2020 to April 2022 were matched 1:1 on week of earliest test, sex, age band, index of multiple deprivation and area of residence. Follow up was censored at end of study and death date, with negatives further censored if subsequently testing positive. Individuals could contribute to both testing groups. Rate of death was calculated, and Cox proportional hazards models stratified by matched sets calculated risk after adjustment for ethnicity, acute severity and comorbidity indicators. Severity was determined by hospital attendances within -7 to +14 days of test, categorised as A&E attendance, hospital admission, or critical care, prioritising most severe category. Attendances for planned procedures, pregnancy, injury, or poisoning- were dismissed. Comorbidities diagnosed prior to test were categorised using the Charlson comorbidity index (CCI) as mild, moderate, or severe.

Results: The matched cohort comprised 8,419,205 positives and 8,419,205 negatives. Post-acute rate of death was 9.21 per 1,000 in positives and 6.05 in negatives, with an absolute rate difference (ARD) of 3.16. The unadjusted Cox model produced a Hazard Ratio (HR) of 1.68 (95% CI:1.66-1.70). After adjustment, the HR attenuated to 1.39 (95% CI: 1.37-1.41). The highest adjusted HRs were in females (HR:1.51, 95% CI:1.48-1.30; ARD 3.49), and those aged over 80 (HR: 1.81, 95% CI:1.78-1.85; ARD:113.26).

Conclusion: In this large, national cohort, we see a 40% increase in risk of post-acute death up to one year after testing positive for SARS-CoV-2, accounting for comorbidities and demographic and disease characteristics. This association is stronger in women and older people. These results showcase how this cohort can facilitate the identification of other COVID-19 health outcomes and their predictors and inform the burden of long-term COVID-19 sequelae. These findings emphasise the importance of continued research into long-term COVID-19 morbidity and mortality to inform planning and provision of healthcare services.

ORAL CONTRIBUTED SESSION 28
Electronic Health Data

OC28-4 Emulation of an SLCO1B1*5-guided target trial to guide statin therapy

Faye Baldwin, Ruwanthi Kolamunnage-Dona, Andrea Jorgensen
Department of Health Data Science, University of Liverpool, Liverpool, United Kingdom

Background: Statins are one of the most widely prescribed treatments worldwide for the prevention of cardiovascular disease (CVD). Despite the benefits of statin use on cardiovascular outcomes, adherence to statin therapy is considered suboptimal. Statin associated muscular symptoms (SAMS) are considered a primary reason for statin discontinuation, and can range from mild muscular pain to more severe conditions such as rhabdomyolysis. Several risk factors for SAMS have been identified, including old age, low BMI and the SLCO1B1*5 gene polymorphism. SLCO1B1*5 is a single nucleotide polymorphism resulting in decreased function of the SLCO1B1 gene. As a result, SLCO1B1*5 carriage has been associated with increased risk of SAMS, most consistently in simvastatin and atorvastatin. While the effect of delivering SLCO1B1 pharmacogenetic results has been investigated, there have been no randomised trials evaluating the benefits of guiding statin choice according to SLCO1B1*5 genotype [1]. Here, we utilise a target trial emulation approach to emulate a trial of SLCO1B1*5-guided statin treatment.

Methods: The trial utilises data from UK Biobank, including a prospective cohort of 500,000 participants aged 40-69 years. Firstly, we investigate the difference in risk of SAMS between those commencing simvastatin versus those commencing rosuvastatin in SLCO1B1*5 carriers by emulating a genotype-guided enrichment trial. Secondly, we emulate a genotype-strategy design to compare a SLCO1B1*5-guided treatment strategy versus a non-genotype guided strategy. Inverse probability of treatment weighting is used to adjust for baseline and time-varying confounding, and inverse-probability of censoring weights is used to adjust for selection bias due to loss to follow-up when comparing the treatment effect.

Results: We demonstrate how observational data from UK Biobank can be utilised to emulate a genotype-guided target trial of a personalised approach to statin treatment. The estimated effects are presented in the form of crude and adjusted odds ratios comparing risk of SAMS in atorvastatin and rosuvastatin users amongst SLC01B1*5 in the enrichment trial, and between the genotype-guided and non-guided arms in the genotype strategy trial.

Conclusion: The proposed approach can provide a cost-effective alternative to a randomised controlled trial of statin choice in the event of SAMS. We anticipate that the methods demonstrated in this study can be applied to assess a genotype-guided approach to prescribing other interventions in the future.

References: *[1] Peyser B, Perry EP, Singh K, Gill RD, Mehan MR, Haga SB, et al. Effects of Delivering SLCO1B1 Pharmacogenetic Information in Randomized Trial and Observational Settings. Circ Genom Precis Med. 2018;11(9):e002228.*

ORAL CONTRIBUTED SESSION 28
Electronic Health Data

OC28-5 Evaluating the performance of common benchmarking and outlier classification methods in clinical registries – A simulation study

Jessy Hansen, Ahmadreza Pourghaderi, Susannah Ahern, Arul Earnest
Monash University, Melbourne, Australia

Background: Clinical quality registries (CQRs) are valuable repositories of prospectively collected health data that can be leveraged for observational studies and health service quality monitoring. Registries often operate within existing health record mechanisms with the additional collection of clinical data relevant to a specific patient population. A key purpose of CQRs is to drive quality improvement by benchmarking health provider (such as hospital site) outcomes to identify underperformers (or outliers). Beyond targeting sites with poor outcomes for improvement, there is also increasing interest in the public reporting of outlier detection to inform patients and stakeholders, making accurate flagging vital. Despite this, little research has robustly investigated the accuracy of benchmarking and outlier classification within clinical registries, and there is scarce guidance on methodological decision making for registries of different sizes.

Methods: We conducted a parametric simulation study varying the following registry factors: number of patients, clinicians and sites, site volume inclusion criteria and outcome prevalence. Outlier sites flagged from a combination of four site estimate methods (unadjusted and risk-adjusted using ordinary, fixed and random effects logistic regression) and three outlier classification techniques (confidence interval, control limit and false discovery rate) were compared to simulated 'true' underperformers. The receiver operator curve area under the curve (ROC AUC) calculated as the summary performance measure.

Results: The ability of benchmarking methods to correctly classify outlying sites varied between methods, with ordinary logistic regression and 95% control limits generally exhibiting superior performance across the range of registry scenarios evaluated. While the absolute number of sites and clinicians within a registry was found to have little effect on performance, higher accuracy with increasing number of patients per site was observed until 80-90% ROC AUC was reached, after which performance plateaued. The results varied by prevalence; a threshold of 100-150 outcome events on average per site was needed to reach peak classification performance. The use of a patient volume minimum to determine site analysis eligibility reduced benchmarking performance proportionally to the percentage of sites excluded with a given minimum, though a minimum requirement of 10-50 patients/site only marginally decreased accuracy.

Conclusion: CQRs should consider their size (in particular patient volume) along with outcome prevalence to determine if benchmarking and outlier flagging will be sufficiently accurate, and to inform methodological decision making. This consideration is especially impactful for registries in their initial stages and those dealing with rare diseases.

ORAL CONTRIBUTED SESSION 29
Prediction and Prognostic Models III

OC29-1 Effective sample size of individual predictions: Quantifying uncertainty in machine learning models

Doranne Thomassen[1], Toby Hackmann[1], Hans van Houwelingen[1], Ewout Steyerberg[1], Saskia le Cessie[1,2]
1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
2 Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands

Background: Recently, the effective sample size has been proposed as a measure of sampling uncertainty in individual clinical predictions, with potential implications for model development, validation and implementation[1]. This measure can be interpreted as the number of similar patients that an individual's prediction is effectively based on, given that the prediction model is correct. For predictions based on generalized linear models (GLMs), analytical expressions have been derived for the effective sample size [1]. We aimed to develop a method to compute effective sample sizes for a wider range of prediction models, including machine learning methods.

Methods: We translated the definition of the effective sample size for a prediction to a ratio of two variances: the variance of the outcome given the prediction and the variance of the prediction itself. This suggests that effective sample sizes can be computed based on estimates of these two variances, for example using bootstrap. We applied this computational approach to predictions based on linear and logistic regression and compared resulting effective sample sizes to the analytical approach. The computational approach was then extended to machine learning methods such as random forests and XGBoost. We studied resulting effective sample sizes in a real dataset of patients with myocardial infarction as well as simulated data. Results The computational approach and the analytical approach yielded similar effective sample sizes for linear and logistic regression models. For predicted risks close to 0 or 1 from a logistic regression model, differences were observed between standard errors based on nonparametric bootstrap vs approximations with the delta method; resulting in different effective sample sizes. The behaviour of effective sample sizes based on the computational approach was stable and extended well to machine learning models.

Discussion: The computational approach for effective sample size generalizes well to a wide class of prediction models. An advantage of bootstrap-based prediction variance estimation is that it can incorporate sampling uncertainty in the entire modelling process, including model selection and hyperparameter tuning. Few assumptions are required to express the effective sample size as the ratio of variances applied in our approach. Therefore, effective sample sizes can be applied to express individual uncertainty in a wide variety of prediction settings.

References: *[1] Thomassen D, le Cessie S, van Houwelingen HC, Steyerberg EW. Effective sample size: A measure of individual uncertainty in predictions. Statistics in Medicine. 2024; 1-13. doi: 10.1002/sim.10018*

ORAL CONTRIBUTED SESSION 29
Prediction and Prognostic Models III

OC29-2 Properness of common performance measures for risk prediction models: A simple illustration

Ben Van Calster[1,2]
1 KU Leuven, Leuven, Belgium
2 Leiden University Medical Center, Leiden, Netherlands

Background: A plethora of performance measures exist to evaluate clinical prediction models. A desirable characteristic of any performance measure is properness. A measure is strictly proper if only the true underlying, data generating model optimizes the expected value. A measure is semi-proper if other models may also optimize the expected value. Otherwise, measures a labeled as improper, and may be misleading. The literature on properness is rather mathematical, and focuses on 'scoring rules' for overall performance. We aim to provide illustrations of properness of common performance measures.

Methods: We simulated four correlated continuous predictors with a standard normal distribution. The true model for the binary outcome was a logistic model with coefficients corresponding with a true AUROC of 0.75 and an event proportion of 0.30. We generated 2000 datasets of 1000 patients, and applied the true model and nine incorrect models (using the same predictors) on each dataset. To avoid leakage, the definition of the incorrect models did not depend in any way on the outcome. We averaged model performance over the 2000 datasets to approach their expected values. We calculated performance measures from five categories: overall, discrimination, calibration, classification, and utility.

Results: As expected, loglikelihood, Brier score, and common R-squared measures (McFadden, Cox-Snell, Nagelkerke) are proper measures of overall performance. The discrimination slope and mean absolute prediction error were found to be improper, because incorrect models sometimes had better results than the correct model. Discrimination measures (area under the ROC curve, area under the precision-recall curve, and partial AUROC) were semi-proper. For calibration, the O:E ratio, calibration intercept, and calibration slope were semi-proper, whereas common summary measures of the calibration curve (estimated calibration index, integrated calibration index, expected calibration error) were proper. Of note, all classification measures evaluated at a fixed threshold were improper: accuracy, balanced accuracy, Youden index, diagnostic odds ratio, kappa, Matthews correlation coefficient, F1 score, sensitivity, specificity, positive predictive value, and negative predictive value. Net benefit, an increasingly popular measure for clinical utility, was proper, whereas expected cost was semi-proper.

Conclusions: This illustration cautions against the use and interpretation of some performance measures that are commonly used in the evaluation of prediction models. Specifically, many classification measures are improper and may falsely suggest one model to be better than another. Measures such as sensitivity, specificity, and positive/negative predictive value should only be used descriptively.

ORAL CONTRIBUTED SESSION 29
Prediction and Prognostic Models III

OC29-3 Optimal treatment regimes for the net benefit of a treatment

Francois Petit[1], Gérard Biau[2], Raphaël Porcher[1]
1 Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), Paris, France
2 Sorbonne University, Paris, France

Introduction: One aim of personalized medicine is to identify which treatment among several options is the most beneficial for individual patients. For that purpose, several methods have been developed to construct individualized treatment rules (ITR). An ITR provides a tailored treatment recommendation to each patient, based on their observed characteristics to improve their clinical outcome. Nonetheless, most statistical methods used to construct ITRs only consider a single outcome, while clinical evaluation is usually more complex. For instance, overall survival and progression-free survival are both commonly used as clinical endpoints in oncology. These multiple clinical endpoints should inform the clinicians and patients on how a potential treatment may improve survival, relieve symptoms, and affect the quality of life. As the finality of an ITR is to propose personalized treatment recommendations, these multiple clinical endpoints and their relations should be taken into account when constructing an ITR.

Objective: To devise statistical techniques for constructing ITRs that consider multiple clinical endpoints simultaneously.

Methods: We developed a mathematical setup relying on Rubin's causal model and inspired by Buyse's generalized pairwise comparisons to define the notion of an optimal ITR in the presence of a hierarchy of clinical endpoints, terming such an ITR pairwise optimal. We present two approaches to estimate pairwise optimal ITRs. The first is a variant of the k-nearest neighbors algorithm. The second is a meta-learner based on a randomized bagging scheme, allowing the use of any classification algorithm for constructing an individualized treatment rule. We study the behavior of these estimation schemes from a theoretical standpoint and through Monte Carlo simulations and illustrate their use on trials data.

Results: We proved that the two estimation schemes we proposed are universally consistent and that pairwise optimal rules are optimal in the usual sense in the case of a single binary endpoint.

Conclusion: We have developed new methods that allow to consider simultaneously several endpoints and use censored data when designing ITR. Our simple method extends to personalized medicine Buyse's generalized pairwise comparison method.

ORAL CONTRIBUTED SESSION 29
Prediction and Prognostic Models III

OC29-4 Optimising dynamic treatment regimens using sequential multiple assignment randomised trials data with missing data

Jessica Xu[1], Robert Mahar[1,2], Anurika P De Silva[1], Katherine Lee[2,3], Julie A Simpson[1,4]
1 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia
2 Clinical Epidemiology and Biostatistics Unit, Murdoch Children Research Institute, Royal Children's Hosptal, Melbourne, Australia
3 Department of Paediatrics, University of Melbourne, Melbourne, Australia
4 Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Background: Dynamic treatment regimens are a set of sequential rules that guide clinicians to make treatment decisions over time for patients with chronic or progressive medical conditions. Sequential multiple assignment randomised trials (SMARTs) are studies that can be used to optimise dynamic treatment regimens by repeatedly randomising participants to treatments. Q-learning, a stage-wise regression-based method commonly used to analyse SMARTs, uses backward induction to compare treatments administered as a sequence. Missing data is a common problem in randomised trials, and in SMARTs with sequential randomisation, the missing data patterns can be complicated. Common methods for handling missing data such as complete case analysis (CCA) and multiple imputation (MI) have been widely explored in single-stage randomised trials, however, the only study that explored these methods in SMARTs did not consider Q-learning. We aim to evaluate the performance of CCA and MI on the estimation of Q-learning parameters.

Methods: We simulated 1000 datasets of 500 participants, based on a simulation scheme in SMARTs of two stages first used by Chakraborty et al. [1], under a range of missing data scenarios defined by different missing directed acyclic graphs (mDAGS), percentages of missing data (20%, 40%), stage 2 treatment effects, and strengths of association with missingness in stage 2 treatment, patient history and outcome. We compared bias and precision for CCA and MI using a simulated two-stage SMART and used retrospective data from a longitudinal smoking cessation SMART for empirical illustrations of the results.

Results: When there was no treatment effect at either stage 1 or 2, we observed close to zero absolute bias and similar empirical standard errors for MI and CCA under all missing data scenarios. When all participants had a relatively large stage 2 treatment effect, we observed some minimal bias, with slightly greater bias for MI. Resulting empirical standard errors were higher for MI compared to CCA under all scenarios except for when data were missing not dependent on any variables. When the stage 2 treatment effect varied between participants, we observed greater bias for MI, which increased with the percentage missingness, while the bias for CCA remained minimal. Resulting empirical standard errors were lower/similar for MI compared to CCA under all missing data scenarios.

Conclusion: Results showed that for a two-stage SMART, MI failed to capture the differences between treatment effects when the stage 2 treatment effect varied between participants.

Reference: *[1] Chakraborty B, et al. Stat Methods Med Res. 2010;19(3):317-43*

ORAL CONTRIBUTED SESSION 29
Prediction and Prognostic Models III

OC29-5 Extended sample size calculations for external validation of a machine learning model for individual risk prediction of a binary outcome

Rebecca Whittle[1,2], Joie Ensor[1,2], Lucinda Archer[12,], Paula Dhiman[3], Gary Collins[3], Richard Riley[1,2], Kym Snell[1,2]
1 Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom
2 National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, United Kingdom
3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Muscuskeletal Sciences, University of Oxford, Oxford, United Kingdom

Background: When evaluating the predictive performance of a model for individualised risk prediction, the sample size needs to be large enough to precisely estimate the performance measures of interest. Current sample size guidance is based on precisely estimating calibration, discrimination and net benefit [1]. However, in recent years there has been an exponential increase in the use of machine learning methods to develop models and these are often evaluated using other metrics, so here we extend our guidance to these metrics.

Methods: We propose to develop closed-form solutions to estimate the minimum sample size required to target sufficiently precise estimates of the accuracy, specificity, recall, precision and F1-score in an external validation study of a prediction model with a binary outcome. This approach requires the user to pre-specify the target standard error and the expected value for each performance measure. A large synthetic dataset can then be simulated approximating the assumed distribution of the model's predicted risk in the validation population, so that the 'true' values of each of the measures at a given probability threshold can be calculated. From this, a minimum required sample size can be calculated for a desired confidence interval width.

Results: We describe how the sample size formulae were derived and demonstrate their use in a worked example planning to externally validate a published model for predicting in-hospital clinical deterioration, targeting a confidence interval width of 0.1 for each of the measures (i.e., standard error=0.0255), at a probability threshold of 0.1 for clinical use. Using the formulae, the minimum required sample size was calculated as 388, 339, 43, 422 and 38 individuals for accuracy, specificity, recall, precision and the F1-score respectively. Whereas using previous criteria [1], it was identified that a larger sample size of at least 949 would be needed to estimate the calibration slope precisely. Extension to time-to-event outcomes is also considered.

Conclusion: Our formulae enable researchers to calculate the minimum sample size for an external validation study, aiming to target a particular confidence interval width, for performance measures commonly used to evaluate machine learning models for risk prediction. In our example, the minimum sample size required was lower than what would be required to precisely estimate the calibration slope, and we expect this would most often be the case.

Reference: *[1] Riley et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. BMJ 2024;384:e074821*

ORAL CONTRIBUTED SESSION 30
Survival Analysis III

OC30-1 Regression models for recurrent events in presence of terminal events with efficient estimation

Giuliana Cortese[1], **Thomas Scheike**[2]
1 University of Padova, Padova, Italy
2 University of Copenhagen, Copenhagen, Denmark

Background: Recurrent events, i.e., the repeated occurrence of the same event over time, are often encountered in the biomedical setting and provide rich information to study disease progression over time. Often observation of further recurrent events can be precluded by the presence of a terminal event, which thus it is important to be considered in estimating key summary measures of interest. Given the complex dependence structure of these data, efficient methods for their analysis are in need.

Methods: One parameter that is frequently of main interest in this context is the marginal mean of the cumulative number of recurrent events up to a certain time t. We here consider the popular GhoshLin regression model for the marginal mean number of recurrent events, in presence of a terminal event. We propose a novel approach to estimate efficiently the parameters of such model. The fully efficient estimator is not practically available and we therefore suggest a dynamic linear censoring augmented estimator for the regression coefficients.

Results: The proposed estimator is simple to compute and does not rely on any additional assumptions. We conducted simulation studies and showed that the suggested estimation approach has better properties and improves over the standard GhoshLin estimator. In particular, the improved efficiency leads to more precise estimates with narrower pointwise confidence intervals. A motivating application on patients with chronic intestinal failure is also shown, where we model how the mean of repeated bloodstream infections depends on several covariates.

Conclusion: We suggest to use a novel efficient estimation method for making inference under the GhoshLin regression model. It provides a considerable improvement over the standard GhoshLin estimator in terms of precision. The augmentation approach therein is based on dynamic linear models for the censoring and has the advantage that no explicit assumptions about the dependence among recurrent events and the terminal events are needed.

ORAL CONTRIBUTED SESSION 30
Survival Analysis III

OC30-2 Prediction of incident cardiovascular events using omics-derived latent factors

Andrea Mario Vergani[1,2], Soumick Chatterjee[1], Andrea Lampis[1,2], Michela Carlotta Massi[1], Sara Ometto[1], Carlo Andrea Pivato[1,3], Alessia Visconti[1,4], Craig A. Glastonbury[1], Marco Masseroli[2], Francesca Ieva[1,2], Emanuele Di Angelantonio[1,5]

1 Human Technopole, Milan, Italy
2 Politecnico di Milano, Milan, Italy
3 IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
4 Università di Torino, Turin, Italy
5 University of Cambridge, Cambridge, United Kingdom

Background/Introduction: Risk scores for cardiovascular disease (CVD) traditionally rely on a limited set of vascular risk factors (VRF, e.g., sex, diabetes). Omics modalities, such as genomics and radiomics, provide additional information predictive for CVD; however, their integration in the forecast of incident cardiovascular events was traditionally restricted to conventional quantities (i.e., Polygenic Risk Scores – PRS – for genomics, cardiac measures for heart magnetic resonance imaging – MRI). This study leverages deep representation learning and multi-omics fusion to develop more accurate risk scores for incident CVD, integrating and evaluating the impacts of genetic information and cardiac MRI.

Methods: We conducted a survival analysis study on approximately 25,000 UK Biobank subjects healthy at MRI; the incidence of CVD was about 4.5% through a median follow-up of 3.5 years (until October 2022). The analysis was repeated considering specific subtypes of CVD as endpoints, i.e., atrial arrhythmia, coronary artery disease, structural heart disease. We extracted latent vectors from Single Nucleotide Polymorphism (SNP) data using a Latent Diffusion Model, and deep representations of long-axis heart magnetic resonance acquisitions with a Diffusion Autoencoder; these latent factors were integrated with traditional VRF into our survival model via Multi-Omics Factor Analysis. We assessed latent model performance through reconstruction error and survival performance through Concordance index.

Results: Our latent feature extraction models achieved state-of-the-art reconstruction quality, with an average error rate of 80/2048 SNPs and a structural similarity index measure of 0.95 for cardiac MRI. The survival models with omics-derived latent variables achieved better performance over traditional risk models based on VRF, PRS and cardiac measures: test Concordance index increased up to 10% compared to classical risk scores entirely built with VRF, and up to 5% in the case of integrated data modalities. While the impact of genomics in the forecast of incident CVD was stable across disease subtypes, the inclusion of cardiac MRI proved particularly predictive for atrial arrhythmia and structural heart disease.

Conclusion: Our approach improves the prediction of incident CVD by integrating deep learning-based representations of omics information. Specifically, this study evaluated the predictivity of genomics- and cardiac MRI-derived latent factors for cardiovascular endpoints and interpreted them in terms of CVD risk. To the best of our knowledge, this is the first work about the integration of a comprehensive latent representation of genomics and cardiac MRI into prediction models for incident CVD.

ORAL CONTRIBUTED SESSION 30
Survival Analysis III

OC30-3 A novel approach to multiple survival curves comparison based on machine learning: Utilising random forests with reducing assumption dependency

Lubomir Stepanek[1,2], Filip Habarta[2], Ivana Malá[2], Luboš Marek[2]
1 Department of Biomedical Statistics, Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, Prague, Czech Republic
2 Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

Background/Introduction: Survival analysis routinely involves comparing survival curves, which is traditionally performed through well-established methods such as the log-rank test and Cox proportional hazard models. However, these methods are constrained by strict statistical assumptions, and their outcomes may be biased if these assumptions are violated.

Methods: In this work, we address the limitations of the classical techniques in survival analysis due to their assumptions and propose a new method dedicated to multiple survival curve comparisons based on a random forest algorithm, which is almost assumption-free. Compared to random survival curves, which require a covariate structure and a set of predictors on input, the introduced method uses only time-to-event variables. The time to event is used for generating variables based on usual estimators, e.g., the Kaplan-Meier estimator, which remain constant within a group but are individualized by missing values imputations for time points after the individual experienced the event. The principle of the method uses the fact that a random forest model contains multiple decision trees, and each tree, according to its complexity and number of leaf nodes, can classify into either one, two or more groups, defined by their survival curves.

A number of trees classifying into two or more groups, i.e., contradicting null hypothesis about no difference between the groups' survival curves, is related to p-value, here denoted as ϕ-value. We derive the number of trees contradicting the null hypothesis following the Poisson-binomial distribution and show that the ϕ-value is a consistent and efficient estimator. Also, we use the Le Cam theorem and Chernoff bound to estimate an upper bound for the probability that the ϕ-value is lower than $\alpha \in R$, which helps to insight into the method's statistical power. Tree complexity can be reduced by pruning, which helps to minimize the first-type error rate but may decrease the statistical power of the technique, too.

Results: We simulated multiple pairs and triplets of different and non-different survival curves and compared them using the proposed method and log-rank test or Cox model. This illustrated the method's reduced risk of first-type error rate for higher tree pruning levels and lowered statistical power.

Conclusion: We introduced a method for comparing multiple survival curves based on random forests as an alternative to the traditional toolbox in survival analysis. The method is almost assumption-free, has minimal and adjustable risk of first-type error rate, but may suffer from lower statistical power.

ORAL CONTRIBUTED SESSION 30
Survival Analysis III

OC30-4 Flexible rate models for recurrent event data

Elsa Coz[1,2,3], Hadrien Charvat[4,5], Delphine Maucort-Boulch[1,2,3], Mathieu Fauvernier[1,3,2]
1 Université Lyon 1, Villeurbanne, France
2 Hospices Civils de Lyon, Pôle Santé Publique, Service de Biostatistique et Bioinformatique, Lyon, France
3 CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé, Villeurbanne, France
4 Faculty of International Liberal Arts, Juntendo University, Tokyo, Japan
5 Division of International Health Policy Research, Institute for Cancer Control, National Cancer Center, Tokyo, Japan

Background/Introduction: In case of recurrent events, the individual counting process may be completely described by the intensity, i.e., the instantaneous probability of event conditionally on the history of the process and covariates. The conditioning makes the intensity potentially very complex to model. Alternatively, marginal quantities like the rate, defined by Lin et al. (2000) as the derivative of the cumulative mean, may be considered. Unlike intensity, modelling the rate does not require any assumption regarding the covariance structure between events. Moreover, it is appealing from an epidemiologic point of view to describe the association of the process with covariates synthetically at the population level. While flexible intensity models have been proposed, only semi-parametric approaches are available for the rate. The aim of this work is to propose a flexible modelling framework for the rate in both the unpenalized and penalized settings.

Methods: Lin's semi-parametric log-rate model is fitted using an estimating equation derived from the partial likelihood of a Poisson process. We extended the approach to a fully parametric log-rate model using natural cubic splines, with inference based on the full-likelihood of a Poisson process estimating equation. For comparison purposes, we also considered a spline-based Poisson process intensity model with Gaussian subject-specific frailty. In line with development for flexible log-hazard models, we further considered models including a penalty term to control the smoothness of the estimated rate. We evaluated and compared the models in an extensive simulation study, with different ways to generate the recurrent process, and illustrated the approach using real data.

Results: In the unpenalized setting, the log-rate model showed good inference properties in all simulated scenarios. The frailty-intensity model performed almost as well, but tended to show either over- or under-estimation of the variance of the parameters in the most complex scenarios. The penalized setting allowed a substantial reduction of the variance of the estimators of the log-rate model at the cost of small under-coverage.

Conclusion: The proposed flexible log-rate models offer a description of the rate over time, handling potential non-proportional and non-linear effects of the covariates. The penalized setting greatly simplifies model choice while limiting overfitting.

Reference: *[1] Lin DY, Wei LJ, Yang I, Ying Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc Ser B Stat Methodol 62:711–730. https://doi.org/10.1111/1467-9868.00259*

ORAL CONTRIBUTED SESSION 30
Survival Analysis III

OC30-5

StCA winner

Bayesian analysis of restricted mean survival time adjusted on covariates using pseudo-observations

Léa Orsini[1,2], Emmanuel Lesaffre[2], Guosheng Yin[3], Caroline Brard[4], David Dejardin[5], Gwénaël Le Teuff[1]
1 CESP, INSERM U1018, Université Paris-Saclay, UVSQ, Villejuif, France
2 I-Biostat, KU-Leuven, Leuven, Belgium
3 Department of Mathematics, Imperial College London, London, United Kingdom
4 Ipsen Innovation, Clinical Development Organisation, Les Ulis, France
5 Product Development, Data Sciences, F. Hoffmann-La Roche AG, Basel, Switzerland

Background/Introduction: The difference in restricted mean survival time (dRMST) at a specific time is an appropriate measure of treatment effect in a 2-arm randomized clinical trial (RCT) when the proportional-hazards (PH) assumption does not hold. This is common with immuno-oncology therapies. One straightforward approach to estimating RMST is to integrate the Kaplan-Meier curve numerically. However, this approach does not allow for covariate adjustment. Several frequentist methods estimate the RMST with covariate adjustment by modeling and integrating the survival function. A more natural approach is to consider a regression model on the RMST directly, using pseudo-observations that allow for a direct fit without modeling the survival function. Only two Bayesian methods exist, and both require non-trivial modeling of the survival function with a nonparametric prior process.

Methods: We developed a new method to analyze pseudo-observations in the Bayesian framework. This method is based on the Bayesian generalized method of moments (GMM) [1], analog to the frequentist Generalized Estimating Equations often used to analyze pseudo-observations [2]. We focus here on the Bayesian analysis of pseudo-observation to estimate RMST. This novel approach combines the simplicity of pseudo-observations for survival analysis with the benefits of the Bayesian approaches. The performance of this approach is compared to existing frequentist and Bayesian RMST estimators through a simulation study of 2-arm RCTs mimicking different patterns of time-dependent treatment effect (early and delayed effect) with prognostic variables (binary, continuous). The PSA progression-free survival (PFS) of the Getug-AFU 15 trial, a phase 3 RCT in prostate cancer (n=384), is chosen as an illustrative example.

Results: Simulations showed that our approach produces similar performance for unadjusted and adjusted RMST estimations compared to existing methods in the different scenarios with better precision after covariates adjustment. For the Getug-AFU 15 trial, the 5-dRMST estimation for the PSA PFS with a diminished treatment effect is similar across the methods.

Conclusion: We propose a novel Bayesian approach using pseudo-observations for analyzing RMST at a specific time adjusted on covariates. This method does not require specifying the survival function, making it attractive compared to Bayesian methods, and can be extended to the joint analysis of RMST at multiple times.

References: *[1] Yin, G. (2009). Bayesian generalized method of moments. Bayesian Analysis, 4(2), 191–208. [2] Andersen, P. K., Hansen, M. G., & Klein, J. P. (2004). Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations. Lifetime Data Analysis, 10(4), 335–350.*

ORAL CONTRIBUTED SESSION 31
Clinical Trials V

OC31-1 Group-sequential methods for generalised pairwise comparisons

Stephen Schüürhuis, Frank Konietschke
Charité - Universitätsmedizin Berlin, Germany

Common nonparametric statistical tests, such as the Brunner-Munzel test, are inherently based on pairwise comparisons between observations of a treatment and control group i = T,C. Buyse (2010) extended the concept of pairwise comparisons to include multiple prioritized outcomes. The resulting statistics are essentially centered around the Net Treatment Benefit , which aggregates the contribution of multiple outcomes into a single effect measure and interprets as the net probability of a favorable outcome in the treatment group. Since these tests can take multiple outcomes into account, they provide more patient-centric methods for clinical trials. To further ensure timely benefits from promising new drugs and early identification of futile treatments, careful planning of clinical trials is crucial. Group-sequential designs are well-known methods to enhance the efficacy of clinical trials by incorporating interim analyses into ongoing trials, enabling early termination for efficacy or futility while controlling for multiplicity. These methods have long been established for continuous, binary, and survival outcomes. Recently, Nowak et al. (2022) have expanded group-sequential designs to the Brunner-Munzel test by demonstrating that the sequential test statistics asymptotically follow a multivariate normal distribution. However, current group-sequential theory remains centered on effect measures and tests including only a single primary endpoint. Building upon the work by Nowak et al. (2022), we expand their research within the framework of generalized pairwise comparisons introduced by Buyse (2010). This enables interim testing in ongoing trials based on multiple prioritized outcomes on varying scales, rather than just a single outcome. In this presentation, we establish that test statistics derived from generalized pairwise comparisons asymptotically conform to the canonical multivariate normal distribution. Through an extensive simulation study, we demonstrate that the proposed methods accurately control the type-I error rate, even in small sample sizes. By incorporating multiple outcomes, we illustrate the potential to enhance the power of group-sequential trials. In summary, our study highlights the promising integration of generalized pairwise comparisons into group-sequential methods, offering a comprehensive approach for clinical trial design and analysis. By leveraging data from multiple outcomes, our findings emphasize the utility of these methods in enhancing both statistical power and clinical relevance. Keywords — Generalized pairwise comparisons, group-sequential designs.

References: *[1] Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Statistics in Medicine, 29(30), 3245-3257. [2] Nowak, C. P., Mütze, T., & Konietschke, F. (2022). Group sequential methods for the Mann-Whitney parameter. Statistical Methods in Medical Research, 31(10), 2004-2020.*

ORAL CONTRIBUTED SESSION 31
Clinical Trials V

OC31-2 Response rate estimation in single-stage basket trials: A comparison of estimators that allow for borrowing across cohorts

Antonios Daletzakis[1,2], Rutger van den Bor[3], Vincent van der Noort[1], Kit CB Roes[2]
1 Biometrics Department, Netherlands Cancer Institute, Amsterdam, Netherlands
2 Department of Health Evidence, Section Biostatistics Radboud University Medical Center, Nijmegen, Netherlands
3 Dept. of Data Science and Biostatistics, Julius Center for Health Sciences and Primary Care University Medical Center Utrecht, Utrecht, Netherlands

Therapeutic advancements in oncology have transitioned towards targeted therapy based on specific genomic aberrations. This shift necessitates innovative statistical approaches in clinical trials, notably in the emerging paradigm of master protocol studies. Basket trials, a type of master protocol, evaluate a single treatment across cohorts sharing a common genomic aberration but differing in tumor histology. While offering operational advantages, the analysis of basket trials introduces challenges with respect to statistical inference. Basket trials can be used to decide for which tumor histology the target treatment is promising enough to move to confirmatory clinical evaluation and can employ a Bayesian design to support this decision making. In addition to decision making, estimation of the cohort-specific response rates is highly relevant to inform design of subsequent trials. This study evaluates seven Bayesian estimation methods for basket trials with a binary outcome contrasted with the (frequentist) sample proportion estimate, through a simulation study. The objective is to estimate cohort-specific response rates, with a focus on average bias, average mean squared error, and the degree of information borrowing. A variety of scenarios are explored, covering homogeneous, heterogeneous and clustered response rates across cohorts. The evaluated methods present trade-offs in bias and MSE, emphasizing the importance of method selection based on trial characteristics. Berry's method excels in scenarios with limited heterogeneity. No clear winner emerges in a more general scenario, with method performance influenced by the amount of shrinkage towards the overall mean, bias and the choice of priors and tuning parameters in more complex settings. Challenges include the computational complexity of methods, the need for careful tuning of parameters and prior distribution specification, and the absence of clear guidance on their selection. Researchers should consider these factors in designing and analyzing basket trials.

ORAL CONTRIBUTED SESSION 31
Clinical Trials V

OC31-3 Patient-centred dose-finding trials using safety, efficacy and patient-reported outcomes

Emily Alger[1], Sumithra Mandrekar[2], Jun Yin[3], Christina Yap[1]
1 Clinical Trial and Statistics Unit, Institute of Cancer Research, London, United Kingdom
2 Department of Quantitative Health Sciences, Mayo Clinic, Rochester, United States
3 Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, United States

Background: Most dose-finding oncology trials (DFOTs) aim to identify a maximum tolerated dose (MTD) which adopts the traditional dose-efficacy paradigm that efficacy increases with treatment dose. The MTD is often determined by observing dose-limiting toxicities (DLTs). Whilst established cytotoxic agents generally conform to this paradigm, new investigational therapies may no longer exhibit such a relationship. In such instances, emphasis should be placed on the methodological advancements required for trial designs to identify optimal biological doses rather than maximum tolerated doses. Incorporating Patient-reported Outcomes (PROs) within DFOTs is increasingly advocated to enrich our understanding of a treatment's tolerability profile over extended tolerability assessment windows. The introduction of PROs within DFOTs will cater for the extended tolerability assessment required for prolonged therapies, potentially until treatment resistance or disease progression is observed.

Methods and Results: The current trial designs which incorporate PROs, including PRO-CRM and PRO-ISO, use PROs to define a Patient-assessed DLT. Such designs reduce this rich high-dimensional source of PRO data to one binary endpoint. Following the identification of some target P-DLT rate, these trial designs escalate dose until a MTD is identified. However as new immunotherapy and targeted agents emerge from drug discovery, it is important to assess patient's tolerability to treatment over extended tolerability assessment windows and to evaluate the trade-off between a treatment's tolerability and additional endpoints such as efficacy.

The proposed new DFOTs integrates three distinct endpoints:

(1) Clinician-assessed DLTs – to determine a dose is clinically safe during the first cycle of treatment,

(2) Patient-reported tolerability – to determine a dose is tolerable to a patient over the course of treatment administration and beyond,

(3) Activity – to determine a dose is efficacious.

Extensive simulations are conducted to assess its performance in comparison to existing trial designs across various clinically relevant scenarios. We demonstrate how the proposed design appropriately assigns doses throughout the study by considering the three distinct endpoints, offering investigators valuable insights into how it works.

Conclusions: Within this work, we explore the directions of future PRO implementation. This includes the investigation of novel PRO endpoints and development of innovative trial designs that optimise treatment dose by integrating PRO data with efficacy endpoints. As the field evolves, patient-centric dose-finding approaches incorporating PROs will become crucial in advancing our understanding of treatment tolerability, thereby shaping the future landscape of DFOTs.

ORAL CONTRIBUTED SESSION 31
Clinical Trials V

OC31-4 Using multi-state models to design single-arm adaptive trials with progression-free and overall survival endpoints

Zhulin Yin[1], Haiyan Zheng[2], Christina Yap[1]
1 Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, United Kingdom
2 Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

Background: The landscape of oncology clinical trials is changing with the fast development of new immunotherapies. Unlike chemotherapy, immunotherapy often takes longer to show benefits, and the survival or clinical benefits are not necessarily reflected in response rates. These findings have brought up the need to develop efficient phase II trial designs incorporating survival endpoints.

Methods: We propose a new class of adaptive two-stage designs for phase IIa oncology trials that exploit the use of information on both the intermediate and long-term endpoints such as progression-free survival (PFS) and overall survival (OS). Specifically, we relate the PFS and OS via a three-state multi-state survival model, which depicts the causal trajectory of the disease progression. We have further derived the respective survival functions for PFS and OS and construct the log-rank test statistics using stagewise survival data. Futility stopping can be planned at the interim. Upon completion of the trial, an inverse normal combination test is used to maintain a strong control of the overall type I error.

Results: We compare our design with four other designs, each considering distinct endpoints, through extensive simulation studies. Simulation results suggest the proposed design outperforms the other in most cases, achieving sufficiently high power in identifying an effective treatment. In scenarios where the treatment has demonstrated relatively low PFS improvement but dramatic improvements in OS, however, the proposed design tends to stop early for futility.

Conclusion: Single-arm phase II designs often encounter challenges in assessing long-term outcomes which may take extended periods to emerge, leading to prolonged trial duration and increased costs. Being both efficient and intuitive, our design can be utilised broadly in phase II two-stage, single-arm trials, wherein both short-term and long-term survival outcomes are available for efficacy evaluation.

ORAL CONTRIBUTED SESSION 31
Clinical Trials V

OC31-5 A pragmatic approach to confidence interval estimation for a complier average causal effect for cluster randomised trials

Tansy Edwards, Jennifer Thompson, Charles Opondo, Elizabeth Allen
London School of Hygiene & Tropical Medicine, United Kingdom

Background: Compliance (also referred to as adherence) in individuals in cluster randomised trials (CRTs) is often less than 100%. Estimating a treatment, or intervention, effect according to assigned treatment (intention-to-treat) ignores non-compliance in individuals and could underestimate efficacy (how well the treatment, or intervention, works in individuals who receive it). Estimating the effect of the intervention among compliers (the complier average causal effect, CACE) provides an unbiased estimate of efficacy but can be complex in CRTs.

Methods: We evaluated the bias of a simple approach to estimating the CACE, and the coverage and power of a bootstrapping approach that accounts for clustering in the data, to obtain a 95% confidence interval for a CACE for CRTs. We varied prevalence of the outcome in the control arm and prevalence of non-compliance (5%, 10%, 20%, 30%, 40% for each). CRTs were simulated with the minimum number of clusters to provide at least 80% and 90% power, to detect an ITT odds ratio of 0.5 with 100 individuals per cluster.

Results: Under all non-compliance scenarios (5-40%), there was negligible bias in CACE odds ratios. In the worst case of bias, a true CACE OR of 0.18 was estimated as 0.15 for the rarest outcome (5%) and highest non-compliance (40%).

There was no under-coverage of bootstrap confidence intervals. Confidence intervals were the correct width for an outcome prevalence of 20-40% but too wide (over-coverage) for a less common outcome.

Power loss for a CACE bootstrap analysis, compared to ITT regression analysis increased as the prevalence of the outcome in the control arm decreased across all non-compliance scenarios, particularly for an outcome prevalence of less than 20%. Power was similar with 40% prevalence of the outcome in the control arm.

Conclusions: Clustered bootstraps are an accessible and intuitive method to estimate confidence intervals for the CACE to support effectiveness analyses in CRTs.

ORAL CONTRIBUTED SESSION 32
Statistics in Epidemiology II

OC32-1 Win statistics in observational cancer research: Integrating clinical and quality of life outcomes

Maria Vittoria Chiaruttini[1], Giulia Lorenzoni[1], Gaya Spolverato[2], Dario Gregori[1]
1 Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences, University of Padova, Padova, Italy
2 Department of Surgical Oncological and Gastrointestinal Science, University of Padova, Padova, Italy

Introduction: Quality of life (QoL) considerations among oncological patients are increasingly recognized as pivotal metrics alongside traditional endpoints such as mortality and disease progression. In this regard, the Win Ratio (WR), Win Odds (WO), and Net Benefit (NB) emerge as statistical tools capable of integrating both efficacy and QoL through composite endpoints. They enable a nuanced assessment of treatment benefits and drawbacks, facilitating the prioritization of outcomes based on clinical significance rather than temporal occurrence. In randomized trials, Win Statistics offer an equitable means of comparing treatment and control cohorts through pairwise comparisons. However, their application in observational settings introduces complexities related to confounding variables. Herein, the Propensity Score (PS) emerges as a valuable strategy to mitigate confounding by matching treated and control subjects with similar baseline risks. Nonetheless, PS matching often leads to sample size reduction, particularly evident with increased between-group imbalances.

Methods: Our study investigates the practical implications of employing robust inference techniques for matched win statistics, as proposed by Matsouaka et al. (2022) rather than the Pocock approach (2012). Additionally, we compare matched win statistics with stratified win statistics computed within strata of similar treatment likelihood, then pooled via the Dong method (2018). Through a real-world application in the colorectal cancer domain and simulation analysis, we elucidate the long-term implications of these statistical approaches on decision-making processes.

Results: Our findings underscore the superior performance of matched win statistics, particularly in terms of coverage probability. Moreover, the Net Benefit statistic emerges as the most reliable measure when employing PS stratification.

Conclusion: In conclusion, our research provides valuable insights into the practical considerations surrounding the application of win statistics in observational settings, aiding researchers in navigating and optimizing their utilization for enhanced oncological patient outcomes.

References: *[1] Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European Heart Journal. 2012;33(2):176-182. doi:10.1093/eurheartj/ehr352. [2] Matsouaka RA. Robust statistical inference for matched win statistics. Stat Methods Med Res. 2022;31(8):1423-1438. doi:10.1177/09622802221090761. [3] Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. Journal of Biopharmaceutical Statistics. 2018;28(4):778-796. doi:10.1080/10543406.2017.1397007*

ORAL CONTRIBUTED SESSION 32
Statistics in Epidemiology II

OC32-2 Joint disease mapping for bivariate count data with residual correlation due to common cases

Edouard Chatignoux[1], Zoé Uhry[1,2], Laurent Remontet[2,3], Isabelle Albert[4]
1 Sante Publique France, Saint Maurice, France
2 Hospices Civils de Lyon, Pôle Santé Publique, Service de Biostatistique-Bioinformatique, Lyon, France
3 Laboratoire de Biométrie et de Biologie Évolutive, Équipe Biostatistique-Santé, CNRS UMR 5558, Villeurbanne, France
4 UMR MIA Paris-Saclay, Université Paris-Saclay, AgroParisTech, INRAE, Palaiseau, France

Background/Introduction: Studying the joint spatial distribution of two count outcomes (e.g. two diseases) have received much attention in last decades. Bivariate count data are usually analyzed with Poisson shared component model (P-SCM) in which geographically structured latent variables are used to model spatial variations that are specific and shared by both outcomes; the dependence between the outcomes is assumed to be fully accounted for by the latent variables.

However, in some contexts (e.g. when it is frequent to have both diseases or when analyzing two proxies of the same disease), common (unobserved) cases are counted in both sources, resulting in an additional positive residual correlation of the counts. In this situation, P-SCM wrongly attributes this residual correlation to the covariance of the latent variables, leading to biased inference and degraded predictive performance.

In this presentation, we propose a new model, the Bivariate Poisson SCM (BP-SCM), to study such correlated bivariate data.

Methods: BP-SCM relies on decomposing each disease observed count (Y1 and Y2) into common cases (X3) and distinct cases (X1 and X2); a model is then specified on each of these three (unobserved) components accounting for their shared and specific geographical variations. We used simulation to evaluate the performances of P-SCM and BP-SCM and we then applied the methodology to jointly model the spatial variations of two proxies of cancer incidence in France.

Results: Simulations and application show the good performance of BP-SCM and confirm that P-SCM overestimates the shared variability, especially when residual correlation is important. They also illustrates the ability of BP-SCM to retrieve incidence levels and spatial variations of the unobserved common and distinct cases and to estimate their shared variation.

Conclusion: Residuals correlation due to common cases should not be overlooked in joint models and BP-SCM ensures correct inference in this case. Beyond this challenge, and interestingly, the BP-SCM makes it possible to partition the bivariate counts into common and distinct cases, using only the margins of the two diseases, thanks to the information contained in the residual correlation. This partitioning provides valuable epidemiological information, allowing, for example, to describe the spatial variation of the co-occurrence of two diseases without the need to link data sources. This model also offers very interesting prospects for predicting cancer incidence in areas not covered by a registry (through joint modelling of a medico-administrative proxy and of registry data).

ORAL CONTRIBUTED SESSION 32
Statistics in Epidemiology II

OC32-3

On estimation of relative risk regression model for binary data

CASc Winner

Mahnaz Ibrahim, M. Shafiqur Rahman
Institute of Statistical Research & Training, University of Dhaka, Dhaka, Bangladesh

Background: Relative risks are often preferred over odds ratios for quantifying the association between a predictor and a binary outcome due to their more intuitive interpretation. In relative risk regression, the log-link binomial generalized linear model imposes constraints on the parameter space to prevent the fitted probabilities from exceeding 1. These constraints lead to numerical problems with standard maximum likelihood estimation and have prompted the development of various estimation approaches. However, no viable approach exists that addresses all numerical problems. The current study proposes a penalized likelihood method for estimating the relative risk regression model and compares its performance with the existing estimation methods for binary data.

Methods: This study has applied Firth's penalized likelihood approach to the relative risk regression model and derived a modified score equation to overcome the boundary problem and improve the robustness of the estimation of the standard errors of the estimates, alongside a Simulated Annealing estimation procedure to overcome the repelling fixed-point problem. Two series of extensive simulations were conducted to assess the performance of the proposed penalized estimator and compared the results with the existing methods (COPY, Poisson, Combinatorial Expectation Maximization, etc.) based on the following evaluation criteria: convergence rate, relative bias, mean squared error, coverage probability, and the ratio of standard error and simulated standard error. Simulation Series I was based on arbitrary true models, while Series II was based on a true model from Stress Echocardiography data. Results The simulation results revealed that the existing methods have failed to achieve convergence of the likelihood concerning infinitely large values of the recession coefficient and standard errors. The convergence problem often exists even for large sample sizes. In contrast, the proposed penalized method has been shown to achieve convergence in all aspects of simulation scenarios in both series. Although all the estimates produced by the existing methods perform gradually better as the sample size increases, exhibiting the usual asymptotic behavior, only the proposed estimation method achieved optimal performance for small sample sizes. Application to real data shows a similar pattern to the simulation findings.

Conclusion: The most impressive feature of the proposed estimation procedure is that no additional assumptions are required to provide reliable estimates. Therefore, the proposed estimation procedure can be an optimal choice for estimating the relative risk regression model across a wide spectrum of scenarios, encompassing small to large sample sizes and scenarios with both rare and common outcome events while accommodating varying numbers of covariates.

ORAL CONTRIBUTED SESSION 32
Statistics in Epidemiology II

OC32-4 Accounting for the recruitment process into Bayesian modelling of vaccine data

Mauro Gasparini, Marco Ratta
Politecnico Di Torino, Torino, Italy

Introduction: The recruitment process may have an impact on the censoring mechanism affecting the data in vaccine clinical trials. In this work we try to model explicitly such impact in order to improve estimation of Vaccine Efficacy (VE), a measure of effect vastly employed in vaccine clinical research.

Methods: We introduce a comprehensive Bayesian approach that improves upon existing methodologies, explicitly considering patient recruitment processes to inform parameter estimation. In particular - in contrast to most methods currently used – in our proposed methodology the total number of cases as well as the exposure times are seen as recruitment-dependent statistics, with underlying distributions which are used to derive a full likelihood for the model.

Results: The methodology is validated through extensive numerical simulations, demonstrating substantial improvements in parameter estimation across diverse scenarios and under multiple recruitment plans. An application of our novel approach to the now famous Pfizer/BioNTech anti-Covid-19 vaccine data reveals increased precision in VE estimation even when the details of the recruitment process are not explicitly known, but rather only estimated.

Conclusion: In this work we prove that modeling explicitely the recruitment process in vaccine trials is beneficial in improving the precision of the Vaccine Efficacy estimation.

While our results focus on recruitment with a linear density, our framework is fully flexible and may be easily extended to any accrual distribution, proving to be conceptually robust and practically advantageous, and offering a promising alternative despite the need for Markov Chain Monte Carlo (MCMC) simulations, which can be conducted very efficiently.

References: *[1] M. Ewell. Comparing methods for calculating confidence intervals for vaccine efficacy. Statistics in Medicine, 15:2379–2392, 1996. [2] F.P. Polack et al. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. The New England Journal of Medicine, 383:2603–2615, 2020. [3] S.J. Thomas et al. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine through 6 months. The New England Journal of Medicine, 385:1761–1773, 2021.*

ORAL CONTRIBUTED SESSION 32
Statistics in Epidemiology II

OC32-5 Stochastic search variable selection in network meta-analysis: A novel approach for detecting inconsistencies

Georgios Seitidis[1,2], Stavros Nikolakopoulos[1,3], Ioannis Ntzoufras[4], Dimitris Mavridis[2]
1 Department of Psychology, University of Ioannina, Ioannina, Greece
2 Department of Primary Education, University of Ioannina, Ioannina, Greece
3 Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands
4 Department of Statistics, Athens University of Economics and Business, Athens, Greece

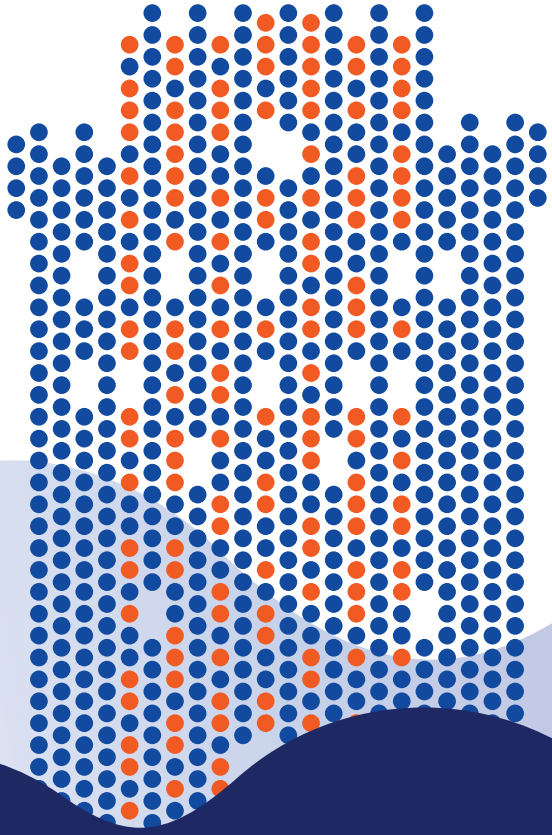
Introduction: The reliability of network meta-analysis (NMA) findings relies on the transitivity assumption, which implies a similar distribution of effect modifiers across all treatment comparisons. Statistically, this assumption is tested by comparing direct and indirect evidence for NMA treatment effects (consistency assumption). Several strategies exist to test this assumption, with inconsistency factors often introduced into the NMA model and subsequently testing if these factors are equal to zero.

Method: We present the Stochastic Search Inconsistency Factor Selection (SSIFS) method, which adds inconsistency factors connecting direct and indirect evidence, as candidate covariates in the NMA model and whose inclusion relies on a Bayesian variable selection technique. The posterior inclusion probability for each inconsistency factor gauge the likelihood of a specific comparison being inconsistent. A pivotal element of our method is the development of a reasonably informative prior regarding network consistency, drawing insights from historical data encompassing 201 published network meta-analyses. The performance of SSIFS is demonstrated through its application in two published network meta-analyses.

Findings: SSIFS can both globally and locally test network consistency. Decision-making regarding the importance of inconsistency is facilitated through posterior model odds or the application of the median probability model. Practically significant differences between direct and indirect evidence can be incorporated into the inconsistency detection process. SSIFS demonstrated robust performance, yielding to stable conclusions regarding network consistency.

Conclusion: SSIFS offers a novel Bayesian technique for the detailed examination of the consistency assumption in the NMA model, on both a global and local scale. The proposed method is accessible through the 'ssifs' R-package which is published on CRAN.

Reference: *[1] Seitidis G, Nikolakopoulos S, Ntzoufras I, Mavridis D. Inconsistency identification in network meta-analysis via stochastic search variable selection. Statistics in Medicine.2023;42(26):4850–4866. doi:10.1002/sim.9891*



21-25 July 2024
Thessaloniki Concert Hall

45th Annual Conference of the International Society for Clinical Biostatistics

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-01 Adaptive design implementation as a way to cope with recruitment difficulties

Danai Andreadi
Nestlé, Lausanne, Switzerland

Introduction: Around 55% of trials manage to successfully recruit their pre-specified sample size [1]. Performing an interim analysis for sample size reassessment allows researchers to evaluate the ongoing study's progress, estimate the treatment effect, and determine if the initially planned sample size is appropriate. This abstract provides an overview of the methods used to conduct an interim analysis on a Randomized Controlled clinical trial dealing with recruitment issues, the results obtained, and draws conclusions based on the findings.

Methods: The primary objective of the trial was to demonstrate improved survival of a probiotic in a new nutritional intervention compared to a standard nutritional product. The initial sample size was determined based on statistical power calculations, assuming a certain effect size and significance level. The interim analysis was performed when a predetermined fraction of the total planned sample size was reached.

The interim analysis design performed was an unblinded sample size re-assessment with 2 stages. At stage 1 (interim stage) the primary endpoint was computed and conditional rejection probability [2] used to re-calculate the trial size. There were three potential scenarios: recruit less subjects than planned, continue the trial as planned or stop trial recruitment for futility. The final trial primary endpoint estimates were calculated by combination of stage 1 and stage 2 estimates via the inverse normal method.

Results: The interim analysis results determined that the originally planned sample size was not sufficient to detect a statistically significant effect and there was need to increase the sample size in order to show an effect. Therefore, the trial was stopped since the trial team had agreed on the original sample size as the maximum number for recruitment.

Conclusions: In this particular study, the adaptive design implementation revealed that the initially planned sample size was insufficient and the trial was stopped. This adjustment ensured no further resources were spent and the study was completed in a timely manner, ultimately benefiting both researchers and patients.

References: *[1] Sully, Ben GO, Steven A. Julious, and Jon Nicholl. "A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies." Trials 14.1 (2013): 1-9. [2] Müller, Hans-Helge, and Helmut Schäfer. "A general statistical principle for changing a design any time during the course of a trial." Statistics in medicine 23.16 (2004): 2497-2508.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-02

A practical seamless phase II/III design for special populations and rare disease settings

Daniel Bonzo, Shenghua Yuan
LFB USA, Inc., Framingham, United States

Background/Introduction: The difficulty in conducting a clinical trial that involves pediatric subjects in rare disease settings leads to practical considerations – use of seamless Phase II/III designs, event-type of endpoints within subjects, e.g., successful treatment of bleeding episodes or presence of treatment-related adverse events, and analyses based on hierarchical models. In addition, the statistical methodology employed is enriched by data augmentation techniques such as extrapolation which may involve borrowing of information from another population to enlarge the subject base.

Methods: A practical seamless Phase II/III design is presented. The design allows for dose optimization and generation of confirmatory data which is expected of a pivotal Phase III trial. The Phase III portion of the design allows for multiple cross-over of subjects using the 2 best dose regimens tested in the Phase II portion of the design. Use of event-type endpoints in this setting leads to analytic models that consider correlation of events within subjects.

Results: An example of a model that allows for testing a success proportion for binary events is proposed. The resulting test does take into account the correlation of events within subjects and is shown to have a limiting normal distribution. The power of the test is examined through simulations considering small sample settings, different event rates, and varying subject exposure times.

Conclusion: In conclusion, one can see that the proposed design can facilitate the generation of the needed marketing authorization data in special populations and rare disease settings.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-03 A response-adaptive multi-arm design for normal endpoints using entropy-based allocation rule

Gianmarco Caruso, Pavel Mozgunov
MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom

Multi-arm trials are gaining interest in practice given the statistical and logistical advantages that they can offer. The standard approach is to use the fixed (throughout the trial) allocation ratio, but there is a call for making the allocation ratio adaptive and skewing the allocation of patients towards better performing arms. This is motivated by offering the most benefit to the patients in the trial and collecting the most of information on the most promising arms. However, among other challenges, it is well-known that these approaches might suffer from low statistical power.

We present a response-adaptive design which explicitly allows to control the trade-off between the number of patients allocated to the "optimal" arm and the statistical power. Such a balance is controlled through the calibration of a tuning parameter, and we explore various strategies to effectively perform it. We consider the general setting of a normally distributed endpoint and the design that targets a desirable value of that endpoint (which can be, for example, a composite endpoint or a score). The proposed allocation rule naturally arises from a weighted version of Shannon's differential entropy, a context-dependent information measure which gives a greater weight to those treatment arms which have characteristics close to the pre-specified clinical target. We also introduce a simulation-based hypothesis testing procedure to assess whether the best performing treatment arm is significantly superior to the second-best. This emphasises a primary focus on selecting the optimal arm rather than on a comparison to the control.

The proposed design may offer a good balance between patient benefit and statistical power, by leading to the selection of the most effective arm among currently used competitors. Insights on the proposed design are shown through an extensive simulation study, where both sequential and fixed allocation rules are considered. Among the response-adaptive alternatives, particular emphasis is placed on the one based on the Gittins Index for the multi-armed bandit problem [1], adapted here to align with the trial main objective. The simulation study highlights the potential advantage of the proposed design over the considered competitors, emphasising its capability of making the right decision in a remarkable proportion of replicated trials.The proposed design may offer a good balance between patient benefit and statistical power, by leading to the selection of the most effective arm among currently used competitors. Insights on the proposed design are shown through an extensive simulation study, where both sequential and fixed allocation rules are considered. Among the response-adaptive alternatives, particular emphasis is placed on the one based on the Gittins Index for the multi-armed bandit problem [1], adapted here to align with the trial main objective. The simulation study highlights the potential advantage of the proposed design over the considered competitors, emphasising its capability of making the right decision in a remarkable proportion of replicated trials.

Reference: *[1] Williamson, S. F., & Villar, S. S. (2020). A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. Biometrics, 76(1), 197-209.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-04

Exploring efficient drug discovery for Major Depressive Disorder: A Phase II platform trial design

Michaela Maria Freitag[1], Dario Zocholl[1], Elias Laurin Meyer[2,3], Stefan Gold[1,4], Marta Bofill Roig[2], Heidi De Smedt[5], Martin Posch[2], Franz König[2]
1 Charité - Universitätsmedizin Berlin, Berlin, Germany
2 Medical University of Vienna, Vienna, Austria
3 Berry Consultants, Vienna, Austria
4 Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany
5 Janssen Research and Development, Beerse, Belgium

Background: Major depressive disorder (MDD) is one of the most common causes of disability worldwide and the leading cause of death by suicide. Despite existing treatments, around 50% of patients do not benefit sufficiently from first-line treatment, highlighting the need for novel therapeutic options. Platform trials provide a new approach for clinical study designs that, compared to separate clinical trials, allows more compounds to be tested in a shorter period of time by, e.g., sharing controls, reducing clinical trial activation times, as well as recruitment times.

Methods: As part of the Innovative Medicines Initiative (IMI) project EU PEARL we address the design process of a Phase II platform trial for MDD, focusing on key elements such as allocation ratios, early stopping rules, and the number of concurrently active treatment arms. Extensive simulations investigated the effectiveness of different design options under varying scenarios, including potential time trends and the introduction of new compounds, selecting the optimal design elements for the specific use case. The resulting platform design was then compared to traditional two-armed randomized controlled trials to assess its potential advantages.

Results: Compared to conventional trials, the proposed platform design offers two key advantages. Simulations across different effect sizes consistently demonstrated higher power to detect genuine treatment effects and the ability to test multiple compounds concurrently reduced the sample size needed per comparison, leading to faster and more cost-effective trials.

Conclusion: This Phase II platform trial design proposes a promising approach for efficiently evaluating numerous potential MDD therapies. This could potentially accelerate treatment development and ultimately improve patient outcomes.

Reference: *[1] Williamson, S. F., & Villar, S. S. (2020). A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. Biometrics, 76(1), 197-209.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-05

Quantification of allocation bias in clinical trials under a response-adaptive randomisation procedure for normal response variables

Vanessa Ihl, Ralf-Dieter Hilgers
RWTH Aachen University, Aachen, Germany

Background: Allocation bias describes the conscious or unconscious selective allocation of patients by the recruiter in a clinical study on which arm is best and has a higher probability of being allocated. It is based on patient characteristics that influence the expected response, causes biased response, and hence influences the results of a study significantly.

Response-adaptive randomization (RAR) promises being less affected by allocation bias compared to classical approaches. Affected by the rise of platform trials, it recently enjoyed higher interest as well as first applications in studies as RAR has shown promising results in simulation studies. It is expected that more patients receive the better treatment, is more cost effective and is more attractive for possible trial participants which makes recruitment of patients easier. Especially in the field of rare diseases, where one expects to include a significantly larger proportion of all diseased persons, it is desirable to treat more patients with the better treatment during the phase II/III trials. There is nothing on allocation bias in that field up to now. Our goal is to quantify allocation bias for clinical trials with RAR procedures.

Methods: We will consider a single-centered two-arm parallel group design with a continuous primary endpoint, normally distributed, in which the doubly-biased coin design is applied for allocating the patients to a treatment arm. Further, we implement the procedure for simulations and quantify the allocation bias under a rather small patient number and special focus in rare diseases. Different assumptions for the allocation bias are investigated including strict biasing policies and higher values for the effect of the bias value.

Results: Our results so far indicate that even if the allocation bias can be very strong some approaches, the displayed procedure is hardly influenced and the responses in a simulation of a study seem to be weakly affected by allocation bias. Further simulations are still in process, the upcoming results are expected to strengthen this hypothesis.

Conclusion: RAR trials can diminish the worry about allocation bias under specific procedures. For specific strategies however, it is important to model since they are affected. It is either way useful to be able to include it in the study analysis for quantification of the allocation bias.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-06

Bias adjustment in an adaptive seamless design with different binary outcomes

Ryota Ishii[1], Kenichi Takahashi[2], Kazushi Maruo[1], Masahiko Gosho[1]
1 University of Tsukuba, Tsukuba, Japan
2 MSD K. K., Chiyoda-ku, Japan

Introduction: Adaptive seamless design combining phases II and III into a single trial consisting of two stages is gaining interest for efficient drug development. The first stage involves selecting promising treatment group(s) based on surrogate endpoint, followed by comparing the efficacy between the selected and control groups based on primary endpoint in the second stage. This design has advantages over conventional drug development, such as a shorter development timeline and fewer total number of patients. However, the presence of treatment selection leads to biased treatment effect estimator. We assume that the surrogate and primary endpoints are binary and develop bias adjustments of treatment effect estimator.

Methods: We considered two approaches that are frequently used in adaptive designs. The first involves estimating the conditional bias of the maximum likelihood estimator (MLE) and subtracting the bias estimator from the MLE. This adjusted estimator is referred to as the conditional mean-adjusted estimator (CMAE). The second approach is an unbiased estimator based on Rao-Blackwell's theorem and is referred to as the uniformly minimum variance conditional unbiased estimator (UMVCUE). We developed CMAE and UMVUE in adaptive seamless design using different binary endpoints. We also performed simulation studies to evaluate the performance of our proposed method.

Results: Our simulation results showed that MLE had an upward bias in all simulation settings. In contrast, CMAE and UMVCUE substantially reduced the bias of MLE. CMAE had a little bias because the estimator of the conditional bias of MLE had a bias. UMVCUE was unbiased in all settings which coincides with the unbiasedness from Rao-Blackwell's theorem. However, the mean squared error of UMVUE was larger than that of CMAE in most cases.

Conclusion: We developed CMAE and UMVCUE in adaptive seamless design using binary surrogate and primary endpoints. In this setting, MLE has a notable bias and CMAE and UMVCUE are appropriate estimators.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-07 Covariate-adjusted response adaptive designs for semi-parametric survival models

Ayon Mukherjee
Director and Head of Novel Trial Design Methodology, IQVIA, Berlin, Germany

Background: Covariate-adjusted response-adaptive (CARA) designs use the available responses to skew the treatment allocation towards the treatment found to be best at an interim stage of a clinical trial, for a given patient's covariate profile. In clinical trials, where the primary outcomes follow heteroscedastic or nonlinear models, CARA designs merit investigation as the optimal allocation may not be balanced across treatment arms. There has recently been extensive research on CARA designs with parametric distributional assumption on the patient responses. However, the range of application for such designs become limited in real clinical trials. Sverdlov, Rosenberger and Ryzenik (2013) has pointed out that irrespective of a specific parametric form of the survival outcomes, their proposed CAR4A designs based on the exponential model provide valid statistical inference, provided the final analysis is performed using the appropriate accelerated failure time (AFT) model. In real survival trials, however, the planned primary analysis is rarely conducted using an AFT model.

Methods: The proposed CARA designs are developed obviating any distributional assumptions about the survival responses, relying only on the proportional hazards assumption between the two treatment arms. To meet the multiple experimental objectives of a clinical trial, the proposed designs are developed based on optimal allocation approach solving a constrained optimization problem. These proposed designs serve as the optimal target allocation proportions that are functions of the Cox regression coefficients and are estimated sequentially with the arrival of every new patient into the trial. The covariate-adjusted doubly-adaptive biased coin design (Zhang and Hu, 2009) and the covariate-adjusted efficient randomised adaptive design (Mukherjee et.al. 2023) are used to randomise the patients to achieve these derived targets on expectation.

Results: The merits of the proposed designs are validated using extensive simulation studies assessing their operating characteristics for a Phase III clinical trial scenario. A power curve has also been derived demonstrating the average sample size needed to achieve the desired benefits from the proposed design. Extensive simulation study has been used to demonstrate the validity of the statistical inference from the proposed design when the assumption of proportional hazard is violated. The proposed designs have also been implemented to re-design a real-life confirmatory clinical trial to demonstrate its benefits.

Conclusion: It is concluded that the proposed CARA designs can be suitable alternatives to the traditional balanced randomisation designs in survival trials, provided response data are available during the recruitment phase to enable adaptations in the design.

References: *[1] Mukherjee A, Coad DS, Jana S. Covariate-adjusted response-adaptive designs for censored survival responses. Journal of Statistical Planning and Inference 2023; 225: 219-242. [2] Sverdlov O, Rosenberger WF, Ryeznik Y. Utility of covariate-adjusted response-adaptive randomization in survival trials. Statistics in Biopharmaceutical Research 2013; 5(1): 38–53. [3] Zhang LX, Hu Ff. A new family of covariate-adjusted response adaptive designs and their properties. Applied Mathematics- A Journal of Chinese Universities 2009; 24(1): 1–13.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-08

A fast, flexible simulation framework for Bayesian adaptive designs with time-to-event endpoints

Rainer Puhr[1], Dominique Couturier[2], Liz Ryan[1], Thomas Jaki[2,3], Stephane Heritier[1]
1 Monash University, Melbourne, Australia
2 University of Cambridge, Cambridge, United Kingdom
3 University of Regensburg, Regensburg, Germany

Introduction: In clinical trials with time-to-event endpoints, adoption of Bayesian adaptive designs has frequently faced challenges due to the lack of readily available software and the prohibitive computational demands of Markov Chain Monte Carlo (MCMC) often employed to calculate the posterior distributions. Here we present a major extension for time-to-event data to the Bayesian Adaptive Trial Simulator (BATS) R package which provides a flexible, modular structure for the fast simulation of Bayesian adaptive multi-arm multi-stage (MAMS) designs.

Methods: We take advantage of integrated nested Laplace approximation (INLA) and its efficient implementation in R to perform approximate Bayesian inference in latent Gaussian models. The versatile simsurv package is used to simulate survival data from standard parametric distributions (exponential, Weibull, and Gompertz), two-component mixture distributions, or user-defined hazard functions. The package also allows for standard or user-defined accrual and censoring functions with fixed and time-varying rates, a variety of adaptations such as stopping arms or trials for efficacy and/or futility and fixed or response-adaptive randomisation, and interim analysis schedules based on observed events, calendar time or participants recruited.

Results: We demonstrate that BATS is an effective tool to study the operating characteristics of designs with time-to-event endpoints and most common adaptations – all customisable with user-defined rules. Further benefits include flexibility, fast computation via parallel processing and possible use on stand-alone or cluster computers.

Conclusions: The BATS survival extension provides a fast and flexible simulation framework for Bayesian adaptive designs with time-to-event data.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-09 Speeding up the clinical studies with biomarker-based randomisation or adaptive Bayesian design with biomarker enrichment

Hong Sun, Irina Irincheeva
Bristol-Myers-Squibb, Boudry, Switzerland

Background: In the development of immunotherapies and/or target therapies in oncology, identification of biomarker-based patient groups is a crucial step that enables delivering the right medicine to the right patients in relatively early time. However, it takes several years for a biomarker to be identified and validated as a criterion for group assignment in both randomizations and data collections in a clinical trial. It is challenging to statistically elaborate on the choice of the threshold value for continuous biomarkers, usually a sparse number of values are being evaluated based on the available knowledge using "naïve" statistical approach. For example, CD8 is a well-established biomarker in breast cancer and colorectal cancer, multiple studies use biomarker stratified design with CD8 threshold 1% or 2% to test the hypothesis about the investigational treatment performance in CD8 expressed and non-expressed subgroups. The early dichotomization leads to ignoring actual distribution of the continuous values as well as the "grey zone" of values with potential supportive argument.

Method: In this work we investigate different scenarios of chosen thresholds of CD8 with simulations inspired by existing studies with survival endpoints. For each given threshold value, with fixed considered sample sizes and similar biomarker distributions, if the measurement standards change due to different labs, the variability of the values may impact the power of the study. The range of variabilities and the corresponding power of the study are being tested using an adaptive (Bayesian) biomarker enrichment design.

References: *[1] Simon N, Simon R. Adaptive enrichment designs for clinical trials. Biostatistics. 2013 Sep;14(4):613-25. doi: 10.1093/biostatistics/kxt010. Epub 2013 Mar 21. PMID: 23525452; PMCID: PMC3769998. [2] Ruitao Lin, Peter F. Thall & Ying Yuan (2021) BAGS: A Bayesian Adaptive Group Sequential Trial Design with Subgroup-Specific Survival Comparisons, Journal of the American Statistical Association, 116:533, 322-334, DOI: 10.1080/01621459.2020.1837142.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-10 LB

Adaptive quantification of prior impact in terms of effective sample size

Silvia Calderazzo
German Cancer Research Center, Heidelberg, Germany

Background: Bayesian clinical trials facilitate incorporation of external information via the specification of informative prior distributions. One main concern in this respect is the potential for prior data conflict, i.e., the possibility that inferences about the current trial are worsened by borrowing of heterogeneous external information. An additional concern is that of prior domination, i.e., that inferences in the current trial are mostly driven by prior information. Various Effective Sample Size (ESS) measures have been proposed in the literature to translate prior informativeness in terms of a certain number of samples added to the analysis. These measures are however typically independent of the current data, and thus cannot reflect the potential for 'harmful' information.

Methods: In previous work, we built on Reimherr et al (2021) to relate prior information to a number of (virtual) samples from the current data model. We now further develop such a dynamic ESS measure, and explicitly additionally tailor it to the aim of hypothesis testing. The measure quantifies the number of added or subtracted (if potential conflict is observed) samples to the current analysis.

Results: We show applicability of the methods on simulated data at both a trial planning stage, assuming different degrees of prior-data conflict, and after observing hypothetical data. Additionally, a case study example is investigated. Here, the method is proposed as a tool to provide a complementary characterization of the role of the prior weight when a robust mixture prior is adopted for the analysis, and can be used in addition to, e.g., a tipping point analysis.

Conclusions: Assessment and communication of prior impact is of great importance, and easily interpretable metrics can facilitate such task. Prior information can have different impact on different inferential targets. We propose an effective sample size metric which aims to characterize the impact of prior information on specific targets, while reflecting observed prior-data conflict.

Reference: *[1] Reimherr, M., Meng, X. L., & Nicolae, D. L. (2021). Prior sample size extensions for assessing prior impact and prior-likelihood discordance. Journal of the Royal Statistical Society Series B: Statistical Methodology, 83(3), 413-437.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-12

Mind the Gap: A scoping review on Machine Learning and Artificial Intelligence in the design of Clinical Trials

Ajsi Kanapari, Giulia Lorenzoni, Honoria Ocagli, Dario Gregori
Unit of Biostatistics, Epidemiology and Public Health, UBEP, University of Padova, Italy

Background: Machine Learning (ML) and Artificial intelligence (AI) methods have not only gained popularity in many sectors, but in the last decade also in healthcare and clinical research (Weissler et al., 2021). Several reviews and US Food and Drug Administration (FDA) discussion papers have examined the potential contributions that advanced techniques may have in clinical studies and particularly by improving their efficiency. The application of ML and AI has been encouraged in site selection, patients' recruitment, treatment assignment, monitoring for adherence, detection of endpoint and retention, and analysis. This scoping review aims to provide insight on the usage of advanced ML and AI techniques in the design of clinical trials and acknowledge the gap between potential and actual applicability.

Methods: A scoping review framework was chosen. The search has been performed on PubMed, Embase (Ovid), Scopus and grey literature without time constraints and by using three main concepts: Randomized Clinical trials (RCT). ML and AI methods and potential applications in design and conduct. Screening has been performed on Covidence and supported by semi-automatic procedures.

Results: We have found 5 articles that implemented AI methods in recruitment, all in oncology in which the matching of patient and trials. As concerns the prediction and stratification of patients by risk group and similar probability of outcome 8 articles have been found, of which 5 regard ai assisted procedures for prediction. 1 application is found directedly on trials for stratified randomization, in fact it is rather performed on known risk groups from pre-clinical studies. One of popular contribution regards AI-assisted procedures as trial interventions (n=7), however should be involved a discussion if these trials are appropriate for the aim of this work. It was followed by imaging (n=6) used for diagnostic and screening purposes. Monitoring for adherence with AI/ML has been applied in 7 studies thanks to the passive collection of data from wearable and other devices, mainly in cardiology, psychiatry, and for monitoring physical activity. As concerns, the analysis of data with ML and AI methods only 2 articles have been found. As for retentions and site selection, no application was found in Clinical Trials even though they were cited in many reviews.

Conclusions: There is a mismatch between the largely discussed potentialities of ML and AI for the design of clinical trials and their actual application: the new methodologies seem to be still limited to preclinical studies or post-hoc analysis.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-13

Stratified propensity score randomisation: The effect of integrating machine learning in the clinical trial design

Ajsi Kanapari, Giulia Lorenzoni, Dario Gregori
Unit of Biostatistics, Epidemiology and Public Health, UBEP, University of Padova, Padova, Italy

Introduction: The choice of the randomization method is a core step in the design of Randomized Clinical Trials (RCT), that aims to minimize imbalances and predictability in data. Stratified randomization is an appropriate solution in the case of prognostic factors and biomarkers that may influence the outcomes and therefore cause confounding on the treatment effect. However, an increased number of factors and consequently of strata may lead to sparsity in the data. We propose a randomization process by strata of propensity score (PS) for the outcome, to enhance balance of multiple covariates particularly in small trials.

Methods: A simulation in the context of a Nutritional study explores stratified propensity score randomization by training and testing Machine Learning models on an independent registry of historical data. Consequently, these models are applied on patients of two-arm trial setting with binary outcome for estimating their PS. Quintiles of PS have been built and equal randomization has been performed by each stratum. To evaluate the ability of different algorithms to capture complex relationships with the outcome have been considered logistic regression, random forest, naïve bayes, boosted regression trees. Moreover, for assessing the impact of differences that the study population may have from historical data also a Gamma generating distributions for covariates have been used. For each configuration 10000 simulations have been performed.

Results: The study highlights the gains with PS stratified randomization, in terms of power, particularly in the context of small trials (n<200) with Naïve Bayes that outperforms the others ML techniques.

Conclusion: The gains of integrating ML techniques leads to an increase in power particularly in small studies in which the balance of multiple covariates is enhanced. This is particularly true in the case of heterogeneity between registry and trial data.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-14 Augmenting control arm of randomised-controlled trials by incorporating information across multiple external sources with stratified propensity score and data-driven mixture prior

Xun Xu, J. Jack Lee
University of Texas MD Anderson Cancer Center, Houston, United States

Background: Interest in augmenting randomized controlled trials by supplementing the concurrent control arm with external data sources in drug development has been rapidly growing. However, external data may lack between-population exchangeability due to varying eligibility criteria, different distributions of confounding covariates, and other factors. Despite these challenges, to facilitate proper information borrowing, we propose approaches with stratified propensity score and data-driven mixture prior.

Methods: We stratify patients in both concurrent and external controls based on their propensity scores to address the effect of heterogeneity and mitigate observed confounding. Within each stratum, we calculate the propensity score overlap coefficient to inform the parameters of the meta-analytic predictive (MAP) prior to account for the between-group heterogeneity. To reduce unobserved confounding and address prior data conflict at the outcome level, we construct self-adaptive mixture prior and calibrated elastic mixture prior. Combining the two steps, the stratified propensity score self-adaptive mixture (SPS-SAM) prior and the stratified propensity score calibrated elastic mixture (SPS-CEM) prior are proposed. The mixture prior is composed of an MAP prior and a vague prior. To obtain the weight measuring the extent of congruent between the current and external data, SPS-SAM prior uses the likelihood ratio test and SPS-CEM prior uses the scaled t test, respectively. Adaptive weight dynamically controls the proportion of the MAP prior.

Results: Simulations show that, under various data heterogeneity scenarios, SPS-SAM prior and SPS-CEM prior can reduce the bias and mean square error in estimating both stratum-specific and overall treatment effects compared to other available methods. SPS-CEM outperforms SPS-SAM in reducing the bias and resulting in larger calibrated power. This superiority can be attributed to the more accurate measurement of the within-stratum weight of the MAP prior component in the mixture prior.

Conclusions: By combining the stratified propensity score and data-driven mixture prior, multiple external sources can be incorporated to augment the control arm of randomized-controlled trials to obtain accurate, efficient, and robust estimation of the treatment effect.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-15 Efficient group sequential design: Harnessing dynamic historical borrowing in medical device trial

Maria Vittoria Chiaruttini, Giulia Lorenzoni, Dario Gregori
University of Padova, Padova, Italy

Introduction: The field of medical device technology evolves rapidly with shorter lifecycles compared to pharmaceuticals. This acceleration emphasizes the need for swift safety and efficacy assessments to match technological advancements. In producing new evidence, leveraging clinical data from previous device versions is crucial, especially as medical devices are often built upon similar mechanisms. In this context, Bayesian statistics represents a valuable approach since it naturally combines prior information with current information on a quantity of interest. Mainly, Bayesian Dynamic Borrowing (BDB) has emerged as an approach that can adjust the weight of historical information as current data accumulates, depending on the congruence between past and new data allowing for unbiased data augmentation.
Moreover, at the design stage, the synergy between BDB and group sequential design (GSD) can improve the operating characteristics of the trial and reduce the sample size.

Methods: This study delves into BDB employing two innovative methods—Normalized Power Prior (NPP) and Self-Adapting Mixture (SAM) prior—in conjunction with the GSD theory within a noninferiority trial design to assess the operating characteristics under congruence and incongruence scenarios between past and current data. A cardiovascular medical device trial serves as a motivating example, showcasing the practical application of these methods.

Results: Simulations highlights the efficiency of the GSD-BDB design due to augmented data and early stopping rules while keeping the error rate at the nominal level. Specifically, simulating the congruent scenarios, under the Null hypothesis 90% of trials stop for futility and under the Alternative, 78% stop for efficacy. On the other hand, simulating the incongruent scenarios, NPP outperforms SAM, demonstrating greater control over Type I error.

Conclusion: The results emphasize the advantages of BDB, especially when integrated with GSD, in the rapidly evolving landscape of the medical device development chain. The analyzed scenarios underscore the critical importance of carefully selecting the most suitable dynamic approach for specific trial hypotheses and the GSD boundaries, with the aid of simulations.

References: *[1] FDA. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices (2017). [2] Yang, P., Zhao, Y., Nie, L., Vallejo, J. & Yuan, Y. SAM: Self-adapting mixture prior to dynamically borrow information from historical data in clinical trials. Biometrics biom.13927 (2023) doi:10.1111/biom.13927. [3] Ibrahim, J. G., Chen, M., Gwon, Y. & Chen, F. The power prior: theory and applications. Statistics in Medicine 34, 3724–3749 (2015).*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-16

Single-arm hierarchical testing with historical controls

Dimitris Karlis[1], Urania Dafni[2,3], Tereza Dellaporta[2,3]
1 Athens University of Economics and Business, Athens, Greece
2 University of Athens, Athens, Greece
3 Frontier Science Foundation Hellas, Athens, Greece

Regulatory agencies now recognize single-arm trials with external historical controls to assess promising treatments for rare or specific indications. This has been quite common in oncology trials, for example. Incorporating historical data in single-arm trials has the potential to accelerate drug development and reduce the number of patients enrolled, compared with standard randomized controlled clinical trials. Single- arm trials do not include concurrent controls, and therefore counter-factual information needs to be found in external sources. This has led to a range of different methodologies utilizing information from historical controls. In this work we suggest an alternative design that allows for both utilization of HC and early termination of a non-promising new phase II trial. The approach is based on hierarchical testing of different hypotheses based on accumulated information. We derive the optimal sample size required for assessing the efficacy of a new drug. This assessment occurs at two different time points: after the enrollment and monitoring of a small number of patients and once all patients are evaluable provided that the study satisfies the criteria for continuation. Results show that the approach can reduce the expected number of patients while control for the uncertainty around a value derived from historical controls.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-17 A 1:1:1 matching design to account for centre-specific prophylaxis/treatment administration in observational studies

Cristina Marelli[1], Alessio Signori[2], Daniele Roberto Giacobbe[1,3]
1 Clinica Malattie Infettive, IRCCS Ospedale Policlinico San Martino, Genoa, Italy
2 Section of Biostatistics, Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy
3 Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy

Introduction: Propensity score matching (PSM) is an established method employed in observational studies to reduce the bias due to differences in the distributions of baseline measured characteristics between cases and controls. In multicenter studies, centers may adopt different therapeutic guidelines to treat the same disease. We hypothesised a setting where one center administered a specific prophylaxis/treatment to patients affected by some disease (case group), while two other centers administered a different prophylaxis/treatment or no prophylaxis/treatment to patients affected by the same disease (control groups). Therefore, we developed a 1:1:1 matching method to jointly match a case to two controls, one per each control group, to maintain the center effect on a given outcome of interest. We aimed to compare this method to the common 1:1 PSM and assess its clinical implications.

Methods: The proposed 1:1:1 PSM algorithm employed a multinomial logistic regression to predict the generalized propensity scores, which were then used to compute the perimeter of the triangle formed by each triplet of patients. The resulting matched cohort was composed by the trios showing the smallest perimeter. We performed a simulation study varying the prevalence of disease in the case group and the risk difference between controls and cases. The two matching methods were compared assessing the standardized difference for measured confounders, a measure of overall covariate distance (standardized distance, SD), and the relative bias, the percentual reduction in bias, and the mean square error for the treatment effect. Finally, we preliminarily applied these methods to a real cohort of patients treated for Klebsiella pneumoniae bloodstream infection in three different centers.

Results: On the 750 simulated datasets, in average, the SD was equal to 0.05±0.01 and 0.15±0.07, and the percentual reduction in bias was equal to 83.94% and 71.91%, for 1:1 and 1:1:1 PSM respectively. Only 1:1:1 PSM enabled to also balance the control groups, showing an average SD equal to 0.14±0.07 (vs. 0.20±0.07 in 1:1 PSM). With the limitation of the small sample (18 cases matched with 36 controls), the case study only partially supported what was observed in the simulation study.

Conclusion: While the simulation study showed that both matching methods are efficient, balance between the two different control groups was more pronounced in the 1:1:1 matching. A larger case study is nonetheless warranted to support the proposed 1:1:1 PSM as the preferred method when also balancing controls from different centers is deemed as crucial.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-18

Random weights for historical control enrichment

Guido Moreira
Bavarian-Nordic, Munich, Germany

Background: Using historical data to enrich the control branch of clinical trials has become viable and intuitive with the use of Bayesian inference and adequate prior information. Part of the prior knowledge that is necessary for the either the ratio of within and between dataset variances or the weights of the historical data. Although the use of weights is easier to arbitrate, practical examples show that weight choice may induce too much influence from the historical data, sometimes even obfuscating the collected data. This work addresses this issue in an attempt to balance arbitration and robustness.

Methods: By setting a prior on the weights, the posterior distribution becomes more complex, but reduces the impact of the arbitration. An MCMC approach is used to obtain samples of the posterior distribution and make inference. A Dirichlet distribution is chosen as a prior, not only due to its adequate domain, but also because it can be reparametrized in a way that permits explicitly defining its uncertainty.

Results: The outcomes of a series of carefully crafted simulation studies are presented. In each scenario, it becomes clear how much the prior weights and uncertainty influence the posterior.

Conclusion: A list of recommendations is provided for a few scenarios based on the simulated studies.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-19

Basket trials in very rare diseases: Are they feasible?

Giulia Risca[1], Haiyan Zheng[2], Maria Grazia Valsecchi[1], Stefania Galimberti[1]
1 Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, Department of Medicine and Surgery, University of Milan-Bicocca, Monza, Italy
2 Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

Background: In recent years, innovative designs have been proposed in the context of personalized medicine to study the effect of single/multiple drugs on multiple/single sub-populations simultaneously. Specifically, basket trials are used to study a single targeted therapy in multiple diseases or disease types sharing common genetic characteristics. This approach is very useful in rare diseases, where basket trials allow for a more efficient analysis due to borrowing of information across sub-trials. Explicitly, the treatment effect in each sub-trial may provide information on the treatment effect in other sub-trials. Our aim is to assess the feasibility and robustness of a basket design in the setting of very rare diseases, an even more challenging situation.

Methods: The clinical setting of interest involves a few trials, each one using the same therapy on very small numbers of patients with different diseases. We evaluated the standard Bayesian approaches on a binary and a continuous endpoint. To achieve concurrent borrowing of information, several methods for a binary outcome have been considered: i) the standard Bayesian hierarchical model [1], ii) the exchangeability-nonexchangeability model by Neuenschwander [2] and iii) its recent improvement given by Daniells [3]. The case of a continuous endpoint, which might be very informative when the innovative treatment is highly effective, has been addressed with the strategy proposed by Ouma [4]. Our simulation study considered various scenarios characterised by the treatment effects of about five sub-trials, each involving a maximum of 15 subjects.

Results: The results of the simulation study suggest that basket trials are feasible even in the context of very rare diseases, especially when the effect size is large and consistent across sub-trials.

Conclusions: Operating characteristics of the different approaches showed promising results. This encourages our further work to investigate leveraging external information both on controls and on trials conducted with the same treatment on different diseases.

References: *[1] Berry SM, Broglio KR, Groshen S, Berry DA. Bayesian hierarchical modeling of patient subpopulations: Efficient designs of Phase II oncology clinical trials. Clinical Trials. 2013;10(5):720-734. doi:10.1177/1740774513497539. [2] Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. Pharm Stat. 2016 Mar-Apr;15(2):123-34. doi: 10.1002/pst.1730. [3] Daniells L, Mozgunov P, Bedding A, Jaki T. A comparison of Bayesian information borrowing methods in basket trials and a novel proposal of modified exchangeability-nonexchangeability method. Statistics in Medicine. 2023;42(24):4392-4417. doi: 10.1002/sim.9867. [4] Ouma, L.O., Grayling, M.J., Wason, J.M.S. & Zheng, H. (2022) Bayesian modelling strategies for borrowing of information in randomised basket trials. Journal of the Royal Statistical Society: Series C (Applied Statistics), 71(5), 2014–2037. Available from: https://doi.org/10.1111/rssc.12602.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-20

U-DESPA: A utility-based Bayesian approach for dosage optimisation handling PK, PD safety and efficacy in oncology clinical trials

Marie-Karelle Riviere[1], Anais Andrillon[1,2], Moreno Ursino[3], Pavel Mozgunov[4], Anne Bellon[5], Esteban Rodrigo Imedio[5], Sandrine Micallef[5]
1 Saryga, France
2 INSERM U1153 Team ECSTRRA, France
3 INSERM, Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, Paris, France
4 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
5 Debiopharm International SA, Lausanne, Switzerland

With the development of novel therapies such as Molecularly Targeted Agents (MTAs) and immunotherapies, the MTD paradigm that "more is better" does not necessarily hold anymore. In this context, doses and schedules of novel therapies may be inadequately characterized and oncology drug dose finding approaches should be revised. In January 2023, in the frame of the interdisciplinary Project Optimus, FDA issued a draft guidance requiring new strategies of dosage optimization prior to initiating registration trials in oncology. In this guidance, the dosage optimization is proposed to rely on a quantitative assessment of the relationship between dosage and relevant endpoints. We developed a Bayesian dose finding design allowing to 1. Directly determine the optimal dosage at the end of the dose escalation phase, or 2. Use of dedicated dose finding cohorts randomizing patients to candidate optimal dosages after safe dosages have been found. This Bayesian dose finding design relies on a dose exposure model built from pharmacokinetic data using non-linear mixed effect modeling approaches. Three models are also built to assess the relationships between exposure and the probability of different relevant endpoints on safety, pharmacodynamics and anti-tumor activity. These models are then combined to predict the different endpoints for every candidate dosages. A utility function is finally proposed to quantify the trade-off between these three endpoints and to determine the optimal dosage. We perform an extensive simulation study to evaluate the operating characteristics of the method. Based on these outcomes, this approach is planned to be applied on a dose finding clinical trial to support decision on the dosage to be further used for late-stage development.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-21

An efficient approach to operational prior specification in Phase I dose-combination escalation trials

Weishi Chen, Pavel Mozgunov
University of Cambridge, Cambridge, United Kingdom

Background: Recent years have seen a rising interest in combining several drug agents and their schedules. The selection of the right doses of each agent is a challenging task. Several methods for Phase I dose-escalation combination trials have been proposed, such as the two-dimensional Bayesian logistic model (2BLRM) and the partial ordering continual reassessment method (POCRM). However, these methods require a number of hyperparameters to be specified. These parameters are calibrated based on the statistical considerations (aka operational prior). This is conventionally done via "grid search" that specifies a grid for each parameter and considers all possible combinations. However, the computational cost of grid search grows multiplicatively with the number of parameters. Very little attention has been given into investigating more efficient calibration methods of combination designs and what combination-toxicity scenarios should be used with these.

Methods: We present a Bayesian partial ordering logistic model (POBLRM), which combines the idea of partial ordering and the more flexible (than CRM) two-parameter logistic model. We will also introduce a novel "cyclic" calibration method, which optimise the parameters one-at-a-time. Compared to grid search, the computation cost of cyclic calibration reduces from multiplicative to additive. A new method to reduce the number of calibration scenarios based on scenario complexities has also been proposed.

Results: Simulation studies show that the cyclic algorithm based on only 2 calibration scenarios converges quickly, can reduce the computation cost by more than 500 folds, while still selects operational prior parameters that give equally good operational characteristics compared to a "grid search" approach. A comparison between our POBLRM under novel calibration methods with the 2BLRM and POCRM models will be presented. We show that, on average, the POBLRM has better operational characteristics than the other methods, and it gives more even selection of the target combinations than the 2BLRM.

Conclusion: To meet various needs of drug development, more flexible model are proposed that leads to computational challenges during the calibration process. The cyclic calibration method, together with the calibration scenario selection criteria, can largely reduce the computational cost and makes calibration feasible for more complicated models.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-22 Bayesian hierarchical model for dose-finding trial incorporating historical data

Linxi Han[1], Qiqi Deng[2], Zhangyi He[3], Frank Fleischer[4], Feng Yu[1]
1 University of Bristol, Bristol, United Kingdom
2 Boehringer Ingelheim Pharmaceuticals Inc, Ridgefield, United States
3 Imperial College London, London, United Kingdom
4 Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

Background/Introduction: The Multiple Comparison Procedure and Modelling (MCPMod) approach has been shown to be a powerful statistical technique that can significantly improve the design and analysis of dose-finding studies under model uncertainty. Due to its frequentist nature, however, it is difficult to incorporate information into MCPMod from historical trials on the same drug. Bayesian MCPMod, a recently introduced Bayesian version of MCPMod, is designed to take into account historical information on the placebo dose group. However, there are shortcomings when extending to multiple dose groups because the historical information on different dose groups is assumed to be independent.

Methods: To address this problem, we introduce a Bayesian hierarchical framework capable of incorporating historical information on an arbitrary number of dose groups, including both placebo and active ones. It assumes that responses at different dose levels for current and historical trials come from the same underlying dose-response curve while accounting for between-trial heterogeneity in prognostic and predictive effects. We expect our proposed model to automatically adjusts the level of information borrowed from historical trial based on the homogeneity of treatment effects in the current and historical trials. When the historical and current trials are homogeneous, strong information borrowing occurs, resulting in increased power; when the two trials are heterogeneous, little or no historical borrowing occurs, thereby controlling for type I error rates. This model is particularly useful in situations where the effect sizes of two trials are different.

Results: Through simulation studies, we compare the performance of our proposed Bayesian hierarchical model (BHM) against the Bayesian pooling model (BPM), where trial data are pooled without differentiation. We show that BHM demonstrate superior performance in scenarios with homogeneity in treatment effects across two trials, enabling strong information borrowing and resulting in increased statistical power. Conversely, in cases of heterogeneity between trials, BHM effectively controls type I error rates, outperforming the BPM. And BHM yields more accurate estimates of treatment effects compared to the BPM, particularly when effect sizes differ between trials. The use of heavy-tailed hyperpriors for prognostic heterogeneity and truncated normal priors for predictive heterogeneity ensures robustness against prior distribution sensitivity, ensuring that power is not adversely impacted by prior-data conflicts.

Conclusion: The proposed BHM offers a robust and flexible framework for incorporating historical data across multiple dose levels, reducing the necessary sample size in the dose-finding trial while maintaining its target power.

P-A01-23 Joint TITE-CRM: A design for dose finding studies for therapies with late-onset safety and activity outcomes

Thomas Jaki[1,2], Oliver Boix[3], Dimitris Kontos[3], Helen Barnett[2,4]
1 University of Regensburg, Regensburg, Germany
2 University of Cambridge, Cambridge, United Kingdom
3 Bayer AG, Germany
4 Lancaster University, Lancaster, United Kingdom

In Phase I/II dose-finding trials, the objective is to find the Optimal Biological Dose (OBD), a dose that is both safe and shows sufficient activity that maximises some optimality criterion based on safety and activity. In cancer, treatment is typically given over several cycles, complicating the identification of the OBD as both toxicity and activity outcomes may occur at any point throughout the follow up of multiple cycles. We present and assess the Joint TITE-CRM, a model-based design for late onset toxicities and activity based on the well-known TITE-CRM. It is found to be superior to the currently available alternative designs that account for late onset bivariate outcomes, as well as being both intuitive and computationally feasible.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-24 Allocation bias in group sequential designs

Daniel Bodden[1], Nicole Heussen[1,2], Ralf-Dieter Hilgers[1]
1 RWTH Aachen University, Aachen, Germany
2 Sigmund Freud Private University, Vienna, Austria

Background: Even though randomized controlled clinical trials are considered the gold standard, they are still susceptible to various biases. The FDA and EMA emphasize the need to mitigate theses biases, even in randomized controlled trials. While allocation bias has been investigated for various trial designs, its impact on group sequential designs remains unexplored.

Methods: We investigate the influence of allocation bias on the testing decisions in group sequential clinical trials across different randomization procedures. A statistical model based on a biasing policy is introduced to calculate the Type I Error probability for different randomization sequences. Monte Carlo simulations are used to evaluate Type I Error probabilities under different randomization procedures when allocation bias is present.

Results: For a Lan & De Mets alpha spending function mimicking the Pocock or the O'Brien and Fleming design, we found the following results: among the evaluated randomization procedures when allocation bias is present, permuted block randomization with a block length of 4 yielded the highest Type I Error probability. This was followed by Chen's design with p=0.67 and an imbalance factor of 3 and Efron's biased coin design with p=0.67. Complete randomization and random allocation rule only led to a very slight increase of Type I Error probability.

Conclusion: In group sequential trials, the selection of an appropriate randomization procedure is crucial. For clinical trials adopting a Lan & DeMets approach with an alpha spending function mimicking Pocock or O'Brien and Fleming designs, our findings suggest the utilization of the random allocation rule or complete randomization when allocation bias cannot be ruled out. In scenarios where additional forms of bias are anticipated, a thorough evaluation of advantages and disadvantages of each randomization procedure becomes essential to ensure trial integrity and validity.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-25 Internal pilot sample size re-estimation for the B-free trial - A cluster randomised crossover trial

Shun Fu Lee[1,2], **Jessica Spence**[1,3]
1 Population Health Research Institute, Hamilton, Canada
2 Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada
3 Anesthesia, McMaster University, Hamilton, Canada

Background: The intra-cluster correlation (ICC) is the key parameter used to determine sample size in cluster randomized trials. Estimating ICC values can be challenging; estimates are often informed by epidemiologic patterns rather than derived from clusters participating in a planned trial. Even slight variations in ICC can significantly impact the sample size of a cluster randomized trial, such that incorrect assumptions can lead to an underpowered trial. We describe a method for mid-study sample size re-estimation, illustrated using our experience in the Benzodiazepine-free Cardiac Anesthesia for Reduction of Postoperative Delirium (B-Free) with a cluster randomized crossover design.

Methods: Our blinded sample size re-estimation can be summarized as follows: 1) calculate the initial sample size; 2) collect a portion of the data; 3) estimate the ICC using available data; 4) re-calculate the sample size based on the observed ICC; 5) increase the number of cluster-periods if needed. We used simulation to ensure that the Type I error remained at 5% at the time of full trial analysis. Furthermore, we utilized simulation to explore the performance of trials with the same design but varying amounts of available data, ICC, and prevalence of the outcomes.

Results: Our dataset included 11,222 patients at 20 clusters, with an average of 9.5 periods per cluster. We observed an event rate of 16.6%, a projected average cluster size of 750, an ICC of 0.06 (95% confidence interval (CI) 0.05, 0.07), and power of 70%. Simulation showed that if 11 clusters completed an additional 64 crossover periods, power increased to 81%, with a type I error of 5% within the reasonable range. According to the simulation results, we maintained a reasonable Type I error rate and achieved sufficient power if trials were underpowered when applying this approach across various scenarios.

Conclusions: Blinded sample size re-estimation can be used to evaluate the validity of assumptions used to inform the sample size of cluster randomized trials.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-26

Sample size formula for Phase I dose-escalation combination trials

Pavel Mozgunov
University of Cambridge, Cambridge, United Kingdom

Background/Introduction: One of the most pressing tasks for a trial statistician is to provide the sample size calculation to answer the particular research question. For many trial settings, either closed formulae for the sample size or analytical (i.e. fast) implementations are available. However, in Phase I dose-escalation trials, the sample size calculations are typically performed via simulations: various sample sizes are tried until the desirable operating characteristics are obtained. This, however, can be computationally intensive and resource-demanding due to the sequential nature of such trials (and many analysis conducted after each small cohort of patients). The simulation-based approach also depends on the design/model and simulation scenarios (that might not be defined yet). These challenges are amplified further in the setting of combination trials where more designs parameters should be selected and broader range of combination-toxicity scenarios should be considered.

Methods: In this talk, we propose a sample size formula for Phase I dose-escalation combination trials. The sample size formula is based on the concept of the non-parametric optimal benchmark for dose-escalation trials that provides the upper bound of a design accuracy in a given scenario. As a result, the proposed approaches do not depend on the particular design or parameter (or semi-parametric) model but only on the i) target toxicity level, ii) number of combinations, and (iii) on the toxicity margin.

Results/Conclusion: We then illustrate on how this sample size formula can be used in several combination setting and compare its performance to a simulation-based approach. Specifically, we will show that the proposed approach can reliably estimate the proportion of correct combination selection for various different commonly used dose-escalation designs.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-27 Quantifying the impact of allocation bias in randomised clinical trials with multi-component endpoints

Stefanie Schoenen[1], **Nicole Heussen**[1,2], **Ralf-Dieter Hilgers**[1]
1 RWTH Aachen University, Aachen, Germany
2 Sigmund Freud University, Faculty of Medicine, Vienna, Austria

Background: In randomised clinical trials, prior knowledge of treatment group allocations may introduce allocation bias, affecting trial validity. The impact of allocation bias depends on the chosen randomisation procedure. Therefore, selecting a suitable randomisation procedure in the planning phase of a clinical trial should be based on mitigating bias. Approaches exist for quantifying allocation bias for normally distributed single endpoints and for multiple endpoints assessed individually. However, no previous studies have analysed allocation bias in trials with multi-component endpoints. Multi-component endpoints combine several pre-defined components into a single global test problem. This offers the advantage of capturing the full risk-benefit effect of a treatment without creating multiplicity issues. Typical applications of multi-component endpoints are multifaceted diseases.

Methods: To evaluate the impact of allocation bias in trials with multi-component endpoints, we developed a biasing policy based on the guessing strategy of Blackwell and Hodges [1], which assumes that the next patient will be allocated to the group with fewer prior assignments. EMA recommends in the ICHE9 guideline [2] that bias effects should be analysed based on their potential contribution to the p-value. Therefore, we used simulations to determine the biased type I error rates of the O'Brien Ordinary Least Square test and the Wei Lachin test, two traditional methods for assessing treatment effects of multi-component endpoints, under different randomisation procedures.

Results: Allocation bias inflates the type I error rates of the O'Brien Ordinary Least Square test and the Wei Lachin test. Even small bias effects cause exceeding the 5% significance level. The amount of inflation depends on the chosen randomisation procedure and the number of components combined in the multi-component endpoint. More components lead to increased error inflation.

Conclusion: Analysing bias effects during the planning phase of a clinical trial and choosing a bias-mitigating randomisation procedure enhances the validity of the trial. Thus, the developed methodology can be applied to base the selection of a randomisation procedure on scientific arguments and to facilitate the design of more robust and valid trials, especially in the context of trials with multi-component endpoints.

References: *[1] Blackwell, D., and Hodges, J. L. (1957) "Design for the Control of Selection Bias." The Annals of Mathematical Statistics 28, no. 2: 449–60. [2] ICH E9: Statistical principles for clinical trials (1998). https://database.ich.org/sites/default/files/E9 Guideline.pdf Accessed 28 Feb 2024*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-28 Sample size determination considering the functional relationship between co-primary endpoints in clinical trials

Takashi Sozu[1], Hitoseki Takasu[2]
1 Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Japan
2 Department of Information and Computer Technology, Graduate School of Engineering, Tokyo University of Science, Japan

Background: Multiple primary endpoints (variables) are often used to evaluate the efficacy of a test treatment against that of a control treatment in clinical trials. Typically, these are observed and evaluated independently. However, one primary endpoint is often defined as a function of another, depending on the strategy used to evaluate multiple aspects of the treatment effect. For example, the following two co-primary endpoints were used in a randomised, placebo-controlled clinical trial assessing the efficacy and safety of 2.4 mg semaglutide versus 1.0 mg semaglutide and placebo for weight management: (1) percentage change in body weight from baseline to week 68 and (2) achievement of weight reduction of at least 5% of baseline weight at week 68 [1]. Under such conditions, no studies exist on power and sample size (the required number of subjects in clinical trials) calculations that consider the correlation between the two co-primary endpoints.

Methods: We proposed a sample size determination method that considers the functional relationship between two co-primary endpoints wherein one primary endpoint is determined as a dichotomous variable of the other continuous primary endpoint. We theoretically derived the correlation coefficient between the two endpoints and between the two test statistics after formulating the primary endpoints and framework of hypothesis testing including the test statistics. We calculated the overall study power (the joint probability of demonstrating treatment effects on the two endpoints) using the derived correlation coefficients and assuming that the two test statistics followed a bivariate normal distribution. The validity of the calculated power was confirmed through simulation experiments.

Results: We evaluated the operating characteristics of the power and sample size calculated using the proposed method through numerical examples. The overall power was almost equal to the marginal power of the binary endpoint because the conditional power of the continuous endpoint was almost equal to 1 when the marginal power of the binary endpoint achieved the targeted power (0.8 or 0.9).

Conclusion: The cutoff value for the continuous variable that determines the probability of success of the binary endpoint should be carefully determined from clinical and statistical perspectives, because the cutoff value strongly influences the overall study power.

Reference: *[1] Davies M, Færch L, Jeppesen OK, et al. Semaglutide 2·4 mg once a week in adults with overweight or obesity, and type 2 diabetes (STEP 2): a randomised, double-blind, double-dummy, placebo-controlled, phase 3 trial. Lancet. 2021;397(10278):971-984.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-29

Non-inferiority trials with evidence of assay sensitivity considering effect modification

Satomi Okamura[1], Eisuke Hida[2]
1 Department of Medical Innovation, Osaka University Hospital, Suita-shi, Japan
2 Graduate School of Medicine, Osaka University, Suita-shi, Japan

Introduction and Objective: Selection of non-inferiority (NI) margins and assay sensitivity assurance are well-known issues in 2-arm NI trials. Three-arm NI trials that include placebo, control, and treatment are strongly recommended to assess assay sensitivity. However, concerns about the ethics and feasibility of including placebo have prevented the practical application of 3-arm NI trials. Therefore, new methods are needed to quantitatively evaluate the assay sensitivity of NI trials.

We propose one approach to confirm assay sensitivity in a 2-arm NI trial using a historical trial with aggregate data. Conventionally, the population adjustment method is often used as an indirect comparison between trials while considering the difference of effect modification. However, this method needs the "shared effect modifier assumption" for setting the NI trial population as the target population (Phillippo, 2018), which would be unavailable when the NI trial is treatment vs control and the historical trial is control vs placebo. For this reason, we considered another approach that can account for the presence of effect modification.

Method and Results: To assess assay sensitivity considering effect modification, we specified a certain set value as an interaction parameter between the medication and the modifier and a range of the parameter. This is because the historical trial does not necessarily have the information of effect modification. The value set as the interaction parameter was determined with reference to feasibility, and after the value's standardization, the range was made based on the standard deviation. Following this setting, we would estimate the effect of control on placebo in the NI trial population, which admitted of confirming assay sensitivity that the effect size of control is more than that of placebo plus margin in the target population.

The performance of this approach was evaluated through simulations under various types of situations, such as strength and distribution of effect modification, sample size, and correlation between the modifier and other covariates. We also performed a tipping point analysis to find how extreme the range had to be to overturn the significance of the test. As the performance measures, joint power and type I error rate were investigated.

Conclusions: Although the level of evidence for the proposed method may be lower than that for a 3-arm NI trial, the results of various simulations suggest that the performance of the proposed method is useful as one of the more feasible methods to assess assay sensitivity of the 2-arm NI trial.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-30

Pseudo-value regression for the design of non-inferiority studies in Paediatric Oncology

Ulrike Poetschger[1], Evgenia Glogova[1], Helga Arnardottir[1], Paulina Kurzmann[1], Martina Mittlböck[2]
1 St. Anna Kinderkrebsforschung - CCRI, Vienna, Austria
2 Medical University of Vienna-Center for Medical Data Science, Vienna, Austria

Background/Introduction: Regression-models based on pseudo-values (PVR) are especially suitable when the outcome at specific time-points is of interest. E.g. in pediatric oncology the main medical interest is in cures rates, that is the outcome at a long-term time point. Motivated by an international randomized non-inferiority trial in childhood leukemia, statistical design considerations for non-inferiority studies are investigated with a special focus on PVR.

Methods: Non-inferiority trials aim to show that the effect of an experimental treatment is not inferior to the standard therapy by more than a pre-specified margin. Most commonly hazard ratios from a Cox-regression model or differences of Kaplan-Meier estimates at a given time point (KME) are evaluated by comparing the limit of a confidence interval to the specified non-inferiority margin.

Similar decisions can be performed with estimates from a PVR. Depending on the link-function either the survival difference, survival ratios or cumulative hazard ratios can be estimated, and their confidence intervals are then compared to the corresponding non-inferiority margin.

Both, KME and PVR can directly address long-term survival without relying on proportional hazards. However, ignoring important covariates for the statistical evaluation makes the results of KME sensitive to random imbalances. Thus, covariate adjustments are recommended to prevent underestimated group differences and a type 1 error inflation in non-inferiority studies. Thus, regression models like Cox-regression and PVR are preferable. However, proportional hazards assumption must hold for Cox-regression but not for PVR.

Clinical trial simulation is used to assess sample size requirements, compare the performance of different approaches as well as potential options for early stopping. The simulation setup considers departures from proportional hazard as well as covariate–outcome relationships of different extent.

Results: The clinical trial simulation helps to define the study design (sample size, stopping rules) and assesses the performance of several statistical methods. Here, PVR is chosen as primary analysis as it allows direct modelling of survival probabilities with covariate adjustment independent from proportional hazards. Differences in sample size requirements are negligible between PVR and Cox-regression, when the proportional hazards assumption holds. Furthermore, PVR also allows to assess non-inferiority in terms of survival probabilities, which may be more intuitive to physicians than hazard ratios.

Conclusion: For non-inferiority studies in pediatric oncology, PVR is a useful tool that allows adjusting for covariates without relying on proportional hazards. PVR should more often be used for non-inferiority trials in pediatric oncology that have a special interest in long-term outcome.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-31 Optimising oncology clinical trial strategies: Comparative analysis of single versus dual endpoint designs

Valeria Mazzanti, Boaz Adler, Pantelis Vlachos
Cytel Inc., Cambridge, United States

Background/Introduction: Adaptive clinical trial designs have become increasingly complex, driven by methodological advancements and the increased demands for robust clinical evidence. An interesting example is the exploration of dual endpoint assessments in oncology studies, instead of relying on a single confirmatory endpoint. This strategy is particularly relevant due to the prolonged observation periods associated with pivotal endpoints like Overall Survival (OS).

Methods: By including a surrogate endpoint such as Progression-Free Survival (PFS), and therefore allowing for earlier data collection, we can provide insights into the anticipated benefits of the drug under development. This study conducts a comparative analysis between single and dual endpoint designs to evaluate their respective efficiencies in terms of trial costs and duration, while also examining the inherent trade-offs. Additionally, we explore different approaches for assigning statistical significance to one or both endpoints, enhancing our understanding of our comparison.

Results: Clinical trials that examined OS with an early analysis on PFS showed lower average study durations than those only examining OS, and consequently lower average costs. Among the dual-endpoint options, the optimization exercise also found that assigning statistical significance to both endpoints improved the overall probability of success of the trial, making this specific adaptive design the more compelling choice for their final protocol.

Conclusion: Incorporation of co-primary or surrogate endpoints in Oncology clinical trials offers a strategic tool for both timely data collection, and potential savings in overall trial duration and average sample size.

P-A01-32 The use of two-sided tolerance interval testing with considering the variability of batches in the assessment of biosimilarity

Chieh Chiang[1], Siao-Ping Hong[1], Chin-Fu Hsiao[2]
1Tamkang University, New Taipei City, Taiwan
2 National Health Research Institutes, Zhunan, Taiwan

Introduction: In recent years, due to the expiration of several innovative drug patents, the corresponding biosimilar products are prepared to be manufactured. However, the statistical rule for evaluating the biosimilarity between an innovative biologic and its biosimilar product is not clear yet. Moreover, it is known that the between-batches variability plays a crucial role for the response of a biosimilar product, however, the involvement in the statistical analysis is seldom discussed in the literatures.

Methods: In this study, we assume that the therapeutic response can be explained by a nested random effect model with the between-batches variability. A two-sided tolerance interval-based hypothesis test is constructed with corresponding sample size determination. The statistical properties are investigated by simulation studies with various parameter components.

Results: Our simulation verifies that the coverage probability and power under various combinations of parameter components are satisfied, especially when the batch size is large enough.

Conclusion: Under the nested random effect model assumption, there are infinite solutions of the batch size and the sample size to reach a desired level of power. Two strategies are suggested for the optimization of the sizes – minimize the total sample and minimize the total cost. If the batch size is restricted, the limits of parameter components to achieve a desired level of power can still be investigated via the proposed method.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-33

Estimating treatment effect in randomised controlled trials with continuous outcomes subject to non-compliance via a CACE framework: A logistic regression based multiple imputation approach

Ruochen Du, Jialiang Li, Junxian Zhu, Mark Chan, Bee Choo Tai
National University of Singapore, Singapore

Background: Non-compliance in randomised controlled trials (RCT) poses a significant challenge to accurately estimate treatment effects, potentially leading to violation of the randomization assumption and biased estimates of the treatment effect. Traditional methods of analysis, such as intention-to-treat (ITT) and per-protocol (PP) analyses, are known to have limitations. The ITT analysis tends to underestimate the effect of treatment, while PP is prone to selection bias. The Complier Average Causal Effect (CACE) framework has been proposed to address non-compliance with several existing methodologies.

Methods: In this paper, we propose a novel method that assumes the CACE framework based on logistic regression using multiple imputation technique (LMI) to estimate the unknown compliance status in control group. Specifically, for RCTs with one-sided non-compliance issues where the compliance statues are observed in the treatment group while they are unknown in the control, a logistic regression was firstly fitted in the subset of treatment group only to obtain the parameters of coefficients that can be used to predict the compliance status, based on which new parameters were simulated for the subset of control group using multiple imputation technique. With the observed compliance status in the treatment group and predicted status in the control, treatment effect was estimated using the subset of the compliant subjects. Its performance is compared with ITT, PP, two other methods under the CACE framework, namely Instrumental Variable (IV) method and Latent Class Regression (LCR) method. The performance of these five methods were evaluated via intensive simulations under different scenarios of varying compliance rates, sample sizes, and effect sizes. Scenarios that account for selection bias are also considered with special interest. The methods were also compared using data from the JOBS II randomised clinical trial on depression, where the non-compliance rate was 55%.

Results: The simulation results demonstrate that our proposed method has smaller bias and mean squared error, with wider coverage and larger power, with robust and consistent performance even in the presence of significant selection bias. The treatment effect sizes estimated by these methods from the real data set align with our simulation results.

Conclusion: We have demonstrated the potential of LMI, in estimating the treatment effect of RCTs with continuous outcomes in the presence of non-compliance, it may be a useful alternative for researchers in the analysis of RCTs with non-compliance.Conclusion: Incorporation of co-primary or surrogate endpoints in Oncology clinical trials offers a strategic tool for both timely data collection, and potential savings in overall trial duration and average sample size.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-34

Extension the outcome of two stage design to survival time

Mizuna Itagaki[1], Ayano Takeuchi[1], Isao Yokota[2]
1 Chuo University, Bunkyo-city, Japan
2 Hokkaido University, Sapporo, Japan

Background: The "two-stage design" proposed by Simon (1989) made it possible to evaluate response midway through the trial and make a decision to stop the trial early. In particular, for fatal diseases such as malignant tumors, early discontinuation is desirable if no effect is observed (Green, 2003, etc.), and two-stage designs are often used in single-group trials without a concurrent control group. However, the two-stage design is a frequentist approach based on hypothesis testing and estimation, and when unexpected situations occur, it is difficult to respond flexibly, such as revising the design of the number of samples (Teramukai, 2008). On the other hand, the Bayesian approach predicts probabilities in advance and updates the information, so it is flexible and efficient, and there are many proposals for Bayesian-based methods (Teramukai, 2008, etc.). Although the Bayesian approach is computationally complex and difficult, its application is gradually becoming more widespread. The design proposed by Teramukai (2008) is intended to be applied to binary outcomes such as "effective" or "ineffective" for each patient group in the statistical analysis of single-arm clinical trials, so data on treatment failures such as these will be excluded as unevaluable (Ohashi and Hamada, 1995), and there is a risk of overestimating the response rate.

In this research, we will propose a Bayesian two-stage design that extends the end point of the design proposed by Teramukai to survival time. By using progression-free survival as an end point, it is possible to quickly determine whether the drug is ineffective or discontinued, and it is possible to evaluate with a high level of safety.

Methods: For the Beta distribution of the Product-Beta distribution, we estimated the arguments of the Beta distribution using the method of moments (the evaluation criterion is Kullback-Leibler divergence), construct a 90% and 95% confidence intervals, and evaluate the accuracy. Based on the survival curve, a simulation was performed in which the censoring rate was varied from 5% to 50%. We used SAS9.4 and R for simulation.

Results/Discussion/Future prospects/Conclusion: A discussion will be made based on the simulation results. We will confirm whether the accuracy remains the same even if the censoring pattern is changed, and conduct research with an eye toward practical application. Our extension allows the two-stage design to be applied to trials with survival time as an primary endpoint.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-35

Covariate and block adjustment in modelling treatment effects. Some useful theory to guide simulation

Stephen Senn[1,2], Franz Koenig[1], Martin Posch[1]
1 Medical University of Vienna, Vienna, Austria
2 University of Sheffield, Sheffield, United Kingdom

Background/Introduction: Popular approaches to examining properties of clinical trial designs and analyses are to simulate from one of three generating mechanisms

1) Sampling from a hyper-population (for example a multivariate Normal, MVN)

2) Randomly permuting treatment labels for an illustrative data-set

3) Resampling with replacement (bootstrapping) an illustrative data-set.

However, in doing this a useful guide, that is often overlooked, is to make use of the theory available.

Methods: There are three aspects of adjustment in randomised clinical trials under the linear model [1]. First, the expected variance penalty due to loss of orthogonality, second the expected reduction in the mean square error and third the effect of loss of degrees of freedom.

We demonstrate the application of the theory using three forms of resampling on a clinical trial fitting 32 possible models corresponding to different combinations of five possible predictors, using the ADEMP framework of Morris et al [2]. We also illustrate different choices of link functions for a stepped wedge design with a binary outcome.

Results: The expected inflation factor, λ, given covariates having a MVN distribution can be shown to be

$$E[\lambda] = 1 + \frac{k}{n-k-3}$$

where n is the number of subjects and k is the rank of the covariates fitted. Given the rank, further aspects of the correlation structure are irrelevant. For the example chosen, the theory using the MVN predicts extremely well what the effect on efficiency is of adjusting not only for sampling from the MVN but also for random permutation and bootstrapping. The key aspect of the stepped wedge design is shown to be the factors adjusted for and not the link function.

Conclusion: Of course, there is no guarantee that the examples are typical. Nevertheless they are sufficient to highlight two underappreciated functions of theory. First, it enables one to highlight cases where investigations using simulation show departures from standard theory, thereby providing added value in terms of insight and second it provides a way that simulations carried out by different authors may be linked to each other, easing comparison of competing claims.

References: *[1] S. Siegfried, S. Senn and T. Hothorn (2023) On the relevance of prognostic information for clinical trials: A theoretical quantification. Biom J. [2] T. P. Morris, I. R. White and M. J. Crowther (2019) Using simulation studies to evaluate statistical methods. Stat Med, 2074-2102.*

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-36 Should we use correlation between clinical trial outcomes as we do for a single trial outcome with repeated measures?

Dominic Stringer[1,2], Ben Carter[1,2,3], Richard Emsley[1,2,3]
1 Department of Biostatistics and Health Informatics, King's College London, United Kingdom
2 King's Clinical Trials Unit, King's College London, United Kingdom
3 NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, United Kingdom

Introduction and Objectives: Randomised controlled trials, particularly in mental health, often collect and report a large number of outcomes. Outcomes which are correlated, and measure similar underlying constructs, such as subscales of a psychometric measure, are usually analysed individually. This would seem inefficient.

Repeated measures of the same outcome over time are analysed in a single model, taking advantage of the within-person correlation that may increase precision, particularly in the presence of missing data. Such methods are unbiased on average under a missing at random assumption. This concept could be extrapolated further for treatment effect estimates of the primary outcome in the presence of correlated outcomes, using multivariate modelling techniques.

The objective was to develop and evaluate such methods, extending previous work (e.g. Vickerstaff, Omar & Ambler 2019) to include both multiple outcomes and repeated measures.

Methods and Results: We proposed structural equation models which extend the MMRM (mixed model for repeated measures) to multiple continuous outcomes at multiple timepoints, with outcomes within and across timepoints assumed to be correlated, modelled using common latent factors. Separate treatment effects were estimated for each outcome at each timepoint.

We carried out a simulation study following the ADEMP framework, comparing these methods to the univariate MMRM. Data was simulated representing a small (n=300) Phase III 2-arm parallel trial with a continuous primary outcome and two correlated secondary continuous outcomes. Data was generated using a multivariate normal distribution with varying correlation structures. Other factors varied were number of timepoints, amount of missing data and missing data mechanism. The estimands were the mean differences between arms for each outcome at each timepoint. The key performance measures were the model-based standard error, convergence, bias, empirical standard error, power and coverage. We apply the method to data from a trial of 362 people with psychosis (SLOW-MO trial).

Simulations showed the methods to be unbiased and reduce model based standard error/increase power particularly with highly correlated outcomes and missing data. However, the method does not always converge and there was a slight reduction in coverage based on a correspondingly smaller reduction in empirical standard error.

Conclusions: We have proposed new methods that can increase power for the analysis of correlated trial outcomes with differential missingness. This could allow trials to have differential missingness by design in outcomes where assessments are burdensome. More work is to extend to mixed forms of outcomes.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-37

Selection of time to measure the RMST: A simulation study

Foteini Tsotra, Dimitris Karlis
Athens University of Economics and Business, Athens, Greece

Background/Introduction: The hazard ratio (HR) is a standard metric in clinical research, typically derived from Cox proportional hazards (PH) models to assess treatment efficacy against standard care. However, its reliance on the PH assumption can limit its utility in scenarios where this assumption is violated. In such cases, the Restricted Mean Survival Time (RMST) presents a robust alternative, representing the average event-free survival time up to a predefined time, □. While widely utilized in oncology trials, RMST faces challenges in selecting the optimal □ due to limited guidance.

Methods: Our study employs a systematic review to investigate RMST properties and applications. Firstly, through a targeted literature review, we illustrate the mathematical properties of existing RMST estimators and explore methodologies for □ selection. Additionally, we provide an overview of RMST applications in clinical trials, illustrating crucial statistical properties such as sample size calculation. Finally, simulations are conducted to evaluate RMST estimator performance across various time points.

Results: Through our review, we expect to identify the key mathematical properties and characteristics of existing RMST estimators and illustrate the strengths and limitations of different estimation techniques. Furthermore, by synthesizing existing approaches and recommendations, we aim to offer guidance on the selection of □ in clinical research settings. This will enhance the interpretability and reliability of RMST estimates in evaluating treatment outcomes. Additionally, our overview of RMST applications in clinical trials will contribute to a better understanding of the practical implications of RMST in clinical research. Finally, the results of our simulations will provide valuable insights into the performance of RMST estimators under varying conditions. By assessing the robustness and accuracy of these estimation techniques across different time points, we aim to identify optimal approaches for RMST estimation in clinical practice.

Conclusion: Our study highlights the use of RMST as a measure in clinical research, particularly when the PH assumption is violated. In this study, we illustrate RMST's mathematical properties, discuss methodologies for selection, and highlight its practical applications. By offering guidance on □ selection and assessing RMST estimator performance, we enhance the reliability and interpretability of treatment outcome evaluations.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-38

A review of methods for optimal utility-based design of oncology clinical development programmes

Haotian Wang[1], Peter Kimani[1], Michael Grayling[2], Josephine Khan[2], Nigel Stallard[1]
1 Warwick Clinical Trials Unit, Coventry, United Kingdom
2 Johnson & Johnson Innovative Medicine, High Wycombe, United Kingdom

Optimal utility-based design is an innovative approach, employing probabilistic dynamic programming to refine and adapt clinical development programmes in response to evolving trial data. Our work provides a comprehensive review of optimal utility-based design approaches applied to oncology clinical trials. Our review scrutinizes available methods within this domain, delineating their principal components and operational mechanics. In the pursuit of elicient oncology clinical development programmes, stakeholders—including physicians, scientists, operations specialists, and regulators—play pivotal roles in shaping the trajectory of clinical trials. The alignment of these diverse perspectives with patient-centric outcomes is critical for their unmet needs and preferences, amidst the complex landscape of oncology treatment risks and benefits. The ability of proposed methods to enable a shift towards patient-centred trial outcomes is also reviewed. By assessing if the current designs reflect aspects of the drug development process such as changing endpoints between phase II and phase III, optimisation on trial speed instead of patient or event numbers and patient-centred trial outcomes, we propose some key inputs to include in utility-based design models and potential statistical and decision theory techniques to develop new methods.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-39

Enhancing endpoint analysis in inflammatory bowel disease trials: A comparative study of different modelling strategies

Xinyue (Sylvia) Zhang, Svetlana Cherlin, James Wason
Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

Introduction: Inflammatory Bowel Disease (IBD) encompasses Crohn's Disease (CD) and Ulcerative Colitis (UC), chronic conditions with diverse clinical presentations and treatment challenges. Evaluation of disease severity and treatment response in IBD often relies on clinical assessment tools, such as the Mayo Score for UC and the Crohn's Disease Activity Index (CDAI) for CD. However, conventional binary models used in IBD research may oversimplify disease complexity.

Methods: To investigate the types of outcomes used in clinical trials for IBD, we conducted a review of clinical trial articles related to IBD published in 2022, identifying 72 studies using clinical remission, Mayo Score, or CDAI as primary endpoints.

Logistic, Poisson, and ordinal models were applied in this study, and models' performance was evaluated performance in simulated data. Poisson and ordinal models used the Partial Mayo Score as the endpoint, while logistic model used clinical remission. Simulation scenarios varied remission rates to assess operating characteristics of the models. Additionally, we conducted real data analysis utilizing phase 3 Ustekinumab trial data from YODA to reverify the simulation results.

Results: From our paper review, clinical remission is the most commonly used primary outcome in IBD trials (47.2%), followed by Mayo Score (34.7%) and CDAI (23.6%).

In the simulation study, the ordinal model exhibited superior performance compared to the Poisson and logistic models in capturing disease severity and treatment effects. While all three models showed improved power with increasing sample size, the ordinal model consistently demonstrated significantly higher power, particularly in smaller sample sizes. The real data analysis also provided additional evidence supporting the superiority of ordinal models in detecting treatment effects.

Conclusion: This study explores the application of logistic, Poisson and ordinal regression, and enhance endpoint analysis in IBD trials. The ordinal model offers a more informative approach to endpoint analysis in IBD trials, providing nuanced insights into disease dynamics and treatment outcomes. By acknowledging the hierarchical nature of disease severity, this model improves statistical power and has the potential to enhance decision-making in IBD management. However, we restricted our analysis to the Partial Mayo Score which excludes the endoscopic subscore, to facilitate simulation calculations in this study. Future research could extend this approach to include endoscopic scores and explore the applicability of ordinal models in CD trials.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-40

Integrating market access considerations in clinical trial design: Leveraging cloud-native software for holistic optimisation

Boaz Adler, Pantelis Vlachos, Valeria Mazzanti
Cytel Inc., Cambridge, United States

Background/Introduction: Cloud-native platforms allow product development teams to consider evidence thresholds from both the regulatory and Health Technology Assessment (HTA) and market access perspectives. This means RCTs can be optimized to meet potential challenges from either lens.

Methods: In this presentation, we will highlight how we've optimized a study design for a BCMA CAR-T treatment for Triple-Refractory Multiple Myeloma (TRMM) patients using two proprietary software offerings, one for endpoint and comparator selection, and the other for adaptive clinical trial design and optimization.

Results: The drug development team was able to select the most appropriate comparator and endpoints from a market access perspective, and further optimize the trial's design to reduce the time to regulatory submission.

Conclusion: From this experience, we are advocating that tackling Market Access considerations early in the design process ensures ultimate launch preparedness and that HTA considerations can and should inform the design process.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-41

U-PRO-CRM: Designing patient-centred dose-finding trials with patient-reported outcomes

Emily Alger[1], Shing Lee[2], Ying Kuen Cheung[2], Christina Yap[1]
1 Clinical Trial and Statistics Unit, Institute of Cancer Research, London, United Kingdom
2 Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, United States

Background: The determination of a Maximum Tolerated Dose (MTD) remains the main objective for the majority of dose-finding oncology trials. Whilst most dose-finding trials rely on clinicians to identify dose-limiting toxicities (DLTs) experienced by patients during the trial, research has suggested that clinicians may under-report patient's adverse events. Therefore, current practice may be vulnerable to recommending intolerable doses to patients for further investigation in subsequent trials. In light of the FDA's Project Optimus initiative, there is fresh interest in leveraging Patient-reported Outcome (PRO) data to supplement the assessment of tolerability for investigational therapies within early phase dose-finding oncology trials. However currently, only three trial designs incorporate PROs formally in dose-escalation decision making: PRO-CRM, TiTE-PRO-CRM and PRO-ISO.

Methods: We present Utility-PRO-Continual Reassessment Method (U-PRO-CRM), a novel trial design which simultaneously utilises both clinician-assessed DLTs (Clinician-DLTs) and patient-assessed DLTs (Patient-DLTs) to make dose escalation or de-escalation decisions and to recommend a MTD that optimises the clinician-patient DLT trade-off. By employing a utility curve, U-PRO-CRM provides greater flexibility to trade-off the rate of Patient-DLTs and Clinician-DLTs to find an optimal dose, whilst containing the published PRO-CRM as a special case. We present a simulation study evaluating the performance of the U-PRO-CRM design compared to the PRO-CRM and a benchmark.

Results: For specified trade-offs between Clinician-DLT and Patient-DLT rate, U-PRO-CRM outperforms the PRO-CRM design by identifying the true MTD more often. For instance, U-PRO-CRM is superior in scenarios where we look to compromise the optimal Patient-DLT rate given a dose's corresponding Clinician-DLT rate. In the special case where U-PRO-CRM generalises to PRO-CRM, U-PRO-CRM performs as well as its published counterpart.

Conclusion: Whilst PRO-CRM constrains dose-selection to maximising either Clinician-DLT or Patient-DLT rate, U-PRO-CRM has greater adaptability. By employing a utility-based dose selection approach, U-PRO-CRM offers the flexibility to define a trade-off between the risk of patient-assessed and clinician-assessed DLTs for an optimal dose. Patient-centric dose-finding strategies, which integrate PROs, are poised to assume an ever more pivotal role in significantly advancing our understanding of treatment tolerability. This bears significant implications in shaping the future landscape of early phase trials.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-42

Cultivating patient preferences in ALS clinical trials: Reliability and Prognostic Value of the Patient-Ranked Order of Function (PROOF)

Ying Lu[1], Ruben van Eijk[2], Leonard van den Berg[2]
1 Stanford University, Stanford, United States
2 UMC Utrecht, Utrecht, Netherlands

Background/Introduction: Integration of individual patient preference with objective outcome measures in clinical trials is an important step for patient-centric evaluation of treatment efficacy. We have used desirability of outcome ranking (DOOR) approach to construct the Patient-Ranked Order of Function (PROOF) for Amyotrophic Lateral Sclerosis (ALS) trials. In this study, we assess the reliability and prognostic value of different sets of patient-reported preferences which can be used for the PROOF endpoint.

Methods: Data were obtained through online surveys over the course of 12 months using the population-based registry of the Netherlands. Patients were asked to score functional domains of the ALS functional rating scale (ALSFRS-R) and rank the order of importance of each domain. Two weeks after the initial invite, the questionnaire was repeated to evaluate test-rest reliability. Vital status was extracted from the municipal population register.

Results: In total, 611 patients with ALS were followed up for survival and 382 were included in the test-retest reliability study. All versions of PROOF, utilizing different sets of preferences, resulted in excellent reliability (ICCs ranged from 0.89 [95% CI 0.87-0.91] to 0.97 [95% CI 0.97-0.98], all $p < 0.001$), without systematic differences between baseline and week 2 (mean rank difference range -1 to -3 [95% CI range -8 to 2], all $p > 0.20$). Preferences about future events were more variable compared to preferences about current symptoms. All versions of PROOF strongly predicted overall survival (hazard ratio per 100 ranks ranged from 0.69 to 0.72 [95% CI range 0.64 to 0.79], all $p < 0.001$) and had a more evenly separation of survival curves between rank-stratified subgroups compared to the ALSFRS-R total score.

Discussion: In a large cohort of patients, we show how patient-reported preferences can be measured and integrated reliably with the ALSFRS-R without leading to systematic bias. Patient preferences may provide unique prognostic information in addition to what is already measured conventionally. This could provide a more comprehensive understanding of how medical interventions effectively address the patient's concerns and improve what matters most to them.

POSTER SESSION A-01:
Clinical Trials Design and Simulations

P-A01-43

Intervention to improve supporting skills of registered dietitians in lifestyle modification: A feasibility study

Misa Adachi[1,2], Asuka Suzuki[1], Kazue Yamaoka[1,3], Mariko Watanabe[4,5], Toshiro Tango[1,6]
1 Teikyo University Graduate School of Public Health, Tokyo, Japan
2 Nutrition Support Network LLC, Kanagawa, Japan
3 Tetsuyu Crinical Research Center, Tokyo, Japan
4 Preffect Ural University of Kumamoto, Kumamoto, Japan
5 Showa Women's University, Tokyo, Japan
6 Center for Medical Statistics, Tokyo, Japan

Introduction and Objective: The indispensable role of registered dietitians in facilitating enhancements in lifestyle habits necessitates the acquisition of advanced competencies to scientifically intervene in nutrition and adeptly encourage behaviour modification. To address this imperative, we introduced a novel 'Skill Check Sheet' for registered dietitians and devised a methodology to evaluate the effectiveness of interventions targeting self-improvement capacities. The primary aim of this investigation was to assess the feasibility of a skill enhancement intervention program employing the Skill Check Sheet to optimize lifestyle modifications among novice registered dietitians.

Method and Results: The research methodology entailed a single-arm, three-month intervention study. Participants comprised registered dietitians (25 individuals) who met specific inclusion and exclusion criteria, possessed less than five years of experience, and conducted self-assessments using the Skill Check Sheet. The overall score for nutrition assessment was computed as the aggregate of the proportion of responses predominantly implemented across all checkpoints. Inter-rater reliability (ICC) was evaluated through objective assessments conducted by three expert evaluators (proficient registered dietitians) for each dietitian's competency. ICC was assessed using a mixed-effects model (SAS NLMIXED procedure). Concurrent validity (score difference between the registered dietitians and proficient registered dietitians) was evaluated by t-test, and reproducibility (scores at 0 week and 2 weeks) was evaluated by Pearson correlation coefficient. The ICC among the three evaluators was 0.52 (0.29 to 0.75). The remaining assessments indicated satisfactory results (both $p < 0.001$).

Conclusion: The feasibility of the program employing the Skill Check Sheet to optimize lifestyle modifications among novice registered dietitians was confirmed, and important basic information was obtained. The findings of this study augur well for elevating the calibers of nutritional interventions administered by registered dietitians, thereby fortifying the efficacy of nutritional strategies aimed at averting the onset and progression of lifestyle-related ailments. Trial registration number: UMIN000052442.

POSTER SESSION A-02:
Using Statistics to Improve the Conduct of Clinical Trials

P-A02-01 Design and analysis of co-enrolment trials

Abraham Richi Contreras
University of Warwick, Coventry, United Kingdom

Background: Research sites that run several studies at one time can come across participants that are eligible for more than one trial. Where compatible, participants can be given the opportunity to co-enrol, either concurrently or sequentially. Co-enrolment also allows multiple studies to reach recruitment targets faster especially where eligible participants are difficult to identify, e.g., time-sensitive settings, or rare incidence. This should maximise research throughput and reduce competition between trials. Implementing co-enrolment poses organisational and logistical hurdles. On the other hand, adequate preparations can drastically improve the efficiency of trial delivery and increase collaboration between healthcare researchers. The latest UK guidance document was a 2019 update within critical care, motivated by the growing occurrence of co-enrolment.

Challenges of CoReCCT: The Confederation of Respiratory Critical Care Trials (CoReCCT) is a co-enrolment model for four large trials (AWAKE PRONE, PROTECT Airways, NAVA, & RELEASE), that promote efficacy in the conduct of critical care trials. CoReCCT faces challenges regarding the thorough considerations to design and statistical aspects due to the anticipated high co-enrolment rates.

Design:
- Data collection and follow-up should follow a common schedule to minimise burden and homogenize patient experience
- Consent strategies must be carefully determined to minimise additional burden to patients and decision-makers as well as homogenize likelihood of co-enrolment
- CRFs and report documentation must be tailored to optimise data collection and trial logistics

Statistical:
- Sequential co-enrolment may imbalance the patient characteristics of a treatment arm of one trial whilst concurrent interventions between the co-enrolment studies may interact and affect study outcomes
- Time between treatments may be considered for potential crossover effects
- Factors of co-enrolment rates should be measured and analysed to ensure balance in likelihood of co-enrolment to specific subgroups

Methods

Design Solutions
- Combined CRFs and bespoke oversight report templates
- Substudy on consent strategies during pilot phases of CoReCCT; opportunity for surveys from research staff and PPI
- Streamline data collection and organisational processes, e.g., one PPI, one ethics committee, etc.

Statistical Analysis Methods
- Interaction testing; subgroups defined a priori
- Univariable analysis of co-enrolment factors, e.g.: co-enrolment details, patient health status and/or severity of illness, research site and staff co-enrolment experience, research site capacities.
- Additional variables collected: time/order of randomisation, time/order of intervention received, co-enrolment rate per trial arm.

We wish to provide an overview of the challenges faced and lessons learned from this novel strategy.

POSTER SESSION A-02:
Using Statistics to Improve the Conduct of Clinical Trials

P-A02-02

Applying the estimands framework in oncology dose escalation - A practical example of improving the MTD estimate in the first-in-human Phase 1 trial of antibody-drug-conjugate M9140

Christina Habermehl[1], Athanasios Pallis[2], Melanie Andreas[1], Marlene Schulte-Goebel[1], Anja Victor[1]
1 Merck Healthcare KGaA, Darmstadt, Germany
2 Merck Santé S.A.S., Lyon, France, an affiliate of Merck KGaA, Darmstadt, Germany

Introduction: Estimands are typically applied for efficacy endpoints in late phase clinical trials. However, there is still reluctance and uncertainty regarding their application in Phase I. This use case demonstrates the benefit of applying the estimands framework in Phase I studies with the goal to establish the maximum tolerated dose (MTD).

Methods: In the M9140 Phase I dose escalation trial (NCT05464030), a Bayesian 2-parameter logistic regression model (BLRM) was applied to guide dose escalation and to suggest an MTD at the end of the trial. The primary endpoint in this setting is the occurrence of dose-limiting toxicities (DLTs). Challenges encountered in this trial in estimating the MTD with the BLRM that will be discussed are:
1) M9140 is an antibody-drug-conjugate (ADC) with an exatecan payload - a chemotherapy with a side effect profile that includes myelosuppression. This side effect is commonly managed with blood transfusions or growth factors, after which the affected blood values increase. In this case, the administration of these countermeasures may prevent the observation of a DLT. If this is ignored in the BLRM, this may lead to an underestimation of the dose-DLT curve. To take this into account, the estimands framework was applied, where the administration of growth factors or blood transfusions during the DLT observation period (first treatment cycle), was treated as an intercurrent event (ICE). These intercurrent event are then considered as if being DLTs in the BLRM (composite strategy).

2) A few patients in the trial received a lower actual dose than the dose level they had been assigned due to a body weight-based dosing cap, which may also lead to underestimation of the dose-DLT curve if they are analyzed with their assigned dose level. As sensitivity analysis, these patients were analyzed in the dose level they actually received.

Results: The implementation of the estimands framework with ICEs and the different estimated dose-DLT curves and MTD suggestions from the BLRM will be presented for different settings, e.g., for different ICE handling strategies.

Conclusion: The estimands framework was applied throughout the trial, and the definition how to treat the intercurrent events helped guiding discussions how to consider these patients in the dose escalation decisions. The final MTD suggestion by the BLRM, taking into account ICEs and the actual dose level patients had received, was also in agreement with the treating investigator's assessment and supported the decision about the MTD by the safety monitoring committee (SMC).

POSTER SESSION A-03:
Infectious Diseases Modelling

P-A03-01 Evaluating pooled testing policies in community and healthcare settings for emerging epidemics

Bethany Heath[1], Stephanie Evans[2,3], David S. Robertson[1], Julie V. Robotham[2,3,4,5], Sofía S. Villar[1], Anne M. Presanis[1,6]

1 MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
2 HCAI, Fungal, AMR, AMU and Sepsis Division, UK Health Security Agency, London, United Kingdom
3 Statistics, Modelling and Economics Division, UK Health Security Agency, London, United Kingdom
4 NIHR Health Protection Research Unit in Modelling and Health Economics at Imperial College London in partnership with the UK Health Security Agency and London School of Hygiene and Tropical Medicine, London, United Kingdom
5 NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with the UK Health Security Agency, Oxford, United Kingdom
6 NIHR Health Protection Research Unit in Behavioural Science and Evaluation at University of Bristol in partnership with the UK Health Security Agency and MRC Biostatistics Unit, University of Cambridge, Bristol, United Kingdom

A coordinated testing policy is an essential tool for responding to emerging epidemics, as was seen with COVID-19. A crucial aspect to consider when implementing a testing strategy as an intervention to manage an emerging epidemic is the number of tests available. This is why pooled testing (a method that involves pooling samples taken from multiple individuals and analysing this with a single diagnostic test) has been suggested. However, pooled testing was rarely used during the COVID-19 pandemic. Therefore, this research aims to understand the impact of a pooled testing policy and the scenarios where it can be best used so it can be better utilised in a future epidemic. To assess pooled testing, a dynamic model compromised of a representative relationship network and an extended SEIR model is used in both the community se6ng of a small town and in a `typical' English hospital. Different pooled testing policies were evaluated in hospitals considering whether to only pool wards together or the whole hospital, whether it is better to use lateral flow devices or PCR tests for follow-up testing and considering the contexts of early in the pandemic and the transition back to `normal' life. Pooled testing in the community was evaluated considering the concerns of a fixed, rather than unlimited, testing capacity, different proportions of cases being symptomatic and individuals not complying with the testing policies. Pooled testing in the community setting was found to be superior in terms of a range of metrics to symptomatic individual testing (including total infections, length of epidemic and peak infections) for when more than a small proportion of the population (10%) do not comply with the testing procedure, while in the hospital setting pooled testing was found to have statistically significant effect on reducing healthcare worker infections. Whether a policymaker should consider pooled testing in a particular setting depends on a range of factors including viral dynamics, compliance levels and, very importantly, which metrics are of highest priority.

POSTER SESSION A-03:
Infectious Diseases Modelling

P-A03-02 Estimation of the effect of Rapid ART on HIV infection

Kazumi Omata[1], Mayumi Imahashi[2], Tatsuya Noda[3], Toshibumi Taniguchi[4]
1 National Center for Global Health and Medicine, Japan
2 Nagoya Medical Center, Nagoya, Japan
3 Nara Medical University, Kashihara, Japan
4 Chiba University, Chiba, Japan

Background: Antiretroviral therapy (ART) is expected to contribute not only to the treatment of HIV-positive people, but also to the reduction of new infections. Currently, the average time from diagnosis to the start of treatment for HIV infection in Japan is estimated to be about 40 days. The purpose of this study is to estimate the future effects of Rapid ART in reducing this duration.

Methods: Assuming HIV infection in MSM (men who have sex with men) communities, estimates were made using a compartmental model which employs five compartments: susceptible, infectious, diagnosed, AIDS, and treated. The input values were adopted from the number of new HIV infections reported by the AIDS Prevention Information Network in Japan, and the analysis of clinical data at Nagoya Medical Center and at the National Center for Global Health and Medicine in Japan.

Results: As a result of estimating the effect of Rapid ART assuming that the average time from diagnosis to the start of treatment for HIV infection is changed from 42 days to 1 day (start of treatment immediately after diagnosis) and this shortening is continued, it was found that the effect of reducing the number of infected people depends on the size of the infection, and the effect is large in high-prevalence countries. In Japan, the number of new infections per year has been changing around 1,000 for the last 25 years, and it is estimated that the effect of Rapid ART is small in such a low-prevalence country.

Conclusion: Rapid ART is effective in reducing new infections in countries and regions where the number of infections is large. In low-prevalence countries such as Japan, recommendations for testing for infection and expanded preventive measures such as PrEP, may be effective in curbing HIV transmission. The effect of Rapid ART is associated with the second and third 95 in the UNAIDS 95-95-95 goal for HIV infection [1], suggesting that it is effective to adjust the amount of effort to achieve this goal depending on whether the country is a high-prevalence or a low-prevalence country.

Reference: *[1] UNAIDS. Global AIDS Strategy 2021-2026 (2021). https://www.unaids.org/en/Global-AIDS-Strategy-2021-2026.*

POSTER SESSION A-03:
Infectious Diseases Modelling

P-A03-03

Evaluating forecasting models for COVID-19: Insights from the Omicron period

Nere Larrea[1,2,3], Dae-Jin Lee[3,4,5], Irantzu Barrio[3,5,6], Inmaculada Arostegui[3,5,6], Eduardo Millan[7], Jose Maria Quintana-Lopez[1,2,3]

1 Hospital Galdakao Usansolo, Usansolo, Spain
2 Biosistemak Institute for Health Systems Research, Barakaldo, Spain
3 Network for Research on Chronicity Primary Care and Health Promotion (RICAPPS), Usansolo, Spain
4 IE University School of Science and Technology, Madrid, Spain
5 BCAM Basque Center for Applied Mathematics, Bilbao, Spain
6 Department of Mathematics University of the Basque Country UPV/EHU, Leioa, Spain
7 Healthcare Services Sub-directorate, Osakidetza-Basque Health Service, Vitoria, Spain

Introduction and Objective: One of the objectives during the pandemic was to provide clinicians and healthcare managers with information on predictions about the evolution of the pandemic, for instance, the short-term forecast of the daily number of individuals infected with SARS-CoV-2, those admitted to the ward, and ICU admissions. The aim of this study was to evaluate three different forecasting modelling approaches used to provide short-term forecasts and evaluate their accuracy.

Methods: The period of study was between December 15, 2021 and January 25, 2022 (Omicron variant higher peak period) in the Basque Country, Spain. We considered penalized regression splines to evaluate daily counts of SARS-CoV-2 positive cases, hospitalizations and ICU admissions. In order to deal with overdispersion, we used the Negative Binomial distribution for the counts. A generalized additive model (GAM) was considered to account for patterns in the data; three types of penalties have been used for the first and second derivatives of this model. The errors of the predictions were evaluated through the RMSE (root median square error) and the relative error of the 2-day and 5-day forecast.

Results: The RMSE ranged between 3.31-6.61 and 3.61-8.70 for the 2- day and 5-day prediction respectively, with the relative error between 0.15-0.32 and 0.16-0.61 for ICU admissions. In the case of hospital admissions, these values increased to 22.37-25.61 and 23.32-30.22 for the RMSE and 0.24-0.29, 0.23-0.35 for the relative error. Finally, the largest errors were obtained in the prediction of SARS-CoV-2 positives, as the errors took values of 2523.45-4362.97 and 3090.71-6787.25 for the RMSE and 0.35-0.55 and 0.45-0.87 for the relative error, respectively. The greatest errors were found when there were sudden trend changes, both for increasing or decreasing trends, and mainly with the model with higher penalty.

Conclusions: Our study on short-term forecasting models during the peak of the Omicron period revealed varying accuracy across different COVID-19 outcomes. While the models performed reasonably well in predicting daily ICU admission counts, their efficacy decreased for hospital admissions and was notably challenged in forecasting positive cases. The highest errors were observed during sudden trend changes, emphasizing the need for more robust models, especially during dynamic shifts in pandemic. These findings underscore the importance of continuous refinement and adaptation of forecasting approaches to enhance their reliability in guiding healthcare responses during pandemics.

POSTER SESSION A-03:
Infectious Diseases Modelling

P-A03-04

An alternative measure of vaccine effect based on Aalen's additive survival model framework

Zdeněk Valenta[1], Julie Dudášová[2,3], Andreea Magalie[4], Jeffrey R. Sachs[4]
1 Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic
2 Quantitative Pharmacology and Pharmacometrics, MSD Czech Republic, Prague, Czech Republic
3 First Faculty of Medicine, Charles University, Prague, Czech Republic
4 Quantitative Pharmacology and Pharmacometrics, Merck & Co., Inc., Rahway, United States

Background: The protective effect of a vaccine is typically evaluated as vaccine efficacy in randomized controlled clinical trials. Efficacy is defined as a proportional reduction in risk of disease among vaccinated subjects relative to a placebo control. Previous work considered immune response biomarker values (immunogenicity) for assessing vaccine efficacy in the context of cross-sectional and time-to-event data, using logistic regression and Cox proportional hazards (PH) or Fine-Gray regression models, respectively.

Methods: In this novel approach, we employ flexible properties of Aalen's additive survival model framework and define an alternative measure of vaccine effect as "relative survival" (RS) of vaccinated versus control subjects. RS compares estimated probabilities of not contracting the disease in vaccinated and control population. Given time-dependent covariates and/or time-varying effects, this measure can vary with respect to time. It extends well to models involving interactions. In this metric, better survival (here meaning lower risk of disease) among vaccinated subjects leads to an RS value greater than 1.

Results: This work studies data on the effect of live attenuated tetravalent dengue vaccine (Dengue Tetravalent Vaccine, Live, CYD-TDV, Sanofi Pasteur, Inc.) on the incidence of DENV2 virologically confirmed dengue disease. We estimate the RS, its standard error and confidence limits based on Aalen's additive model. The best-fitting Aalen's additive model involves time-dependent DENV2-specific immunogenicity, baseline dengue serostatus, vaccination status, and an interaction between baseline dengue serostatus and vaccination status; in this model the effects of the two categorical predictors and their interaction change over time, while the immunogenicity effect is constant with respect to time. When fitting the main effects model only, the effect of vaccination status and dengue serostatus were insignificant while the protective effect appeared to be mediated solely by time-dependent immunogenicity values. When employing the Cox PH model in estimating time-invariant (i.e., averaged) main effects of dengue serostatus, vaccination status and time-dependent immunogenicity, addition of an interaction term resulted in convergence failure.

Conclusions: This work extends the time-to-event data analysis approach to situations where either the assumption of proportional hazards is violated, or the estimation based on the Cox PH multiplicative survival framework fails. Understanding the time-varying effects of vaccination status, immunogenicity and baseline covariates can elucidate persistence (durability) of vaccine-induced protection and its potential heterogeneity. This knowledge is essential to decisions on vaccine recommendations, design of immunization schedules and revaccination strategies.

POSTER SESSION A-03:
Infectious Diseases Modelling

P-A03-05

Redefining the SEIR model for direct modulation of hospital load independent of number of susceptible, exposed, infected, and recovered

Jan Brink Valentin[1], **Simon Grøntved**[1,2]
1 Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
2 Psychiatry, Aalborg University Hospital, Aalborg, Denmark

Background: Hospital load planning have become a major challenge in most western countries with an aging population, increased psychiatric hospital activity, and elevated risk of virus and bacterial infections as a consequence of globalisation and antibiotic resistance. Statistical modelling that can aid future hospital planning or provide insight into the mechanisms for increased hospital activity is warranted. However, investigating the mechanisms for the spread of infectious diseases requires accurate estimation of the basic reproduction number (BRN), which can be provided by theoretical models such as the SIR/SEIR models. The benefit of these models is the ability to both predict and explain. However, the models need to be able to predict the disease spread with high accuracy for trustworthy estimation of the BRN. Unfortunately, the SIR/SEIR have shown to be highly sensitive to even small deviations in input data, leading to inaccurate predictions. The lack of robust training algorithms is therefore an obstacle that needs to be overcome. A further benefit of the SIR/SEIR models is the low number of predictors, which may be attractive for hospital planning in developing countries.
In this study we intended to redefine the SEIR model for prediction of hospital activity following infectious diseases without any other knowledge than prior hospital activity. The primary aim was to optimize the filtering algorithm of the input timeseries to reduce systematic prediction error in terms of time to peak activity for three weeks prognoses.

Methods: The model was trained on Danish national data and included COVID-19 positive in-patients from 2020 to 2021. The algorithm was tested on data from early 2022. Model performance was assessed using Bland-Altman plots.

Results: In the observation period the hospital activity peaked at 1762 in-patients. Our preliminary results revealed a difference of 1 day when predicting "day of peak activity", and the limits of agreement for hospital activity was -196 to 333 with a bias of 69 in-patients.

Conclusion: For three weeks prognoses, our results indicate that our model is well calibrated. Moreover, the model can easily be extended with data on daily number of vaccines given, and advanced statistical prediction models for estimation of the BRN without compromising the explainable value of the theoretical models for spread of infectious diseases.

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-01

Use of a nonparametric Bayesian method to model health state preferences: An application to Lebanese SF-6D valuations

Samer Kharroubi[1,2]
1 American University of Beirut, Beirut, Lebanon
2 The University of Sheffield, Sheffield, United Kingdom

Introduction: Typically, models that were used for health state valuation data have been parametric. Recently, many researchers have explored the use of non-parametric Bayesian methods in this field. In the present paper we report on the results from using a nonparametric model to predict a Bayesian SF-6D health state valuation algorithm along with estimating the effect of the individual characteristics on health state valuations.

Methods: A sample of 249 states defined by the SF-6D have been valued by a representative sample of 577 members of the Lebanese general population, using standard gamble. Results from applying the nonparametric model were reported and compared to the original model estimated using a conventional parametric random effects model. The covariates' effect on health state valuations was also reported.

Results: The nonparametric Bayesian model was found to perform better than the parametric model 1) at predicting health state values within the full estimation data and in an out-of-sample validation in terms of mean predictions, root mean squared error and the patterns of standardized residuals, and 2) at allowing for the covariates' effect to vary by health state. The findings also suggest an important age effect with sex, having some effect, but the remaining covariates having no discernible effect.

Conclusions: The nonparametric Bayesian model is a powerful technique for analyzing health state valuation data and is argued to be theoretically more flexible and produces better utility predictions from the SF-6D than previously used classical parametric model. In addition, the Bayesian model is more appropriate to account the covariates' effect. Further research is encouraged.

Abstract Summary: Typically, models that were used for health state valuation data have been parametric. Recently, many researchers have explored the use of non-parametric Bayesian methods in this field. In the present paper we report on the results from using a nonparametric model to predict a Bayesian SF-6D health state valuation algorithm along with estimating the effect of the individual characteristics on health state valuations.

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-02

Bayesian modelling of health state preferences: Can existing preference data be used to generate better estimates?

Samer Kharroubi[1,2]
1 American University of Beirut, Beirut, Lebanon
2 The University of Sheffield, Sheffield, United Kingdom

Introduction: Valuations of preference-based measure such as EQ-5D or SF6D have been conducted in different countries because social and cultural differences are likely to lead to systematically different valuations. However, there is a scope to make use of the results in one country as informative priors to help with the analysis of a study in another, for this to enable better estimation to be obtained in the new country than analyzing its data separately. This is explored using a case study modelling UK data alongside Lebanon data to generate Lebanon estimates.

Methods: Data from two SF-6D valuation studies were analyzed where, using similar standard gamble protocols, values for 249 common health states were devised from representative samples of the Lebanon and UK general adult populations, respectively. A nonparametric Bayesian model was applied to estimate a Lebanon value set, where the UK results were used as informative priors. Generated estimates were compared to a Lebanon value set estimated using Lebanon values alone using different prediction criterion, including predicted versus actual mean health states valuations, mean predicted error, root mean square error and out of sample leave one out prediction.

Results: The proposed method of modelling utility functions permitted the UK valuations to contribute significant prior information to the Lebanon analysis. The results suggest that using Lebanon data alongside the existing UK data produces Lebanon utility estimates better than would have been possible using the Lebanon study data by itself.

Conclusions: The promising results suggest that the existing preference data could be combined with data from a valuation study in a new country to generate preference weights, thus making own country value sets more achievable for low–middle income countries. Further research and application to other countries and preference-based measures are encouraged.

Abstract Summary: Valuations of preference-based measure such as EQ-5D and/or SF6D have been conducted in different countries. There is potential to borrow strength from existing countries' valuations to generate better representative utility estimates. This is explored using a case study modelling UK data alongside Lebanon data to generate Lebanon estimates.

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-03 Validation of EQ-5D in patients with heart failure with preserved ejection fraction

Kazutaka Nishio, Tomomi Yamada, Yasushi Sakata, Eisuke Hida
Osaka University, Suita, Japan

Introduction and Objective: In cost-effectiveness analysis in the medical field, Quality Adjusted Life Years (QALYs) are often used as measure of effectiveness. QALYs are defined as the product of Quality of Life (QOL) and years of survival. The EQ-5D is commonly used as a measure of QOL to calculate QALYs. The EQ-5D is a scale consisting of 5 items developed by the EuroQol group, and its reliability, validity and responsiveness have been validated in in various disease areas. However, to date, there have been no studies conducted regarding the measurement of population norms of EQ-5D for the Japanese version in cardiology. Therefore, this study aims to construct a model to calculate the population norms of EQ-5D for patients in cardiology, particularly, we focus on patients with heart failure with preserved ejection fraction (HFpEF). Additionally, this study aims to investigate the relationship between patient background:or disease and symptoms and EQ-5D.

Method: This study used PURSUIT HFpEF registry which includ s 1024 patients with HFpEF admitted to Osaka University Hospital and its affiliated hospitals since 2016. Data collected included basic information, such as gender, age and cohabitation status, clinical information, such as 6 minute walk distance and Blood test results and medication information, and questionnaire information, such as EQ 5D 5L, Pittsburgh Sleep Quality Index (PSQI) and Self rating Depression Scale (SDS). Data collection points were at admission, discharge and one year after discharge. We examined the association between clinical information and questionnaire data, such as factors contributing to EQ 5D, for constructing the model representing population norms. Furthermore, to validate the constructed model, we investigated the longitudinal changes in EQ investigated the longitudinal changes in EQ--5d between admission and discharge, as well as 5d between admission and discharge, as well as discharge and one year after discharge.

Results: While there are limitations imposed by the available data, we are able to identify several significant factors among the clinical information and questionnaire data that influence EQ-5D. Furthermore, it was confirmed that it is possible to calculate Quality-Adjusted Life Adjusted Life Years (QALY) using a simpler method based on the relationship between the longitudinal Years (QALY) using a simpler method based on the relationship between the longitudinal trends in QOL values and symptoms.

Conclusion: This study derived population norms of EQ-5D and longitudinal trends for Japanese 5D and longitudinal trends for Japanese patients with HFpEF, which is expected to facilitate future implementation of economic evaluations in the disease.

Reference: *[1] Feng YS, Kohlmann T, Janssen MF, Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021 Mar; 30(3): 647-673.*

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-04

Impact of limited sample size and follow-up on partitioned survival or multistate modelling-based health economic models: A simulation study

Jaclyn Becca[1], David Naimark[2], Petros Pechlivanoglou[1]
1 The Hospital for Sick Children, Toronto, Canada
2 Sunnybrook Hospital, Toronto, Canada

Introduction: Health economic models often rely on extrapolation of clinical time-to-event data for multiple events. Two potential modelling approaches to model such extrapolations in oncology include partitioned survival models (PSM) and multistate models (MSM). The objective of this simulation study was to assess the performance of PSM and MSM across datasets with varying sample sizes and degrees of censoring.

Methods: We simulated trajectories of disease progression and death for multiple hypothetical populations with advanced cancers across different scenarios of progression and death intensities. These populations served as the sampling pool for simulated trial cohorts with multiple sample sizes and various levels of follow-up. We fit parametric survival models on transition-specific data (for MSM) or overall survival and progression-free survival data (for PSM) on these simulated cohorts We conduct scenarios for both MSM and PSM with different approaches to incorporating general population mortality (GPM). Mean survival was compared to population values using root mean squared error.

Results: With robust follow-up, both PSM and MSM accurately estimated mean population survival, while smaller samples and shorter follow-up were associated with larger error across approaches and clinical scenarios, especially for more distant clinical endpoints. MSM was slightly more often not estimable when informed by studies with small sample sizes or follow-up due to a low event rate for downstream transitions. However, when estimable, MSM frequently produced smaller error in mean survival with limited data compared to PSM approaches. Incorporating GPM hazards in PSM marginally improved mean survival estimates, and estimation of time in health states not affected by GPM adjustment were still subject to error.

Conclusions: Caution should be taken with all modelling approaches when the underlying data are very limited, particularly PSM that does not appropriately incorporate GPM hazards. Further research is needed to improve MSM estimation with limited data.

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-05

Comprehensive evidence synthesis: Embracing single-arm and non-randomised studies in network meta-analysis

Rahma Said, Louise Haddon, Ellesha Smith, Stephanie Hubbar
Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom

Background: This work updates a systematic review, drawing on both randomised studies and non-randomised studies (NRS), to evaluate intervention effectiveness in clinical practice. In the original systematic review, NRS and randomised evidence were naively pooled, while single-arm studies were omitted. Given their substantial contribution to the evidence base, constituting 17% and 25% of total participants, respectively, this evidence could offer valuable evidence on intervention effectiveness. However, the incorporation of single-arm and NRS remains challenging. Selecting appropriate methods for their inclusion and addressing inherent study-design biases is crucial. Thus, this work seeks to integrate single-arm evidence and NRS into network meta-analysis (NMA), facilitating informed decision-making in evidence synthesis.

Methods: NMA models are applied to a motivating example within a Bayesian framework. Single-arm studies are firstly incorporated by imputing missing baseline data and their standard errors. A second method involved matching studies with baseline data closely resembling the single-arm study's participant characteristics. To incorporate NRS, power prior, hierarchical, and bias-adjusted models are utilized to assess their impact on effect estimates and uncertainty.

Results: The inclusion of single-arm studies expands the network of evidence by introducing two new interventions; the models found these to be the most effective. Informative priors and naive pooling yield results consistent with RCT-only estimates and reduce the uncertainty of effect estimates compared to RCT-only results. Down-weighting NRS minimally affected outcomes unless substantially discounted. Hierarchical models, considering study design, emphasized NRS results more than naive pooling. However, no method significantly reduces uncertainty compared to NMA.

Conclusion: Where there are high levels of single-arm studies and NRS in a review, it is efficient to incorporate all of the available evidence. However, there is limited application of these methods to real-life datasets in the literature, which restricts the understanding of where the methods can be clinically useful. This work contributes to the evidence and illustrates some of the benefits and limitations of incorporating evidence beyond RCTs. Although single-arm evidence offers additional insights and expands the evidence network, few studies implement the most effective interventions, necessitating further research for informed decision-making. While including NRS is advantageous, it is essential to evaluate the methods for their impact on estimates and uncertainty. Justifying complex models remains challenging when limited NRS exist. Further research is still required to understand where the methods can be most useful.

POSTER SESSION A-04:
Statistical Methods for Health Economic Evaluations

P-A04-06

A review of evidence level and causal support in prediction models of complications in health economic models of type 2 diabetes

Wenjun Wang[1], Zhonglin Chai[1], Junfeng Wang[2]
1 Department of Diabetes, Clinical Centre School, Monash University, Melbourne, Australia
2 Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, Netherlands

Background: Predicting the risk of complications is essential for health economic (HE) simulation models of type 2 diabetes (T2D). Clinical prediction models, both existing ones and newly developed ones, can serve as the risk equations to project changes in incidence of complications due to changes in risk factors. This implicitly requests the prediction models having a casual inference, which is usually not met since most of them are factual prediction models. The aims of this review are: to evaluate the evidence level of prediction models used in HE models of T2D; to assess the causal support of risk factors used in the prediction models.

Methods: Prediction models identified in a previous review of T2D HE models were included in this review [1]. Level of evidence was evaluated with the tool proposed by Reilly et al. [2], and the first 3 levels are relevant for HE modeling. We used a causation score containing 4 of the Hill's criteria (strength, consistency, temporality, and coherence) to assess the causal support of each risk factors included in the prediction models.

Results: Twenty-one prediction models were included in the analysis. Fourteen (66.7%) of them are rated as Level 1 (lowest evidence), five (23.8%) are at Level 2, and only two (9.5%) models are at Level 3. There are 64 different risk factors used as predictors in those prediction models. Among these risk factors, 18 variables (Education, Smoking status, BMI, Physical activity, HbA1c, White blood cell count, Heart rate, DBP, SBP, Triglycerides, Total cholesterol, HDL, Non-HDL cholesterol, LDL cholesterol, Total cholesterol/HDL ratio, Albumin, Urinary albumin/creatinine ratio, eGFR) are 'modifiable risk factors' and were assessed for the causal support.

Conclusion: The applicability of prediction models in HE models needs more consideration. The level of evidence is relatively low in prediction models used in T2D HE models. Intervening on risk factors and predicting based on modifiable risk factors may lead to incorrect incidence estimates when risk factors have less causal support.

References: *[1] Li X, et al. Prediction of complications in HE models of T2D: a review of methods used. Acta diabetologica, 60(7), 861-879. [2] Reilly B M, et al. Translating clinical research into clinical practice. Annals of internal medicine, 144(3), 201-209.*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-01

Assessing the impact of dietary fatty acids on leucocyte telomere length: A substitution model approach using the UK Biobank data

Vasiliki Bountziouka[1,2,3,4], Christopher Nelson[2,3], Veryan Codd[2,3], Nilesh Samani[2,3]
1 Computer Simulation, Genomics and Data Analysis Laboratory, Dept. of Food Science and Nutrition, University of the Aegean, Lemnos, Greece
2 Dept. of Cardiovascular Science, College of Life Science, University of Leicester, Leicester, United Kingdom
3 NIHR Leicester Biomedical Research Centre, Glenfield Hospital, United Kingdom
4 Population, Policy and Practice Research and Teaching Dept., GOS Institute of Child Health, UCL, London, United Kingdom

Background/Introduction: Telomeres are protective caps at the ends of chromosomes, which are responsible for maintaining genome stability. They shorten with age and are linked to several age-related diseases. The variation in telomere length among individuals of the same age suggests that telomere shortening is a modifiable factor that can be influenced by lifestyle factors such as diet. Omega-3 polyunsaturated faty acids (PUFA) may have protective effects on telomeres, whilst high saturated fat intake (SFA) was associated with accelerated telomere shortening. This study aims to assess the impact of substituting SFA with PUFA and/or mono-unsaturated FA (MUFA) on leucocyte telomere length (LTL).

Methods: Data from 143,447 UK Biobank participants (mean (SD) age: 56 (7.9) years old, 47% male, 96% white ethnic background), with valid LTL measurement and energy intake (EI), and complete data in additional covariates (i.e., ethnicity, white blood cell count, education, body mass index, smoking, physical activity, LDL, CRP) were analysed. Multivariable linear regression models were used to explore the association between macronutrients and LTL. As fat intake distribution of 13% SFA; 14% MUFA; 6% PUFA, in UK Biobank, slightly exceeds the recommendations, the substitution models were additionally used to examine the expected change in LTL when substituting 3% and 6% of SFA with the equivalent amount of MUFA.

Results: A 5% increase in EI from total fat, whilst holding constant the proportion of EI from protein and carbohydrate and at the expense of an equal amount of energy from alcohol intake, was associated with 6 months of age-related change in LTL (beta=0.011SD (95%CI: 0.007SD; 0.016SD)). Similarly, a 5% increase in the amount of energy from MUFA intake, at the expense of an equal amount of energy from alcohol intake and whilst holding constant the proportion of energy intake from the remaining macronutrients, was associated with an age-related change in LTL of approximately 1.4 years (0.032 (0.020; 0.044)). Substituting 3% and 6% of SFA with MUFA, whilst holding the overall EI and PUFA constant, was associated with approximately 1 year (0.022 (0.010; 0.033)) and 2 years (0.044 (0.021; 0.067)) of age-related change in LTL, respectively.

Conclusion: These findings suggest that dietary modifications, particularly increasing MUFA intake and reducing SFA intake, may influence LTL dynamics. These potential implications for telomere biology and cellular aging could further mitigate age-related diseases associated with telomere attrition.

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-02

Statistical methods for genetic studies of disease progression

Apostolos Gkatzionis, Robyn Wootton, Jasmine Khouja, Kate Tilling
University of Bristol, Bristol, United Kingdom

The term "disease progression" refers to outcomes that are subsequent of disease incidence, including relapse, hospitalization and mortality. Studies of disease progression traits are often referred to as case-only studies, as they must be conducted exclusively on individuals who have the disease of interest. As a result, these studies are known to suffer from a particular form of selection bias called index event bias. In this talk, we discuss index event bias in genetic studies. We consider both traditional genetic studies aiming to identify genetic variants associated with a progression trait, and Mendelian randomization (MR) studies aiming to use genetic data as instrumental variables to identify causes of progression. We review existing methods to adjust for index event bias in genetic studies and compare them in simulations. We then discuss the previously unexplored case where both the exposure and the outcome of an MR analysis are disease progression traits. We show that existing methods can in principle be used to adjust for index event bias in this "double progression" scenario, but more care must be placed on verifying these methods' assumptions. Two motivating examples include investigating the effects of smoking cessation on depression relapse and lung cancer mortality.

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-03

Identifying metabolic biomarkers associated with risk of pancreatic cancer: A case-subcohort study

Christiana Kartsonaki, China Kadoorie Biobank Collaborative Group
Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

Background: Pancreatic cancer has a very poor prognosis with 5-year survival of 5- 10% and is often diagnosed late. Metabolic risk factors, such as adiposity and diabetes, and biomarkers, such as plasma glucose, fatty acids, and some proteins, have been shown to be associated with risk, but the relationship of other metabolic biomarkers with risk of pancreatic cancer is unclear. We aimed to assess the associations between circulating metabolites and risk of pancreatic cancer.

Methods: We conducted a case-subcohort study on 695 incident pancreatic cancer cases and a subcohort of 900 individuals within the China Kadoorie Biobank, a prospective study of 512,891 adults, with 225 metabolic biomarkers measured in baseline plasma samples using nuclear magnetic resonance (NMR). Cox regression fitted using the Prentice pseudo-partial likelihood was used to estimate hazard ratios (HRs) of pancreatic cancer. Interactions with follow-up time were explored. The approach of Cox and Battey (2017) was used to identify sets of metabolic markers associated with risk. Multivariable models were fitted to assess whether metabolic biomarkers help predict risk of pancreatic cancer.

Results: Among 1573 individuals without a history of cancer at baseline, certain lipoprotein fractions, fatty acids and other metabolites were associated with risk of pancreatic cancer. Several lipid-related biomarkers were associated with risk of pancreatic cancer, with adjusted HRs per 1-standard deviation increase in biomarker of 0.74 to 1.37. Glycoprotein acetyls and ratios of monounsaturated and of saturated fatty acids to total fatty acids were associated with a higher risk, while sphingomyelins, glutamine, estimated degree of unsaturation and ratios of omega-6 fatty acids and of polyunsaturated fatty acids to total fatty acids were associated with lower risk. The discriminatory ability of a model with known risk factors increased when several metabolic biomarkers were included (C-statistic 0.76 to 0.78). Associations between some biomarkers and risk of pancreatic cancer were time-varying; for example fractions of medium and small HDL were inversely associated with short-term risk (within the first year of follow-up). Within the first year, discriminatory ability substantially increased when including metabolic biomarkers (C-statistic 0.88 to 0.94).

Conclusions: Several metabolic blood biomarkers were associated with risk of pancreatic cancer and may be useful for earlier diagnosis. Methodological considerations will be discussed.

Reference: *[1] Cox, D. R. and Battey, H. S. (2017). Large numbers of explanatory variables, a semi - descriptive analysis. Proceedings of the National Academy of Sciences of the United States of America, 114 (32), 8592–8595*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-04 Does time-adjusted admission severity improve stroke prognostication?

Zewen Lu[1,2], Halvor Næss[4], Matthew Gittins[1,2], Amit Kishore[2,3], Craig Smith[2,3], Andy Vail[1,2]

1 Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom
2 Manchester Centre for Clinical Neurosciences, Geoffrey Jefferson Brain Research Centre, Manchester Academic Health Science Centre, Salford Care Organisation, Northern Care Alliance NHS Foundation Trust, Manchester, United Kingdom
3 Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom
4 Department of Neurology, University of Bergen, Haukeland University Hospital, Bergen, Norway

Background: In acute stroke, symptom severity evolves from onset and the time to assessment may affect the recorded severity on admission. Cross-sectional data suggest higher average values of stroke severity with earlier admission following symptom onset. However, much research implicitly considers 'admission severity' as a fixed marker of stroke severity in sharp contrast with analogous statistical scenarios such as gestational age-adjusted birth weight. We therefore want to explore whether there is any statistical benefit in stroke prognostication of adjusting for the timing of baseline severity assessment and, if so, how best to do this.

Methods: We retrospectively analysed data (2006-2021) from the Bergen NORSTROKE study dataset. We included patients diagnosed with ischaemic stroke of known onset time who had admission National Institutes of Health Stroke Scale (NIHSS) scores recorded within 12 hours of onset and before acute interventions.

We compared five approaches in stroke prognostic models using logistic regression: admission NIHSS with no time adjustment; considering both variables; considering both variables with their interaction; continuous centiles and categorical centiles. We developed a centile model for this purpose using generalised additive models for location, scale, and shape (GAMLSS). The best centile estimation was selected by GAIC among models with different distributional assumptions and additive terms. The statistical gain of adjustments was evaluated by Nagelkerke's pseudo-R squared (explained variation), AIC and BIC (model fit) and area under ROC curves (discriminative power).

Results: 2388 patients were included: median (IQR) age 74 (62 to 83) y; 44% male; median (IQR) admission NIHSS and time to assessment 3 (1 to 9) and 2(1 to 4) hrs respectively. The best centile estimation was the zero-inflated gamma model with cubic splines (GAIC=13784). After adjustment for time, the Nagelkerk's R-squared, AIC and BIC were slightly improved for centile (0.338, 2596, 2608, 0.795), centile categories (0.348, 2587, 2639, 0.974), both variables (0.349, 2600, 2618, 0.799), and both variables with their interaction (0.349, 2575, 2598, 0.803) respectively compared with no adjustment (0.337, 2600, 2612, 0.8), except for the area under ROC curves.

Conclusion: We improved stroke prognostication by adjusting for the time of baseline severity assessment. However, the statistical benefit for all three aspects was rather limited. Including both variables with their interaction appeared best for these criteria and is simple to implement. This approach should be considered in future stroke prognostic research.

References: *(DOI links) [1] 10.1176/appi.ajp.157.2.163. [2] 10.1111/ane.12477. [3] 10.1111/j.1467-9876.2005.00510.x.*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-05

Different age-specific time trends of colorectal cancer incidence rates in the Czech Republic and a potential role of colorectal cancer screening: Application of an age-period-cohort model

Ondřej Májek[1,2], Ondřej Ngo[1,2], Kateřina Hejcmanová[1,2], Monika Ambrožová[1,2], Renata Chloupková[1,2], Štěpán Suchánek[3], Miroslav Zavoral[3], Jan Bureš[3], Jan Mužík[1,2], Karel Hejduk[1,2], Ladislav Dušek[1,2]

1 National Screening Centre, Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic
2 Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic
3 Military University Hospital, First Faculty of Medicine, Charles University, Prague, Czech Republic

Background: Whereas recent studies show decline or stabilisation in overall colorectal cancer (CRC) burden in some high-income countries, increasing incidence in young adults was also recently reported in many populations. CRC screening programmes are widely recommended for individuals aged over 50. CRC screening programme was firstly introduced in the Czech Republic in 2000, organised programme was enacted in 2009. We aimed to investigate recent time trends of CRC incidence in younger and older individuals in the Czech Republic.

Methods: We used data on CRC cases diagnosed during the time period 1980-2020 and recorded in the Czech National Cancer Registry. Age-period-cohort (APC) models using restricted cubic splines were fit using apcfit package (Rutherford et al., 2010) in Stata 15.1. Reference cohort was taken as the median date of birth among cases, period effects were averaged at zero. Czech population estimates were obtained from Human Mortality Database (HMD).

Results: In younger individuals (aged between 20 and 49), only small period effects were identified between 1980-2020 (95% CI limits for relative risks between 0.92 and 1.10). Substantial increase was identified for younger individuals born after 1980, estimated relative risk was 2.55 (95% CI 1.80-3.63, reference cohort 1956) for individuals born in 1999. On the contrary, in older individuals (aged between 50 and 79), small cohort effects were identified for individuals born between 1930-1969 (95% CI limits for relative risks between 0.96 and 1.13, reference cohort 1935), but substantial decrease in incidence was shown for recent period after 2012, estimated relative risk 0.77 (95% CI 0.75-0.78) in 2020.

Conclusion: Model-based analysis of CRC incidence using Czech national data confirms unfavourable time trends for younger individuals born after 1980, which could potentially lead to increase in burden of Czech population and health system in near future. At the same time, significant decreasing period effect was shown for individuals aged over 50 in the last decade, suggesting potential beneficial role of organised Czech CRC screening programme.

References: *[1] Rutherford MJ, Lambert PC, Thompson JR. Age–period–cohort modeling. The Stata Journal. 2010 Dec;10(4):606-27. [2] HMD. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on 25/09/2022).*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-06

Risk factors for anticitrullinated protein seropositivity in Older Poles

Anna Chudek[1], Przemysław Kotyla[2], Małgorzata Mossakowska[3], Tomasz Grodzicki[4], Magdalena Olszanecka-Glinianowicz[1], Jerzy Chudek[6], Aleksander Owczarek[1]

1 Health Promotion and Obesity Management Unit, Department of Pathophysiology, Faculty of Medical Sciences in Katowice, Medical University of Silesia in Katowice, Katowice, Poland
2 Department of Internal Medicine, Rheumatology and Clinical Immunology, Faculty of Medical Sciences in Katowice, Medical University of Silesia in Katowice, Katowice, Poland
3 Study on Ageing and Longevity, International Institute of Molecular and Cell Biology, Warsaw, Poland
4 Department of Internal Medicine and Gerontology, Jagiellonian University Medical College, Kraków, Poland
5 Division of Preventive Medicine and Education, Medical University of Gdańsk, Gdańsk, Poland
6 Department of Internal Medicine and Oncological Chemotherapy, Medical University of Silesia, Katowice, Poland

Introduction: Anticitrullinated protein antibodies (ACPA, evaluated usually as anti-cyclic citrullinated peptide – anti-CCP) seropositivity, is a common feature of rheumatoid arthritis (RA). The occurrence of anti-CCP antibodies is more specific than of rheumatoid factor (RF) for the RA diagnosis. The study aims to assess the risk factors of anti-CCP seropositivity in older Poles representative of the Polish population.

Methods: The anti-CCP antibodies titer was assessed in a randomly selected subgroup of 37.5% 1537of 4979) participants from 7 age cohorts (65 90+ 791 (51.5%) women and 746 (48.5%) men one third aged 85 The C reactive protein level 3 mg/dL was considered as a marker of high cardiovascular risk and CRP level 10 mg/dL indicated inflammatory status. Interleukin-6 cut-off value was established as a value for 95% of our study cohort. As hypoalbuminemia, we assumed the level of albumins below 35 g/L.

Results: There were 50 anti-CCP seropositive participants. The frequency of anti-CCP seropositivity was estimated at 3.25% (95% CI: 2.45 4.30%), being higher among women 4.05% (95% CI: 2.83 5.73) than men 2.41% (95% CI: 1.48 3.86). The frequency of anti-CCP seropositivity decreased with age from 4.29% in aged 65 74 years and 4.07% in aged 70 84 years to 1.5% aged 85 years or above ($p < 0.05$). The risk increased with hypoalbuminemia, higher serum CRP level (as well as CPR > 3 mg/dL) or plasma IL 6 level (as well as IL 6 ≥ 10 pg/mL), inflammatory status (CRP > 10 mg/dL or IL 10 pg/mL), inflammatory status (CRP > 10 mg/dL or IL 6≥ 10 pg/mL), inflammatory status (CRP > 10 mg/dL or IL 10 pg/mL), and female gender. In older age (as well as age ≥ 10 pg/mL), inflammatory status (CRP > 10 mg/dL or IL 85 yrs), serum albumin levels and alcohol consumption decreased the risk of seropositivity. Multivariable logistic regression revealed that hypoalbuminemia (OR = 6.29), and inflammatory status (OR = were independently associated risk factors of seropositivity, while age 85+ and alcohol consumption were protective factors. No significant difference in 5-year survival rates was noted between seropositive and seronegative subjects (parallel Kaplan-Meier survival curve analysis p = 0.85).

Conclusion: Hypoalbuminemia and inflammatory status were risk factors, while age 85+ and alcohol consumption decreased the odds of seropositivity. The anticitrullinated protein seropositivity is not associated with shorter survival.

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-08

Environmental factors and their impact on Emergency Department visits in Lyon, France

Levi Monteiro Martins[1,2,3], Elsa Coz[1,2], Celine Piegay[4,5], Arnaud Friggeri[6,5,7], Mohand-Saïd Hacid[3,1], Delphine Maucort-Boulch[8,5,2,1]
1 Université Claude Bernard Lyon 1, Villeurbanne, France
2 CNRS UMR 5558 - Laboratoire de Biometrie et Biologie Evolutive, Equipe Biostatistique Santé, Villeurbanne
3 CNRS UMR 5205 - Laboratoire d'Informatique en Image et Systèmes d'Information, Villeurbanne, France
4 Département d'Information Médicale, Pôle de Santé Publique, Lyon, France
5 Hospices Civils de Lyon, Lyon, France
6 Service d'anesthésie, Service de Médecine Intensive, Service de Médecine péri-opératoire et de Réanimation, Lyon, France
7 Groupement Sud, Lyon, France
8 Service de Biostatistique, Lyon, France

Introduction: Emergency department overcrowding can lead to an increase in hospital admissions, resulting in various issues. Numerous studies have shown a link between environmental factors, such as air pollution and meteorological conditions, and emergency department (ED) visits. Our aim is to explore the delayed and non-linear effects of environmental factors on daily visits to emergency services for reasons related to circulatory or respiratory system diseases (International Classification of Diseases, 10th revision, I00-I99 and J00-J99).

Methods: We propose the use of a Quasi-Poisson generalized linear regression and a distributed lag nonlinear model. Additionally, we propose to conduct subgroup analyses by age (>=18<65, >=65), gender (male, female), French deprivation score (1-3, 4-5), and Charlson comorbidity index (0, 1, 2, and >2). Visits to three hospitals affiliated with the Hospices Civils de Lyon were included for the period from 1st January 2009 to 31st December 2019. Hospital data were extracted from the Program for Medicalization of the Information Systems (PMSI)1. The pollutants included in this study were fine particles PM2.5 and PM10, O3, NO2, SO2, and CO. Meteorological data included temperature, wind speed, relative humidity (RH), and rainfall.

Results: For combined ED visits, i.e. visits with entry reason I00-I99 or J00-J99, extremely cold temperatures (≤-2.77°C) and low RH increased the risk of visits. Extremely cold temperatures at lag 7 showed a positive association with RR of 1.016 (95% CI: 1.001-1.031). A positive association was also found at lags 5, 6, and 7 for the RH value at the 1st percentile (39.58), with RRs of 1.027 (95% CI: 1.003-1.051), 1.03 (95% CI: 1.006-1.055), and 1.033 (95% CI: 1.008-1.058). For visits related to circulatory system diseases, an extreme concentration, 99th percentile (13.10), of SO2 showed a positive association at lags 0, 1, and 2, with RRs of 1.105 (95% CI: 1.033-1.181), 1.076 (95% CI: 1.023-1.132), and 1.049 (95% CI: 1.008-1.093). Furthermore, for visits related to respiratory diseases, extreme concentrations, 99th percentile (106.95), of O3 showed a positive association at lags 0 and 1 with RRs of 1.081 (95% CI: 1.002-1.165) and 1.058 (95% CI: 1.001-1.118).

Conclusion: Our preliminary results for the combined ED visits showed no significant association with air pollutants. However, temperature and RH exhibited an immediate positive effect on initial lags. For visits related to circulatory system diseases, the extreme concentration of SO2 has an immediate positive effect. Furthermore, for visits related to respiratory diseases, extreme concentrations of O3 have an immediate positive effect.

Reference: *[1] https://pubmed.ncbi.nlm.nih.gov/10544715/*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-09

Association between the oxidative balance score and the quality of life based on data from the Korean National Health and Nutrition Examination Survey

Jinheum Kim, Eunjeong Cha
University of Suwon, Hwaseong, Republic of Korea

Background: Oxidative stress is a complex process resulting from an imbalance between anti- and pro-oxidant components. Oxidative balance scores (OBS) are measures that have arisen to assess the overall balance of individuals' oxidation-reduction status. The purpose of this presentation is to investigate the association between OBS and quality of life (QoL).

Methods: The data from the Korean National Health and Nutrition Examination Survey in 2020 were utilized, which included 4,192 people aged 19 and above. OBS consists of twelve dietary (vitamins C, D, E and B9, β-carotene, fiber, zinc, retinol, calcium, saturated fatty acid, poly-unsaturated fatty acid, and iron intakes) and three lifestyle components (smoke, drink, and physical activities). OBS was also defined as the components' simple total and the first principal component score, respectively. In addition, OBS was classified using non-parametric approaches. To investigate the effect of OBS on QoL, three models were analyzed: one with simply OBS (M1), one with OBS and sociodemographic characteristics such as gender, age (19-39, 40-64, and 65 and above), household income (1st to 5th quintile), education (elementary, middle to high, and college and above), and single-person household (M2), and one with OBS plus sociodemographic data, the presence of chronic disease and atopic dermatitis, and sleep quality (<7, 7-8, and 9 and above hours) (M3).

Results: According to the simple totals, OBS was considerably significant on EQ-5D index in all three models (p-value = 0.0028, 0.0208, and 0.0301 for M1, M2, and M3, respectively). Based on the first principal component scores, OBS was significant on EQ-5D index in M1 (p-value < 0.0001) but not in M2 (p-value = 0.1311) and M3 (p-value = 0.1665).

Conclusion: We found from M3 that the higher the OBS (p-value = 0.0301), the female aged 19-39 (p-value < 0.0001), the higher the household income (p-value < 0.0001), the higher the level of education (p-value < 0.0001), the non-single-person household (p-value = 0. 0048), the absence of atopic dermatitis (p-value = 0. 0016), and the sleep time of 7-8 hours compared to less than 7 or more than 9 hours, the higher (better) the EQ-5D index (QoL).

*This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00239958).

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-10

A lesson for understanding parents' perspective: Perception of quality of care and COVID-19 related fears among users of paediatric health services over the COVID-19 pandemic in 11 facilities in Italy

Paolo Dalena[1,2], Alessandro Zago[1], Egidio Barbi[1,2], Marzia Lazzerini[2]
1 University of Trieste, Trieste, Italy
2 IRCCS Burlo Garofolo, Trieste, Italy

Background: the COVID-19 pandemic had an important psychological impact on children and their families. This study aims to explore pediatric health services users' perceptions of quality of care (QOC) and COVID-19 related fears, and their changes over time in relation to COVID-19 pandemic.

Methods: in a multi-center cross-sectional study involving 11 public hospitals providing pediatric care across the Italian territory, we collected data from services users through a validated questionnaire [1] based on the WHO standards for improving QOC for children at facility level [2]. We analyzed four indicators: 1) QOC perceived in relation to COVID-19; 2) Overall QOC perceived; 3) Fear of accessing health services due to COVID-19; 4) Fear of contracting COVID-19 in hospital - and calculated Spearman's correlation indexes (r) with the number of COVID-19 new cases over time. Subgroup analyses were conducted by macroregions and single facility.

Results: Data from 956 services users were analyzed. QOC indicators were stable over time at values closed to the maximum (ranges 77-100 for COVID-19 QOC and 74- 98 for overall QOC), and no correlation were found with the COVID-19 new cases (r = - 0.073 and -0.016, respectively). Fear of accessing care and fear of contracting the infection varied over time in between 0-52% and 0-53%, respectively, but did not correlate directly with COVID-19 new cases (r = 0.101, 0.107 and 0.233, 0.046 respectively). At subgroup analyses, significantly higher frequencies of fear (p-values < 0.05) and lower QOC (p-values < 0.001) were reported in South Italy, and three facilities showed moderate correlation between indicators.

Conclusions: COVID-19 related fears and perceived QOC may be mediated by more complex cultural and geographical factors, than simply by epidemic peaks. Subgroup analyses can unpack major differences within the same country.

References: *[1] Lazzerini M, Mariani I, de Melo E Lima TR, Felici E, Martelossi S, Lubrano R et al; CHOICE Study Group. WHO standards-based tools to measure service providers' and service users' views on the quality of hospital child care: development and validation in Italy. BMJ Open. 2022; doi: 10.1136/bmjopen-2021-052115. [2] Standards for improving the quality of care for children and Young Adolescents in health facilities [Internet]. World Health Organization. World Health Organization; 2018 Available from: https://www.who.int/publications-detail-redirect/9789241565554. Accessed June 2023.*

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-11 The current use of reliability analysis for questionnaires in dental research

Natchalee Srimaneekarn, Kawin Sipiyarak, Tawepong Arayapisit
Faculty of Dentistry, Mahidol University, Bangkok, Thailand

Introduction: Questionnaires are often used as the main tool for collecting data in survey research. The validity and reliability are widely used to test the quality of the questionnaires. This lead to the objective of this study is to survey the type of reliability using in questionnaires of dental research and investigate the relationship between the use of reliability and ranking in dental research.

Methods: 385 dental researches using questionnaire were selected from Scopus and Pubmed by specific inclusion and exclusion criteria. The research information included author name, research title, year of publication, journals, quartile according to SJR indicator, method of creating questionnaire, the use of reliability and its statistics. The use of reliability was categorized into Test-retest reliability, Internal consistency reliability, and Alternate Forms reliability. All statistical analyses were performed by SPSS. Descriptive statistics were used to analyze overall data. Chi square testing was used to analyze the test of correlation between the use of reliability and quartiles.

Results: The percentage of using reliability in questionnaires of dental research was 84 researches 21 8%). According to the researches that use reliability in questionnaire, Internal consistency was one of the reliability test which is used the most from all researches collected 67 4% while Test retest reliability was used for 27 2%. No statistically significant association was found between use of reliability and SJR indicator.

Conclusion: This study found that 1 out of 5 dental research are using reliability in questionnaires but only 2 out of 3 types of reliability was used. This study found that Internal consistency is the most prevalent form of reliability used in research, with Cronbach's alpha being the most commonly approximately 90% used to measure it. Test retest varies in terms of the specific statistics used, with the most commonly is ICC and Cohen s Kappa, respectively.

Alternate Forms reliability was not found in this study. Moreover, this study found that the use of reliability is not related to quartile rank, which mean reliability test has a potential to use in any ranking. In conclusion, we suggest 3 statistical analyses which are Cronbach s alpha, ICC, and Cohen's kappa to be emphasized in reliability learning course, and well--understood using and interpretation in further research.

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-12

Challenges in the establishment of minimal important changes (MIC) for outcomes by patients undergoing an arthroscopic rotator cuff repair in Switzerland

Thomas Stojanov[1], Berend Terluin[2], Andreas Marc Müller[1], Giuseppe Filardo[3], Study Group ARCR_Pred[1], Laurent Audigé[1]
1 University Hospital Basel, Basel, Switzerland
2 Amsterdam UMC, Amsterdam, Netherlands
3 EOC Lugano, Lugano, Switzerland

Background: Establishing minimal important change (MIC) thresholds is crucial in orthopedics for interpreting patient-reported outcome measure (PROM) data. Various methods have emerged to determine MIC, each with its own advantages and drawbacks. However, MIC values for PROM to assess patients undergoing arthroscopic rotator cuff repair (ARCR) in Switzerland remain unclear. This study aims to compare MIC values using three different methods for the Oxford Shoulder Score (OSS) among ARCR patients in Switzerland.

Methods: Between June 2020 and November 2021, primary ARCR patients were prospectively enrolled in a multicenter cohort across 18 Swiss and one German orthopedic center. The OSS, a PROM describing shoulder function and ranging from 0 to 48 points, was collected at baseline (N = 973), 6- (N = 910), 12- (N = 873), and 24-month (N = 854) follow-ups. We used a 5-item Likert anchor question to assess subjective shoulder function. We assessed the ceiling effect of OSS, the proportion of minimally improved patients, and the present state bias (PSB). Three methods (receiver operating characteristics-based, predictive modeling, adjusted predictive modeling) were compared to establish MIC values.

Results: The proportion of minimally improved patients was 86%, 93%, and 94% at 6-, 12-, and 24-month follow-up. Ceiling effect occurred in 9%, 30%, and 42% at the respective timepoints. PSB was 0.56, 0.65, and 0.75. MIC values ranges were 5 to 8, 10 to 13 and 10 to 14 points.

Conclusions: Our study highlighted that nearly all patients with available data at follow-up experienced a minimal improvement after an ARCR in Switzerland. We found that a large improvement in the OSS was necessary to categorize patients as being minimally improved in the Swiss population. However, the large proportion of minimally improved patients, ceiling effects, and PSBs suggest that our MIC estimates may be overestimated. Alternative analytical approaches addressing these issues should be considered.

POSTER SESSION A-05:
Statistics in Epidemiology

P-A05-14 Why has rank-based non-parametric hypothesis testing become the gold standard for comparison of non-normal distributed data?

Jan Brink Valentin
Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Background: Non-parametric rank-based hypothesis tests such as the Mann-Whitney U test (MWUt) were originally developed to estimate whether the pairwise concordance between two intervention groups is statistically different from a 50/50 ranking based on some continuous outcome. Unsurprisingly, this corresponds to testing whether the concordance index is different from 0.5. The test may have merit for some epidemiological comparisons, however, for testing differences in means, such as for time and/or cost optimization trials, the MWUt may not provide an adequate estimate of statistical significance. Nevertheless, a review of 4 high ranking medical journals reveals a continuous usage of the MWUt without specifying the effect measure under consideration. This study aims to investigate to which extent the MWUt can provide an adequate estimate of statistical significance for comparisons of means, medians, and distributions.

Methods: We simulated various data sets with varying distributions and number of observations. For comparisons of means and medians, we used a permutation test with 1000 permutations as gold standard. We used a t-test and Mood's test of medians as benchmarks. For comparisons of distributions, we used a permutation test on a concordance index estimated by applying regression analysis with restricted cubic splines (3 knots) as gold standard. The same approach, but without restricted cubic splines, was used as gold standard for testing whether the concordance index was different from 0.5.

Results: For comparisons of means the p-value of the MWUt and t-test deviated from the gold standard with a maximum value of 0.99 and 0.09, respectively. For comparisons of medians both the MWUt and Mood's test of medians deviated from the gold standard with a maximum value of 1.00. For comparisons of distributions the MWUt deviated from the gold standard with a maximum value of 1.00. However, for testing whether the concordance index was different from 0.5 the MWUt deviated from the gold standard with a maximum value of 0.14.

Conclusion: The results show that the MWUt fails to provide an adequate estimate of statistical significance for comparisons of means, medians, and distributions. In addition, the standard t-test was shown to be robust to large deviations from the assumption of residual normality, while Mood's test of medians was as inaccurate as the MWUt. We recommend that a p-value estimated from the MWUt is presented with the corresponding concordance index to emphasise the measure of effect.

POSTER SESSION A-06:
Causal Inference

P-A06-01

Target trial emulation to leverage randomized trial data: investigate alternative questions of interest in late stage development of monoclonal antibodies in Alzheimer's disease

Rachid Abbas[1], Emilie Lanoy[2,3], Paul Delmar[1], Sandrine Katsahian[2,3,4]
1 F. Hoffmann - La Roche, Basel, Switzerland
2 AP-HP, Hôpital Européen Georges Pompidou, Unité de Recherche Clinique, APHP Centre, Paris, France
3 Inserm, Centre d'Investigation Clinique 1418 (CIC1418) Epidémiologie Clinique, Paris, France
4 HeKA, INRIA PARIS, Université de Paris, Paris, France

Background/Introduction: Target trial emulation, a method to draw causal conclusions from observational data (Hernan, JAMA 2022), can be applied to clinical trial data. Phase III randomized clinical trials (RCTs) evaluating experimental molecules are designed to address a pre-specified question for registration purposes. In case of high attrition, incomplete observance or dose adaptation, leveraging the large amount of data collected to consider counterfactual situations becomes crucial: what would have happened in absence of attrition, perfect observance or in a real-life population. We propose to use trial emulations to address these alternative causal questions enhancing the RCT intent-to-treat (ITT) analysis. We illustrate our strategy with a post-hoc analysis of RCTs in Alzheimer's disease (AD) aligning definitions for exposure and inclusion criteria to those of a competing trial.

Methods: The data source was two RCTs evaluating the efficacy of a monoclonal antibody targeting amyloid beta at full dose after 36 weeks of up-titration in mild to moderate AD patients. The primary endpoint was the change from baseline to 116 weeks in a continuous clinical scale measuring cognition and function. We emulated target trials following three alternative frameworks aligning study characteristics on a competing RCT: all participants on trial at week 36 (OT36), protocol-planned full dose started at week 36 in randomized population (FD36R), and in a population observing stricter inclusion criteria (FD36S). Counterfactual exposure effects were estimated using marginal structural models (MSM) with inverse probability weighting (IPW).

Results: While the ITT analysis in 1959 participants identified a difference in adjusted mean change of -.29 (95%CI: -.58 ; -.01) in the primary outcome associated with experimental treatment after 116 weeks, corresponding effects derived from MSMs were -.30 (95%CI: -.59 ; -.01) in 1805 (92%) observations from OT36, -.33 (95%CI: -.69 ;.02) in 1224 (62%) observations from FD36R and -.27 (95%CI: -.55 ;.01) in 905 (46%) observations from FD36S. Aligning study characteristics to a successful competing RCT did not modify the original RCT conclusions.

Conclusions: Emulating target trials in a less advanced AD population exposed to the protocol-planned dose ruled out strong influence of attrition during the up-titration period and of eligibility criteria on observed outcomes. Trial emulation improves the use of clinical trial data, enriches conclusions from primary analyses, allows pertinent comparisons with competitors trials, and paves the way for the subsequent pivotal trials.

POSTER SESSION A-06:
Causal Inference

P-A06-02

Drug repurposing of metformin via emulating a target trial for preventing chronic kidney disease progression

Nicole Fontana[1,2], Carlo Andrea Pivato[2,3], Piercesare Secchi[1], Emanuele Di Angelantonio[2,4,5,6,7], Francesca Ieva[1,2]

1 MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy
2 HDS, Health Data Science Center, Human Technopole, Milan, Italy
3 IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
4 Blood and Transplant Research Unit in Donor Health and Behaviour, Cambridge, United Kingdom
5 Dept of Public Health & Primary Care, University of Cambridge, Cambridge, United Kingdom
6 British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, United Kingdom
7 Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, United Kingdom

Background/Introduction: Drug repurposing represents an effective strategy to identify new uses for existing drugs, speeding up the transition from research to clinical practice. Leveraging observational data, we aim to emulate a target trial for repurposing metformin as a potential renoprotective agent for chronic kidney disease (CKD) progression. Given its established history and affordability as a primary treatment for type 2 diabetes, metformin offers promise as a candidate to mitigate CKD advancement since other classes of glucose lowering therapies, such as SGLT2 inhibitors and GLP1 receptor agonists, have shown beneficial effects in preventing CKD progression [1]. This study simulates a target trial of metformin use and estimates intention-to-treat (ITT) and per-protocol (PP) effects on CKD advancement.

Methods: Using UK Biobank data, we conducted sequential "trials" in which individuals were classified as metformin initiators or non-initiators and were followed until progression of CKD or loss to follow-up. The treatment effects were estimated via a two-stage process. Firstly, we applied inverse-probability weighting to estimate baseline treatment, follow-up and treatment persistence weights to adjust for baseline confounding, loss to follow-up, and non-persistence with the initial treatment, respectively. In the second stage, we used weighted pooled logistic regression models to estimate the initiation vs non-initiation effects (ITT analysis) and the sustained use vs non-use effects (PP analysis).

Results: Of 81,388 eligible individuals, there were 3,347 initiators and 78,041 non initiators. Our analyses were adjusted for potential confounding factors, including demographics, comorbidities, treatments and laboratory information, all of which are considered direct causes of metformin use or CKD. Comparing metformin initiation with no initiation, the adjusted hazard ratio for the progression of CKD was 0.44 (95% CI (0.21, 0.91) in ITT analysis and 0.39 (95% CI (0.23, 0.65)) in PP analysis.

Conclusions: Metformin initiation after a diagnosis of stage 2 or 3 of CKD was associated with a lower risk of diagnosis progression, with a potentially greater benefit in patients who persisted with metformin over time. This study lays a crucial foundation for future clinical trials, wherein the findings can be validated and the effects comprehensively understood within a randomized setting. By increasing the number of exposures participants and events subsequent trials can yield deeper insights and more reliable conclusions.

References: *[1] Rangaswami J, Bhalla V, et.al. Cardiorenal protection with the newer antidiabetic agents in patients with diabetes and chronic kidney disease: a scientific statement from the American Heart Association. Circulation. 2020;142:e265–e286.*

POSTER SESSION A-06:
Causal Inference

P-A06-03 Target trial emulation to overcome immortal time bias in real-world data of urothelial cancer

Josef Fritz[1,2], Mag. Sarah Maier[1], Renate Pichler[1], Andreas Kronbichler[1], Hanno Ulmer[1]
1 Medical University of Innsbruck, Innsbruck, Austria
2 Lund University, Malmö, Sweden

Background: Immortal time bias is an often-overlooked challenge in longitudinal clinical studies where time zero and exposure assignment are not aligned. To minimise this bias, several approaches have been proposed, including time-dependent analysis. Recently, target trial emulation (TTE) was introduced as an additional promising tool. We contrasted the application of TTE with naïve and time-dependent analysis using real-world data from patients with urothelial cancer, which are highly affected by immortal time bias.

Methods: This study utilised data from a multicentre observational study of patients with metastatic urothelial carcinoma who have been receiving immune checkpoint inhibitors since 2016. The primary aim of our study was to assess the impact of immune-related adverse events (irAEs) on overall survival (OS). Since irAEs could occur anytime during the follow-up, appropriate methodology is needed to avoid immortal time bias. Multivariably adjusted standard Cox proportional hazards regression models were used for the naïve and time-dependent approaches. For the TTE approach, we decided to emulate a sequence of 45 hypothetical clinical studies by using patients repeatedly ('cloning'), starting each month of follow-up. In each study, patients experiencing an irAE in a certain month were defined as exposed, patients with irAEs occurring later (or never) were defined as controls and OS was censored at its occurrence. To account for this artificial censoring, inverse probability censoring weighting was performed. Finally, hazard ratios were estimated using a pooled weighted Cox model, 95% confidence intervals (CIs) were calculated using bootstrapping.

Results: Out of the 335 patients included in the study, 193 died during the follow-up period. The naïve Cox model, which did not account for time-dependent exposure, yielded a hazard ratio (HR) of 0.46 (95% CI: 0.33-0.63), indicating that the occurrence of irAEs during follow-up strongly improves survival. However, the time-dependent Cox model did not show any significant effect of irAEs on OS (HR=1.02, 95% CI: 0.72-1.44). The TTE approach produced a similar result (HR=1.07, 95% CI: 0.71-1.52).

Conclusion: The study showed that TTE is a suitable tool for overcoming immortal time bias, as demonstrated by its agreement with the time-dependent Cox model. In contrast, the naïve Cox model was substantially biased. Although the TTE concept was initially developed to assess treatment effects, this study suggests that it can also be used to eliminate immortal time bias in studies with time-dependent non-interventional exposures.

POSTER SESSION A-06:
Causal Inference

P-A06-05

Using inverse probability of censoring weighting to estimate hypothetical estimands in clinical trials: Should we implement stabilisation, and if so how?

Jingyi Xuan[1], Shahrul Mt-Isa[2], Nicholas Latimer[3,4], Helen Bell-Gorrod[3], William Malbecq[5], Kristel Vandormael[6], Victoria Yorke-Edwards[1,7], Ian White[1]

1 MRC Clinical Trials Unit at UCL, University College London, London, United Kingdom
2 Biostatistics and Research Decision Sciences (BARDS) Health Technology Assessment (HTA) Statistics, MSD, Zurich, Switzerland
3 Sheffield Centre for Health and Related Research, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom
4 Delta Hat Limited, Nottingham, United Kingdom
5 Department of Mathematics, University of Brussels, Brussels, Belgium
6 Biostatistics and Research Decision Sciences (BARDS) Health Technology Assessment (HTA) Statistics, Brussels, Belgium
7 Centre for Advanced Research Computing, University College London, London, United Kingdom

Introduction: In the presence of intercurrent events (ICEs) in clinical trials, inverse probability of censoring weighting (IPCW) can be used to estimate the hypothetical estimand targeting the treatment effect if such ICEs had not occurred. Though IPCW yields unbiased estimates when the "No unmeasured confounders" and "Positivity" assumptions are satisfied, large standard errors (SEs) and wide confidence intervals can be a concern. To deal with this issue, stabilisation is claimed to decrease the extreme weights and reduce the numerical instability of small probabilities by replacing the numerator 1 in the unstabilised weighting model with alternative models accounting for time and/or baseline covariates. We aim to compare the performance of IPCW with different types of weights under situations where the outcome analysis model is correctly specified or mis-specified in different trial settings.

Methods: We simulate datasets informed by realistic trial settings. The data-generating mechanism varies the occurrence of an ICE in one or both arms, the prevalence of the ICE in each arm, the magnitude of the treatment effect, whether the treatment effect is time-varying and sample size. The estimand is the marginal risk difference, to avoid non-collapsibility issues when adjusting for baseline covariates. IPCW implementations with different choices of the weighting model including unstabilised weights and stabilised weights with time and/or baseline covariates in the numerator, and whether the outcome analysis model is correctly specified or mis-specified are investigated.

Results: When the outcome analysis model does not account for the time-varying treatment effect, IPCW with stabilised weights adjusting for time in the numerator reduces SEs compared with IPCW with unstabilised weights. When time-varying treatment effect is accounted for in the outcome analysis model, IPCW stabilising for time does not provide an obvious reduction in SEs. When stabilised weights with baseline covariates are used, smaller SEs are obtained. However, when the outcome model is mis-specified, IPCW with stabilised weights can yield larger bias.

Conclusion: Whether and how to implement stabilisation in IPCW should be considered empirically regarding the estimand of interest and the pre-specified methods for estimation. IPCW with stabilised weights is preferred to reduce standard errors but is more subject to bias when the model is mis-specified.

References: *[1] Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550-60.*

POSTER SESSION A-06:
Causal Inference

P-A06-06

Causal inference for integrating external data in randomised trials

Xiner Zhou[1], **Christiana Drake**[2]
1 Harvard T Chan School of Public Health, Boston, United States
2 University of California, Davis, Davis, United States

Randomized trials often utilize a select group of study participants. This group does not typically represent the general population. Furthermore, sample sizes are often small to reduce cost. To improve power and generalizability, external control groups may be added to the randomized study. We show using tools of causal inference how to incorporate a suitably selected external control group into the study design to improve efficiency. After the randomized phase III study ends and the drug under investigation is found to have efficacy, an observational study sometimes follow. In chronic diseases with a placebo control group, if the drug is deemed effective, all subjects are switched to treatment at the end of RCT. Long term effects of drug use are of interest. There is no longer a randomized control group and a comparison group must come from some other source, the external control group. This type of study is no longer randomized and is subject to confounding. To minimize bias investigators often use historical control groups from similar studies of the same disease. We propose a difference in difference estimator and discuss the conditions required for unbiased causal conclusions. The techniques are applied to a study of a novel drug treatment for Spinal Muscular Atrophy.

POSTER SESSION A-06:
Causal Inference

P-A06-07

Heterogeneous treatment effect estimation for observational data using model-based forests

Susanne Dandl[1,2], Andreas Bender[1,2], Torsten Hothorn[3]
1 LMU Munich, Germany
2 Munich Center for Machine Learning, Germany
3 University of Zurich, Switzerland

Introduction: The estimation of heterogeneous treatment effects (HTE) poses a huge challenge for complex (e.g., survival or ordinal) outcomes that cannot be modelled traditionally with linear models. This is due to the noncollapsibility issue, wherein HTE estimates are shrunken towards zero, and become incomparable between individuals. Model-based forests (MBF) provide a practical solution to this problem by allowing the full estimation of the model - including prognostic effects. So far, they have only been applied to randomized trial data but have not been explored for observational data, for which confounding effects could bias the estimation of HTEs. In our work, we propose and evaluate novel MBF variants for HTE estimation in the case of observational data which are applicable to many clinically relevant outcomes, under generalized linear or transformation models.

Methods: Our proposed approaches are rooted in the orthogonalization strategy by Robinson (1988). Given data on the outcome Y, treatment indicator W, and some covariates X, our approach consists of the following steps: First, a machine learning model is trained to estimate propensities, i.e., $\pi(x) = P(W|X = x)$. Then, the treatment indicator W for each observation is centred by subtracting $\pi(x)$. An appropriate model for the outcome is set up (e.g., a generalized linear or transformation model), given the centred treatment. Additionally, an offset capturing the marginal effect of X on Y can be incorporated into the model. MBFs are then deployed to account for the heterogeneity in estimated effects.

Results: A simulation study with diverse outcomes (continuous, binary, ordinal, and survival) shows that centring W is effective in mitigating biases in HTE estimation in the case of observational data where confounding is present. Adding an offset further improved the performance – especially in cases with very strong confounding. We illustrate the practicality of our proposed approach by estimating the HTE of Riluzole on the progression of Amyotrophic Lateral Sclerosis.

Conclusion: In conclusion, our work, rooted in MBFs and the orthogonalization strategy by Robinson, presents a first holistic approach to address the challenging problem of HTE estimation in observational studies for complex outcomes beyond mean regression. Extensions to multivariate outcomes, the application of cross-fitting, and a sensitivity analysis to violations of underlying assumptions offer exciting opportunities for further research.

POSTER SESSION A-06:
Causal Inference

P-A06-08

Causal updating of prediction models in a pandemic environment

Ilaria Prosepe[1], Anna Kaal[1,2], Claudine Stark[1], Jim Smit[3], Emiel Leegwater[2], Cees van Nieuwkoop[2], Sesmu Arbous[1], Nan van Geloven[1]
1 Leiden University Medical Center, Leiden, Netherlands
2 Haga Teaching Hospital, The Hague, Netherlands
3 Erasmus Medical Center, Rotterdam, Netherlands

Background/Introduction: Predictive algorithms are often used in hospitals to help guide patient management and care planning. However, when treatment policies change over time, they may quickly loose predictive performance. While it is true that predictive algorithms can be updated with new data, often they cannot be updated fast enough, as we need to wait until the new data is collected. An "interventional" prediction model, which incorporates emerging treatment evidence into the prediction model, theoretically allows for faster updating. This is different from a "standard" prediction model, where treatment effects are not incorporated.

Methods: We used data extracted by a natural language processing and text mining-based tool from the electronic health records of patients hospitalised with COVID-19 in four Dutch hospitals in 2020 and 2021. We trained prediction algorithms on first-wave data (February-May 2020) and tested them on second-wave data (August 2020-February 2021). The estimand of interest was the 28-day risk of mortality or need for ICU, assessed at hospital admission. We hypothesized that an interventional model would adapt quicker to new data, addressing the challenge of changing treatment policies over time. We compared the performances of a "standard" and an "interventional" prediction model, by means of discrimination (c-index and AUCt), calibration (plot, root mean squared bias and calibration intercept + slope) and overall performance (Brier score). For both models we allow updating steps using intercept recalibration, refitting and Bayesian updating.

Results: We analyzed data from 4064 patients. After scanning their medication use over time, we found dexamethasone to be the most important change in treatment policy during the 2020-2021 time period. We incorporated evidence on the effectiveness of dexamethasone provided by the RECOVERY trial, based on the results that became available in July 2020. The comparison of performance of the two modelling approaches is currently still in progress.

Conclusions: While it has been theoretically described that "interventional" prediction models may be more robust to changes in treatment policies over time, there are not many applications yet. Data collected during Covid-19 pandemic provides a unique opportunity to put this hypothesis to the test using empirical data. In this Covid-19 model comparison, we shed new light and insights on the feasibility of faster model updating by means of interventional prediction models.

POSTER SESSION A-06:
Causal Inference

P-A06-09

Marginal structural models for estimating causal risk ratio and causal risk difference in longitudinal studies

Hiroyuki Shiiba[1,2], Hisashi Noma[3], Keisuke Kuwahara[4,5,6], Tohru Nakagawa[7], Tetsuya Mizoue[4]
1 The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies (SOKENDAI), Japan
2 Clinical Planning and Development Department, Medical Head Quarters, Eisai Co., Ltd., Japan
3 Department of Data Science, The Institute of Statistical Mathematics, Japan
4 Department of Epidemiology and Prevention, Center for Clinical Sciences, National Center for Global Health and Medicine, Japan
5 Department of Public Health, Yokohama City University School of Medicine, Japan
6 Department of Health Data Science, Graduate School of Data Science, Yokohama City University, Japan
7 Hitachi Health Care Center, Hitachi Ltd., Japan

Background: The marginal structural model (MSM) is a recognized approach for causal inference, designed to handle the issue of time-varying confounders in longitudinal studies. In the analysis of longitudinal data with binary outcomes, the generalized estimating equation (GEE) logistic regression within the MSM framework is widely employed to estimate causal odds ratios. However, the odds ratio is primarily interpretable as an approximate estimate of the risk ratio only in cases of rare events and lacks clear clinical or epidemiological interpretations. Consequently, recent statistical guidelines recommend the use of risk ratio and risk difference instead of odds ratio. Nevertheless, despite the recommendation, there is a lack of standard methods within the MSM methodology for directly estimating causal risk ratio and risk difference. Therefore, we propose straightforward and efficient models to obtain the causal risk ratio and risk difference using the framework of applying GEE to MSM, even in situation where event frequency is high.

Methods: We propose adopting the Poisson and Gaussian GEE regression models in the MSM analyses. Although we can also obtain the risk ratio and risk difference estimates by using the binomial regression models with log or identity link functions, these models often fail to provide the Liang–Zeger-type GEE estimates because the ranges of values for the link functions can be outside of [0, 1]. To circumvent the computational issues, we propose using the Poisson GEE model for estimating the causal risk ratio and the Gaussian GEE model for estimating the causal risk difference within the MSM-GEE framework. To evaluate the effectiveness of these methods, we conducted simulation studies across a total of 64 scenarios designed to emulate practical situations and assessed their practical effectiveness. In addition, we applied the proposed methods to a real-world longitudinal data from the Japan Epidemiology Collaboration on Occupational Health (J-ECOH) study.

Results: In the simulation studies, the proposed methods provided unbiased estimates of the causal risk ratio and risk difference, consistently. Importantly, these methods remain effective for longitudinal data in which the frequency of outcome events exceeds 10%. In the application to the J-ECOH study, the proposed methods also clearly showed their practical effectiveness.

Conclusion: The proposed methods are well suited for providing interpretable estimators for causal parameters in longitudinal clinical and epidemiological studies, even in scenarios where event frequency is high.Conclusion: In conclusion, our work, rooted in MBFs and the orthogonalization strategy by Robinson, presents a first holistic approach to address the challenging problem of HTE estimation in observational studies for complex outcomes beyond mean regression. Extensions to multivariate outcomes, the application of cross-fitting, and a sensitivity analysis to violations of underlying assumptions offer exciting opportunities for further research.

POSTER SESSION A-06:
Causal Inference

P-A06-10

Estimation of the Average Treatment Effect (ATE) in causal survival: Comparison, applications and practical recommendations

Charlotte Voinot[1,2], Julie Josse[1]
1 Inria - Premedical, Montpellier, France
2 Sanofi - R&D, Gentilly, France

Background: Estimating the Average Treatment Effect (ATE) is one of the fundamental measures in causal inference, aimed at assessing the causal impact of a treatment on an outcome variable. Causal survival analysis is at the heart of this approach, seeking to evaluate the effect of a treatment on patient survival over time. However, despite the abundance of literature on causal survival, the use of Cox methods remains predominant for assessing this effect.

Objective: The main objective of this research is to estimate the causal effect of a treatment using survival data which not necessarily derived from randomized trials. Its main aim is to provide users with practical recommendations in the face of the multitude of information available, and to highlight the advantages and differences compared with the classic correlation approaches still widely used such as Hazard ratio to measure the impact of a treatment.

Method: To this end, we will begin by presenting the state of the art in causal survival methods, describing identifiable assumptions and the main estimators, including weighting, regression and triply/doubly robust approaches. Then, an extensive simulation study will then be carried out to compare the different estimators, their preferred regimes and illustrate their theoretical properties on finite sample sizes. Finally, we will examine how the addition of certain variables in the censoring, survival or treatment models can impact the variance of the estimators.

Results & Conclusion: Results will be discussed with a particular focus on the validity of the estimators and its robustness to misspecification on finite sample sizes simulated datasets for practical recommendations in non-randomized settings.

POSTER SESSION A-06:
Causal Inference

P-A06-11

Estimating causal effect on count outcome with excess zeros in randomised clinical trials subject to some degree of noncompliance

Junxian Zhu, Bee Choo Tai, Jialiang Li, Mark Chan, Mark Richards
National University of Singapore, Singapore, Singapore

In public health and medical research, the information collected based on count outcome is relatively common. In randomised clinical trials (RCTs), patients may become non-compliant due to the side effects of treatment. Although there is an extensive body of literature dedicated to exploring non-compliance for continuous or time to event outcomes, this issue is rarely discussed in the context of count outcomes, especially when a significant number of zeros is reported. In this paper, we propose an extension of the inverse probability weighting (IPW) method to address noncompliance in RCTs involving count data with excess zeros. This extension leverages the information on non-compliance to the randomly assigned treatment by recategorizing it to quantify the degree of non-compliance. This differs from the typical approach where compliance is treated as a binary variable. Under regular conditions, the identification and asymptotic distribution of the proposed method are provided. In numerical studies, we evaluate the statistical performance of our method and compare it to the traditional approaches, such as intention to treat, per protocol, instrumental variable and inverse probability weighting. The proposed method performs well in terms of bias, coverage, mean squared error, power, and Type I error, particularly when dealing with selection bias and partial compliance effect. Finally, we illustrate our novel approach using prospective data from the randomised clinical trials of IMMACULATE and TreVasc.

POSTER SESSION A-06:
Causal Inference

P-A06-12

Mediation analysis with imperfectly defined mediators: A microsimulation experiment with breast cancer survival, socio-economic status, and stage at diagnosis

Alessandro Gasparini[1,2], Keith Humphreys[1], Sarah Booth[3], Mark J. Rutherford[3], Elisavet Syriopoulou[1]
1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
2 Red Door Analytics AB, Stockholm, Sweden
3 Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom

Introduction: Mediation analysis methods in survival analysis are used to explore whether the observed association between an exposure and survival time is partly driven by a third variable, named a mediator. One application of mediation analysis in cancer epidemiology is for studying whether survival differences between socio-economic groups can be explained by tumour stage at diagnosis acting as a mediator. Commonly used stages at diagnosis (i.e., categories) may be too crude to fully capture the heterogeneity between patients, resulting in significant leftover within-stage variation. In this study, we explore the impact of ill-defined mediators, such as stage at diagnosis, on the estimation of mediated effects.

Methods: Using microsimulation methods we generated birth cohorts of women according to a breast cancer natural history model developed in Swedish settings. We independently simulated socio-economic status (SES) and imposed a screening programme mimicking that used in Sweden. We then considered four scenarios, assuming that between socioeconomic groups 1) there are no differences in symptomatic detection rates or screening attendance probabilities, 2) there are different rates of symptomatic detection, 3) there are different probabilities of attending screening, or 4) there are both. Nonetheless, across all scenarios, data was simulated assuming no direct effect of SES on breast cancer survival. We performed a mediation analysis under each scenario, adjusting the survival models for age, grade, SES, and stage at diagnosis; in some analyses we also adjusted for mode of detection, tumour size, both, or used a more detailed stage definition.

Results: We showed that using a crude stage definition resulted in biased estimates of the direct effect of SES on survival, with a survival advantage of up to 4% even though the true, simulated direct effect was zero. A more detailed stage definition reduced but did not fully eliminate this bias. Adjusting for mode of detection did not eliminate the bias, but adjusting for tumour size significantly improved the results.

Conclusion: Our results suggest that performing naive mediation analyses with stage at diagnosis as the mediator may lead to significant bias, as tumour stage is often too crude to fully capture the heterogeneity between cancer patients. It is thus fundamentally important to consider the definition of categorical mediators (or adjust for additional characteristics), ensuring that there is no significant unexplained heterogeneity. Failure to do so may lead to biased mediation analyses that should be interpreted with caution.

POSTER SESSION A-06:
Causal Inference

P-A06-13

Causal mediation analysis for a survival outcome with longitudinal mediators, time-varying confounders and left truncation

Katherine Holdsworth[1], Stijn Vansteelandt[2], Diana Bilton[3], Nicholas Simmonds[3,4], Ruth Keogh[1]
1 London School of Hygiene & Tropical Medicine, United Kingdom
2 Ghent University, Ghent, Belgium
3 Imperial College, National Heart and Lung Institute, London, United Kingdom
4 Royal Brompton Hospital, Adult Cystic Fibrosis Centre, London, United Kingdom

Background: Among people with cystic fibrosis (CF), it is well established that females have worse survival than males (estimated median survival age in the UK in 2022 was 54.7 for females, and 57.5 for males). However, the mechanisms underlying this are not well understood. We used causal mediation analyses to investigate the extent to which the effect of sex on survival is mediated through the trajectories of different markers of disease progression in CF. We used data from the UK CF Registry, which has collected longitudinal measures of health status and mortality data on over 11,000 individuals since 1996. Some individuals are observed from birth, however a specific challenge is that there is left-truncation due to individuals joining the Registry in 1996 at older ages. A further challenge is time-varying mediator-outcome confounding.

Methods: The focus was on estimating the counterfactual survival curve for females if the trajectory of the longitudinal mediator in females were shifted to be similar to males. We used the path-specific effects approach [Vansteelandt et al, 2019], which involves fitting a series of regression models for the mortality hazard, and the longitudinal mediator and confounding processes. We also estimated the mediated proportion (MP), defined as the percentage by which the difference between the sexes in survival probability up to a given age would change if the mediator trajectory in females shifted to that of males. To make use of individuals not observed from birth we had to handle left-truncated survival times, such that the entry time depends on birth year, and also the lack of longitudinal data prior to that time. This was done by controlling for birth year and allowing the mortality hazard at a given age to depend on the mediator and confounder history only through more recent values.

Results: With lung health trajectory as the mediator we found an estimated MP at age 30 of 41.1% (95% CI 16.8-86.5%), i.e. at age 30, the female survival probability would be 41.1% closer to that of males if their lung health trajectory had been shifted.

Conclusion: We have shown how left-truncated data can be used in longitudinal mediation analysis for survival outcomes. Several factors were found to mediate the effect of sex on survival in CF, the most important being lung health. Vansteelandt et al. Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. Stat. Med. 2019; 38: 4828–4840.

POSTER SESSION A-06:
Causal Inference

P-A06-14

Estimating correlations when inferring the causal direction between two traits in genetic association studies

Sharon Lutz[1,2], Christoph Lange[2]
1 Harvard Medical School, Boston, United States
2 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, United States

Background: In genetic association studies, Mendelian Randomization (MR) has gained in popularity as an approach to assess the causal relationship between an exposure and outcome using single nucleotide polymorphisms (SNPs) as instrument variables for the exposure. Recently, MR approaches have been extended to infer the effect direction between two phenotypes.

Methods: The causal direction (CD) methods aim to establish both the existence and the direction of a causal relationship. These methods include the causal direction-ratio (CD-Ratio), causal direction-GLS (CD-GLS), causal direction-Egger (CD-Egger), and constrained maximum likelihood approaches (CD-cML, and MR-cML). In order to infer the effect direction between two phenotypes (denoted phenotype 1 and phenotype 2), these approaches first estimate the correlation of the SNP and phenotype 1 and the SNP and phenotype 2 using summary statistics from regression analyses. However, the formula used to estimate the correlations may not be valid depending on the type of model used to obtain the summary statistics (i.e., simple vs multiple linear regressions). Through simulation studies, we assess/quantify the impact of the correlation estimates on the results of the CD methods in the presence of pleiotropy and unmeasured confounding.

Results: For simulations when pleiotropy was generated and there is a causal effect of phenotype 1 on phenotype 2, the results of the CD ratio approach differed if the estimate of the correlations were based on a simple or multiple linear regression. For simulations when unmeasured confounding was generated and there is a no causal effect of phenotype 1 on phenotype 2, the results of the CDcML and MRcML approaches differed if the estimate of the correlations were based on a simple or multiple linear regression.

Conclusions: Estimation of the correlation of the SNP and phenotype 1 and the SNP and phenotype 2 using summary statistics from different regression analyses may impact the results of the CD methods in the presence of pleiotropy or unmeasured confounding.

POSTER SESSION A-06:
Causal Inference

P-A06-15

Mediation analysis with multiple mediators using the relative survival framework: An example exploring socioeconomic differences in cancer survival

Elisavet Syriopoulou, Therese Andersson
Karolinska Institutet, Sweden

Introduction: Mediation analysis methods can be applied to understand how different pathways influence the relationship between an exposure and outcome. For instance, it helps investigate if survival differences in cancer between socioeconomic groups are partly explained by stage at diagnosis. Cancer disparities may be the result of complex mechanisms that involve both cancer-related and other-cause mortality differences, making it difficult to identify the causing factors. A recent extension of mediation analysis to the relative survival framework allows us to focus on factors that account for cancer-related differences instead of all-cause differences. However, the current methodology does not accommodate settings with multiple mediators affecting each other. This study aims to fill this gap and address such complexities that are common in cancer epidemiology.

Methods: We build on existing methodology for randomised interventional analogues of the direct and indirect effects of the exposure that require no cross-world assumptions, extending it to the relative survival framework. The proposed effect decomposition allows to quantify effects mediated through each of the mediators of interest separately, but also through the mediators' dependence. We illustrate the method using an example of socioeconomic survival differences among colon cancer patients. The mediators of interest in this example are the stage at diagnosis and treatment allocation. We discuss identification assumptions and provide an estimation algorithm using regression standardisation. A flexible parametric relative survival model is fitted for the outcome and separate models are required for the mediators. Confidence intervals are obtained via bootstrapping.

Results: We evaluate the impact on all-cause survival under hypothetical interventions aimed at cancer-related mortality rates that are assumed to have no impact on other-cause mortality rates. We partition the total difference in all-cause survival between socioeconomic groups into: i) differences mediated by treatment allocation, ii) differences due to stage at diagnosis, iii) interaction between stage and treatment, and iv) the direct effect of socioeconomic position. For further exploration of the survival differences, the potential impact of interventions can be quantified as the number of avoidable deaths within a time frame e.g., by shifting the stage distribution of the most deprived to match the least deprived while keeping other-cause (background) mortality constant.

Conclusion: The proposed method has the advantage of allowing us to focus on cancer-related differences, the underlying determinants of which may be easier to identify in comparison with all-cause differences and helps improve our understanding of cancer inequalities.

POSTER SESSION A-06:
Causal Inference

P-A06-16

Biomarkers in clinical drug development: A mediation analysis approach

Julia Krzykalla, Elvira Erhardt, Carina Ittrich
Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

Background: Biomarkers are gaining importance in clinical drug development as to characterize the mode of action of new compounds and to detect early signs of efficacy and safety. The anticipated translation of short term treatment effects on biomarkers to clinically meaningful efficacy in established long term clinical endpoints will be an indicator of biomarker integration into early decision making. To this end, it is not sufficient to demonstrate a prognostic value of the biomarkers, investigation of their associations with clinical endpoints is important.

Methods: Establishing this disease link can be achieved by using mediation analysis techniques. Causal mediation analysis aims to characterize an exposure's effect on an outcome and to quantify the indirect effect that acts through a given mediator or a group of mediators of interest. When investigating biomarkers as potential mediators and their indirect effect makes up a substantial proportion of the total treatment effect, these biomarkers may qualify as reliable parameters for decision making in early clinical development.
To gain a better understanding how certain factors influence the identifiability of biomarkers as potential mediators, we explore the associations between treatment, clinical endpoint, and biomarkers for varying numbers of included biomarkers, effect sizes and interdependencies, as well as different sample sizes. The simulation study is based on real clinical trial data.

Results: Joint modelling including all biomarkers results in unbiased estimates for all scenarios. If biomarkers are uncorrelated, separate evaluation of biomarkers also yields unbiased estimation of the proportion of effect mediated since the indirect effect of the omitted biomarker(s) is attributed to the direct effect. However, if biomarkers are correlated, estimation of the proportion of effect mediated suffers from omitted-variable bias. The amount and direction of bias varies with the direction and strength of correlation, as well as with the effect size of the omitted biomarker(s) on the outcome and the relation of variabilities of omitted and included biomarkers.

Conclusion: Omitted-variable bias might be less problematic if this leads to a reasonable representative being selected out of a group of positively correlated biomarkers. But negative correlation might dilute the effect of the mediator and thus hinder its identification.

POSTER SESSION A-06:
Causal Inference

P-A06-17

Comparison of modeling approaches for continuous covariates in explanatory modeling: Does it matter which methods are used?

Nilufar Akbari[1], Georg Heinze[2], Daniela Dunkler[2]
1 Charité - Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology, Berlin, Germany
2 Section for Clinical Biometrics, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria

Determining the appropriate functional form of continuous covariates is crucial in explanatory multivariable model building, as misspecification can lead to biased estimates of the estimand of interest. Common approaches include linear modeling, categorization, fractional polynomials, and several types of splines, like natural splines or P-splines. The assumption of linear effects is often the default choice and unquestioned due to a lack of knowledge rather than careful consideration. This is additionally reinforced by statistical programs where linear effects are the default option for continuous variables. Categorization is often based on cutpoints derived from clinical thresholds or certain quantiles. Flexible modeling techniques, such as fractional polynomials and splines, can capture nonlinear functional forms. However, these techniques are often underused despite being available in most statistical software. Reasons for selecting a specific approach are rarely explained.

We quantified by means of a case study, how much the estimand of interest varies depending on how potentially nonlinear associations of confounders are accounted for. We used Austrian registry data of people with acute myocardial infarction who received either standard care or participated in a rehabilitation program as recommended by guidelines. The outcomes of interest were participation in cardiac rehabilitation and time to mortality. The corresponding estimands were the association of age and sex with rehabilitation prevalence and the effect of cardiac rehabilitation on total mortality within four years of observation time, were age is an important confounder. The abovementioned modeling approaches for continuous covariates with different parameter settings and different degrees of freedom were applied and compared.

We found that different specifications of the association of the continuous covariates or rather confounders can lead to systematic differences in the effect estimate. Both estimands showed that ignoring or overly simplifying the nonlinear functional form of age lead to different effect estimates. For the associations of age and sex with rehabilitation participation, applying different flexible modeling approaches resulted in different prevalences in age ranges where few observations and events existed, apart from that, the results were similar. The effects of participation in cardiac rehabilitation on mortality were similar between different flexible modeling approaches irrespective of number of degrees of freedom.

To minimize the danger of residual confounding in the estimation of causal effects caused by inappropriate specification of functional forms, we recommend the use of flexible modeling approaches for continuous covariates.

POSTER SESSION A-06:
Causal Inference

P-A06-18 LB

Accounting for the between-study heterogeneity when selecting instrumental variables for two-sample Mendelian randomisation studies

Fabiola Del Greco Miglianico[1], Giulia Pontali[2], Luisa Foco[1], Cristian Pattaro[1]
1 Institute for Biomedicine (affiliated to the University of Lübeck), Eurac Research, Bolzano-Bozen, Italy
2 Fondazione Human Technopole, Milan, Italy

Introduction: Two-sample Mendelian randomisation (2SMR) is widely used to explore causation of modifiable exposures on outcomes, thanks to the exploitation of publicly available data from genome-wide association study meta-analyses (GWAMAs).
The instrumental variable (IVs) selection remains the key step to guarantee reliable and robust results, mainly because MR methods developed for summary genetic data are often based on simplifying assumptions hard to be completely rebutted.
Despite formally satisfying all modeling assumptions, whether between-study heterogeneity in GWAMAs has an effect on the quality of the identified IVs is unknown. We investigated this issue by simulations mirroring a real data example.

Methods: We conducted a metabolome-wide 2SMR to assess the causal effect of 129 metabolites on Parkinson´s disease [1]. For each metabolites a GWAMA from 7 cohorts was available [2], where the between-study heterogeneity was quantified by the I2 index [3]. We assessed the effect of removing IVs with elevated I2, both before and after the linkage disequilibrium clumping (LD) step. We evaluated the consistency of the results in terms of reduction of false positives and power gain that we could have had using an optimal set of IVs with an higher homogeneity.
Mimicking the illustrative real data example by simulating summary genetic data under random scenarios of heterogeneity among exposure studies, we estimated power, type I error and bias, varying the IVs number and the MR causal effect, fixing the IVs strength, and considering only independent instruments.

Results: Results of both simulations and real data example suggested that the Introduction: of the I2 pruning step could increase the IVs quality leading to higher power with less biased MR estimates. But, the reduction of the IVs number could cause some power loss. Conducting the I2 pruning before the LD clumping step increases the analysis power.

Conclusions: The I2 pruning should be included in the IVs selection to obtain more reliable MR results.

References: *[1] Draisma HHM, Pool R, Kobl M, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015;6:7208. [2] Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18(12):1091-1102. [3] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557–560.*

POSTER SESSION A-06:
Causal Inference

P-A06-19

Implications of non-compliance in the experimental arm on treatment effect metrics in randomised controlled trials: A methodological assessment

Tianyuan Gu[1], Jialiang Li[2], Yu Yang Soon[3], Joseph Wee[4,5], Bee-Choo Tai[1,6]

1 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
2 Department of Statistics and Data Science, National University of Singapore, Singapore
3 Department of Radiation Oncology, National University Cancer Institute, National University Hospital, Singapore, Singapore
4 Division of Radiation Oncology, National Cancer Centre Singapore, Singapore
5 Duke-NUS, Graduate Medical School, Singapore, Singapore
6 Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Background: In randomised controlled trials (RCTs), participant often fail to comply with the prescribed treatment owing to severe side effects or operational challenges. Intention-to-treat (ITT) and per-protocol (PP) analyses may underestimate the treatment effect in cases of severe or non-random non-compliance. Inverse Probability of Censoring Weighting (IPCW) method and Rank Preserving Structural Failure Time model (RPSFTM) provide more sophisticated approaches to address these problems. However, the literature lacks a broader definition of random and non-random non-compliance, particularly concerning dose-response and treatment cycles. It is uncertain which method yields more appropriate and reliable estimates when non-compliance affects only the experimental arm in RCTs employing different treatment effect metrics.

Methods: Survival times were simulated using the Weibull distribution based on a balanced design with sample sizes of 200 and 500, and true treatment effect of 0.3, 0.55 and 0.8. Low (20%) and high (60%) censoring patterns were considered. Random and non-random non-compliance rates of 40% and 70% were simulated for the experimental arm. Hazard ratio (HR), ratio of restricted mean survival time (rRMST), and milestone survival difference (MSD) at 3-year were estimated and evaluated using ITT, PP, IPCW and RPSFTM (for both re-censoring and no re-censoring, Log-rank and Weibull tests). These methods were examined in terms of bias, coverage and power, and compared using data from a RCT of patients with nasopharyngeal cancer (SQNP01).

Results: The RPSFTM-related models produced smaller bias, and IPCW performed well in scenarios with moderate treatment effect, high non-compliance and low censoring across all metrics. The RMST-based estimates had reduced power than HR-based estimates under identical simulation conditions. PP was susceptible to selection bias with greater bias, and reduced coverage and power. When applied to SQNP01 data, RPSFTM-related models had the largest treatment effect in terms of HR, whereas IPCW had the largest treatment effect for other metrics.

Conclusion: The RPSFTM-related models and IPCW methods were comparable and had small bias. However, it is crucial to meticulously consider the selection of metric to quantify treatment effect and generate suitable models on a case-by-case basis when addressing treatment non-compliance in RCTs, regardless of selection bias.

POSTER SESSION A-06:
Causal Inference

P-A06-20

Assessing patient subgroups specific safety and effectiveness of oral type 2 diabetes treatments using the Local Instrumental Variable method

Laura Güdemann, Trevelyan McKinley, Angus G Jones, Andrew P McGovern, Beverley M Shields, John M Dennis, Jack Bowden
University of Exeter, Exeter, United Kingdom

Background/Introduction: Type 2 diabetes (T2D) management guidelines call for individualized treatment strategies and are mostly based on clinical trial evidence and expert opinions. Adults over the age of 70 years are a large and important patient group, but evidence-based treatment recommendations are lacking due to strict exclusion criteria of clinical trials. Observational studies using for example electronic health care records have the potential to provide additional evidence in this patient group. The quality of observational evidence depends on the data source and choice of suitable analysis methods which can address unmeasured confounding bias inherit due to the lack of treatment allocation randomization. We conducted a causal analysis using the Local Instrumental Variable (LIV) approach to assess the effectiveness of SGLT2-inhibitors regarding the reduction of blood glucose and weight, as well as their safety concerning potentially important adverse events for older people with type 2 diabetes such as genital infections, dehydration, and other osmotic symptoms.

Methods: Hospital-linked UK primary care data (Clinical Practice Research Datalink, 2013-2020) were used to compare adverse events and effectiveness in individuals initiating SGLT2-inhibitors (n=77,229) versus DPP4-inhibitors (n=109,606). We used the LIV approach to estimate person-centered treatment effects and conditional treatment effects for age-stratified subgroups. This IV analysis was based on a continuous provider preference-based Instrument calculated as the proportion of previous prescriptions and allows the estimation of treatment effects addressing unmeasured confounding and treatment effect heterogeneity based on unmeasured prognostic factors.

Results: Results of the causal analysis showed a greater glycaemic efficacy with SGLT2-inhibitors compared to DPP4-inhibitors in both age groups with on average 4.9mmol/mol (95%CI 4.8;4.9) more reduction of HbA1c in younger adults and 2.5mmol/mol (95%CI 2.4;2.6) in older adults. Weight reduction after SGLT2-inhibitor initiation is confirmed from our results for both age-stratified populations for younger adults on average 2.6kg (95%CI 2.5;2.6) and 2.9kg (95%CI 2.8;2.59) for older adults. Incidence risk ratio (IRR) estimated for genital infection were elevated for both subgroups and very similar. Rounded results for both groups were IRR: 1.63 (95%CI 1.62; 1.63). No increased risk of other osmotic symptoms was detected in both age groups.

Conclusion: The Local IV approach has great potential for high quality observational evidence in precision medicine by addressing unmeasured confounding and possible heterogeneity in treatment effects. This causal analysis extends trial evidence and suggests that SGLT2-inhibitors are effective and generally safe in older adults but increase risk of genital infections.

POSTER SESSION A-06:
Causal Inference

P-A06-21

Dealing with highly imbalanced groups from surveys: Life satisfaction among Italian native, migrant and international adopted adolescents

Emanuele Koumantakis[1,2,3], Andrea Ricotti[4], Laura Ferrari[5], Sonia Ranieri[5], Rosa Regina Rosnati[5], Lorena Charrier[1], Paola Berchialla[2]

1 Department of Public Health and Pediatrics, University of Torino, Torino, Italy
2 Department of Clinical and Biological Sciences, University of Torino, Orbassano, Italy
3 Post Graduate School of Medical Statistics, University of Torino, Torino, Italy
4 Clinical Trial Unit - Hospital AO Ordine Mauriziano of Torino, Torino, Italy
5 Family Studies and Research University Centre, Department of Psychology, Catholic University, Milano, Italy

Introduction: Imbalanced data with unequal sample sizes poses a common challenge in drawing conclusions in both experimental and observational studies, including surveys. Matching techniques are often employed to address this issue. However, these methods often reduce the size of the majority group, leading to considerable variance. This study aims to evaluate potential techniques that achieve balance between groups without excessively shrinking the sample size, using real survey data.

Methods: Italian participants from two international surveys were included: (i) Italian students from the Health Behaviour In School-aged Children (HBSC) 2018 survey, categorized as natives or migrants based on parental birthplace, and (ii) internationally adopted individuals surveyed by the Group for European Adoption Research (GEAR). Demographic variables including gender and age were collected for all participants. The focus of the study was life satisfaction (LS) assessed by the Cantril ladder and dichotomised in low (from 0 to 8) and high (9-10). The primary aim was to investigate differences in LS among Italian natives, migrants, and adopted individuals. Three strategies were used: (i) a multivariable logistic regression model was employed, adjusting for age and gender; (ii), coarsened exact matching (CEM) on age and gender to achieve balance between groups [1] followed by logistic regression with and without weighting; (iii) multinomial propensity scores (MNPS) considering age and gender [2], followed by weighted logistic regression with and without trimming. Significance level was set at 0.05.

Results: The analysis included 48,227 Italian natives and 10,341 migrants from the Italian HBSC 2018 survey, and 178 adopted adolescents from the GEAR survey. Comparing moderately unequal groups, such as migrants and Italian natives, yielded similar estimates across different techniques. Conversely, comparing highly unequal groups, like adopted and Italian natives, resulted in less stable estimates. MNPS without trimming had the smallest standard error.

Conclusion: The methodologies applied showed that highly imbalanced groups, compared to moderately imbalanced groups, yielded divergent results that were dependent on the model used. A simulation study could help better investigate the bias introduced by different balancing techniques when groups are highly imbalanced.

References: *[1] Iacus SM, King G, Porro G. Causal Inference without Balance Checking: Coarsened Exact Matching. Political Analysis. 2012;20(1):1–24. doi:10.1093/pan/mpr013. [2] Griffin BA, Ridgeway G, Morral AR, et al. Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) Website. Santa Monica, CA: RAND Corporation, 2014. http://www.rand.org/statistics/twang.*

POSTER SESSION A-06:
Causal Inference

P-A06-22 Comparing covariates balancing methods, a simulation study

Etienne Peyrot, Raphaël Porcher, François Petit
Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), Paris, France

Intro: Comparing treatment effects is a significant aim in epidemiology. One measure commonly used to perform this comparison is the average treatment effect. It aims to quantify the causal relation between a treatment and an outcome. Randomized control trials (RCT) are considered the gold standard for this estimation. Nonetheless, conducting a RCT is not always practical because of cost, time, logistic constraints, etc. For that reason, using observational data is increasingly popular, but this comes with many challenges. One of those challenges is the presence of confounders, which can bias an estimation. Balancing methods were created to alleviate this issue. Recently, several balancing methods have been introduced, but there are currently no real guidelines on which one to use. To this aim, we compare one classical method to three more recent ones using simulated data.

Method: To define an estimation method for an average treatment effect using observational data, one must specify the balancing method, the estimator, and the types of regressor, if needed, for both the balancing and estimation tasks. We compare four balancing methods: Inverse Probability of Treatment Weighting (IPTW), Energy Balancing (EB), Kernel Optimal Matching (KOM) and covariates balancing by Tailored Loss Function (TLF); with three classical estimators: average weighting, double robust scheme, linear regression coefficient; using if required three types of regressor: a correctly specified logistic regression, misspecified logistic regression and a random forest; for a total of 24 estimation methods. Those methods are tested on simulated data with 36 scenarios in which sample size, treatment probability, and confounding level vary.

Results: Results show that the choice of the balancing method matters as much as the choice of the estimator. The type of regressor had a limited impact when estimating the ATE; it was more significant when estimating the ATT. EB demonstrates the best performance overall, TLF yields poor estimation in most settings, and KOM as well as IPTW give good estimations in simple settings. However, the treatment rarity and the confounding level heavily impact these last two methods.

Conclusion: In most scenarios, opting for a double robust estimator proves to be the optimal choice for estimating both ATE and ATT, irrespective of the balancing method employed. Standard methods, like the IPTW with a logistic regression, give good results in settings where the probability of treatment is not too far from 1/2. However, EB with linear regression estimator seems to be the best for settings with rarer treatment according to our simulation.

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-01 Joint modelling of ordinal repeated measurements and multi-state survival data: A simulation study

Behnaz Alafchi[1], Ghodratollah Roshanaei[1], Hossein Mahjub[2], Leily Tapak[1]
1 Department of Biostatistics, School of Public Health, Modeling Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
2 Research Center for Health Sciences, Department of Biostatistics, Faculty of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

In many follow-up studies, a longitudinal response variable is measured over time until the occurrence of an event, which could cause non-ignorable missing data. Moreover, in such studies, a patient may experience a sequence of clinical progression events instead of the occurrence of an event of interest. Hence, the progression of disease should be considered as a multi-state process. Joint models have been increasingly used for the combined analysis of longitudinal and survival data in recent years. However, some few studies have extended the classical joint modelling framework to multi-state survival data [1,2,3]. Furthermore, all of these studies focused on continuous longitudinal responses and there has been no study that extended the joint models to simultaneously model longitudinal ordinal data and multi-state data. Here, we introduce a joint model for ordinal repeated measurements and multi-state survival data. The proposed joint model consists of two sub-models including a mixed-effects proportional odds model for ordinal repeated measurements and a multi-state model with transition-specific proportional intensities for the transitions times between the states. These two sub-models are linked through shared random effects. The proposed joint model was estimated by the maximum likelihood method for a piecewise constant baseline intensity function. The model has been implemented using R. The full likelihood does not have a closed form as the integral with respect to the random effects does not have an analytical solution. The integral is approximated by employing the Gauss–Hermite quadrature. Moreover, the numerical method is used to estimate the parameters using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm of maximization, which is an option in optim function of state package in R. Also, a simulation study was conducted to evaluate the performance of the proposed joint model. In summary, the empirical simulation studies demonstrate a considerable advantage for the proposed joint model regarding the coefficients, biases, and the mean square errors. The simulation results revealed

that the superiority of the joint model in comparison to the separate models is more noticeable in larger value of the association parameters which explain the correlation between the two longitudinal and multi-state outcomes.

References: *[1] Dantan E, Joly P, Dartigues J-F, Jacqmin-Gadda H. Joint model with latent state for longitudinal and multistate data. Biostatistics. 2011;12(4):723-36. [2] Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin‐Gadda H, Proust‐Lima C. Joint modelling of longitudinal and multi‐state processes: application to clinical progressions in prostate cancer. Statistics in medicine. 2016;35(22):3933-48.2. [3] Mwanyekange J, Mwalili SM, Ngesa O. Bayesian Joint Models for Longitudinal and Multi-state Survival Data. International Journal of Statistics and Probability. 2019;8(2).*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-02 Simultaneous modelling of ordinal repeated measurements and multi-state data in the presence of cure fraction: A joint model

Elaheh Talebi Ghane, Behnaz Alafchi
Department of Biostatistics, School of Public Health, Modeling Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Islamic Republic of Iran

In some cohort studies, patients may experience one or more intermediate events (IE) prior to the event of interest. Additionally, there may be a cured fraction of population who will never experience the IE. For instance, progression of cancer prior to death in cancer patients who underwent tumor removal, in which a patient may never experience progression of cancer after resection and be cured for this IE. A small number of studies have developed multi-state models which incorporate latent cured state to model such multi-state data with cured fraction [1,2]. On the other hand, most often in such studies, longitudinal outcomes are collected along with the survival information. These longitudinal variables often contain information about the disease progression and they can be used to monitor patients effectively. However, most of the time these valuable variables are excluded from the analysis. Although some few studies have developed joint models for simultaneous modeling of longitudinal and multi-state data [3,4], there is no study that jointly model the longitudinal data and multi-state data with a cured fraction. Here, we focused on the longitudinal ordinal outcomes and introduce a joint model for ordinal repeated measurements and multi-state data in the presence of cure fraction. The proposed joint model is decomposed into two sub-models: a mixed proportional odds sub-model for longitudinal ordinal data, and a transition-specific proportional intensities model which incorporate a latent cured fraction for the event history data. The two sub-models are linked through a function of the shared random effects. We have four states in the multi-state model: state 1 to be alive and cured of disease, state 2 to be alive and not being cured of disease, state 3 to be alive and experience the intermediate event (IE), and state 4 to be dead. Cured subjects are defined as subjects who will never experience IE. Subjects with an observed IE are known to be not cured of disease, but all other subjects have unknown cure status and the probability of being cure is modeled using a logistic function. The joint model was estimated by the maximum likelihood method and the model has been implemented using R. Also, a simulation study was conducted to evaluate the performance of the proposed joint model. In summary, the empirical simulation studies demonstrate a considerable advantage for the proposed joint model regarding the coefficients, biases, the mean square errors and the Akaike information criterion.

References: *[1] Dantan E, Joly P, Dartigues J-F, Jacqmin-Gadda H. Joint model with latent state for longitudinal and multistate data. Biostatistics. 2011;12(4):723-36. [2] Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust‐Lima C. Joint modelling of longitudinal and multi‐state processes: application to clinical progressions in prostate cancer. Statistics in medicine. 2016;35(22):3933-48.2. [3] Taylor, J M, Conlon, A S. Multistate models for colon cancer recurrence and death with a cured fraction. 2013. [4] Beesley, L J, Taylor, J M. EM algorithms for fitting multistate cure models. Biostatistics. 2019; 20(3), 416-432.*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-03 Swift inference and dynamic risk predictions for multivariate joint models using INLAjoint

Denis Rustand
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Background/Introduction: Clinical research often requires the simultaneous study of longitudinal and survival data. Joint models, which can combine these two types of data, are essential tools in this context. A joint model involves multiple regression submodels (one for each longitudinal/survival outcome) usually linked together through correlated or shared random effects. This makes their estimation process rather complex, time-consuming, and sometimes even unfeasible, especially when dealing with many outcomes. Methods In this context, we introduce INLAjoint (Rustand et al., 2024), a user-friendly and flexible R package designed to leverage the Integrated Nested Laplace Approximation (INLA) method from the INLA R package, renowned for its computational efficiency and speed (van Niekerk et al., 2023). INLAjoint can handle various model formulations and simplifies the application of INLA to fit joint models, ensuring fast and accurate parameter estimation.

Results/Conclusion: Our simulation studies show that INLA reduces the computation time substantially compared to alternative strategies such as Bayesian inference via Markov chain Monte Carlo, without compromising on accuracy (Rustand et al., 2023). We illustrate the use of INLAjoint to the famous PBC dataset, including follow-up data from 312 patients with primary biliary cirrhosis. We fitted a joint model with 9 outcomes (7 longitudinal markers and 2 competing risks of events), and we show how dynamic risk predictions (i.e., incorporate changes in the longitudinal outcomes to update future risk predictions) can easily be computed from this joint model. This makes INLAjoint a valuable tool for analyzing complex health data.

References: *[1] Janet van Niekerk, Elias T. Krainski, Denis Rustand, and Håvard Rue. A new avenue for Bayesian inference with INLA. Computational Statistics & Data Analysis 181, 107692, 2023. [2] Denis Rustand, Janet van Niekerk, Elias T. Krainski, Håvard Rue, and Cécile Proust-Lima. Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. Biostatistics (Oxford, England), kxad019, 2023. [3] Denis Rustand, Janet van Niekerk, Elias T. Krainski, and Håvard Rue. Joint Modeling of Multivariate Longitudinal and Survival Outcomes with the R package INLAjoint. arXiv preprint:2402.08335, 2024.*

P-A07-04 Joint modelling of longitudinal covariates and partly-interval censored survival data - A penalised likelihood approach

Jun Ma
Macquarie University, Sydney, Australia

This talk will focus on a joint modelling of longitudinal covariates and partly interval censored time-to- event data. Longitudinal time-varying covariates play a crucial role in achieving accurate dynamic predictions using a survival regression model. However, these covariates are often measured at limited time points and may contain measurement errors. Moreover, they are usually specific to each individual. On the other hand, the event times of interest are often interval-censored. Accounting for all these factors is essential when constructing a survival model. We will present a new approach for joint modelling of the longitudinal time-varying covariates and the time-to-event Cox model, where the latter is subject to interval censoring. We will develop a novel maximum penalized likelihood approach for estimation of all the model parameters including the random effects. A profile likelihood is used to obtain the covariance matrix of the estimated parameters.

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-05

Investigating the role of longitudinal cytokine concentrations in HIV acquisition among women: A joint modelling approach

Farzana Osman[1,2], Nobuhle Mchunu[3], Temesgen Zewotir[2], Lyle Mckinnon[1,4], Quarraisha Abdool Karim[1,5], Salim S. Abdool Karim[1,5], Nonhlanhla Yende-Zuma[2,3]
1 Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa
2 University of Kwazulu-Natal, Durban, South Africa, 3South African Medical Research Council, Durban, South Africa
4 University of Manitoba, Winnipeg, Canada
5 Department of Epidemiology, Columbia University, New York City, United States

Introduction: Traditionally, researchers have used the time varying Cox regression models to analyse the association between longitudinal data, such as a biomarker and HIV acquisition. Recognizing the limitations of time-varying Cox models in treating longitudinal biomarkers as exogenous covariates, our study opts for joint modelling to predict the association between longitudinal cytokine concentrations and time to HIV infection.

Methods: We studied 774 women aged between 18 and 40 years from the Centre for the AIDS Programme of research in South Africa (CAPRISA) between May 2007 and March 2010. Considering the established link between genital inflammation (defined by ≥ 5 of 9 inflammatory cytokines above the 75th percentile), and the increased risk of HIV acquisition1, we conducted a joint modeling analysis to investigate the effect of longitudinal cytokine concentrations on HIV acquisition. We used proportional hazards regression model to assess the effect of baseline age, study arm and research site on HIV risk and linear mixed effects model to evaluate the effect of age, study arm and research site on log cytokine concentrations over time. We then investigated an appropriate joint model to link the longitudinally measured cytokine biomarkers with the time to HIV infection.

Results: The women's overall median age was 22 years (interquartile range (IQR) 20.0 – 27.0). There was a total of 59 HIV infections from the 774 women over 2139 study visits. The joint model reveals significant associations between specific cytokines and the risk of HIV acquisition. For IL-1B, a unit increase in the log concentration corresponds to a 2.7-fold increase in HIV acquisition risk (HR: 2.73, 95% CI: 1.27-6.0). Similarly, for HGF, a unit increase in the log concentration corresponds to a 3.7-fold increase in HIV acquisition risk (HR: 3.68, 95% CI: 1.03-13.99). While, for IL-15, a unit increase in the log concentration is associated with a decrease in HIV acquisition risk (HR: 0.46, 95% CI: 0.26-0.85). When comparing these findings to those obtained from the time-varying Cox model, smaller estimates were observed from the Cox model.

Conclusion: Joint modelling offers a robust statistical framework for analysing longitudinal cytokine data, effectively addressing challenges such as correlated observations, missing data, and measurement error.

Reference: *[1] J.-A. S. P. Lindi Masson, Lenine J. Liebenberg, Lise Werner, et al, Genital Inflammation and the Risk of HIV Acquisition in Women, Clinical Infectious Diseases 2015 Vol. 61 Issue 2 Pages 260-269.*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-06

Joint modelling of non-time to event longitudinal data: A systematic review of methodology and applications

Rehema Ouko[1], Mavuto Mukaka[2], Eric Ohuma[1]
1 London School of Hygiene & Tropical Medicine, London, United Kingdom
2 University of Oxford, Oxford, United Kingdom

Background: In recent years, the application of joint modelling for longitudinal data has increased, especially for time-to-event data. Joint modelling has been shown to; a) reduce bias in parameter estimation, b) increase the efficiency of statistical inference by incorporating the correlation between measurements, and c) allow the borrowing of information in cases where there is missing data for one variable of interest. Most joint modelling methods and applications are on time-to-event data, focusing primarily on one continuous and one binary variable. However, in practice, there are two or more time-varying variables, e.g., height and weight associated with different types of outcomes ranging from count, ordinal, etc., that are of great interest, and they have not been given much attention in joint modelling. Therefore, in this systematic review, we aim to summarise the current state of joint modelling of non-time-to-event longitudinal data to draw recommendations for future research.

Methods: We will review current methodologies of joint modelling for non-time-to-event data for both univariable and multivariable longitudinal data. The search will include five main databases: Embase, Medline, Scopus, PubMed, and Web of Science. We will extract information including the field of application, i.e., medicine and social sciences, statistical methods and approaches used, such as modelling assumptions and distributions, the structure of association, estimation methods, and the software used for implementation, and also elucidate the gaps and challenges in the application of joint modelling of longitudinal data.

Results: After conducting a thorough search across the five electronic databases, we identified 1912 records that will undergo a screening process. From this screening, we will select eligible studies for further analysis. The approaches to modelling the non-time-to-event longitudinal data will be reported, mimicking the description provided in the methodology section.

Conclusion: Several methodologies for different data types have been proposed, and the value of joint modelling has been established. However, researchers are currently limited in their ability to routinely fit these models. Joint modelling has been proven to be beneficial in producing more accurate dynamic predictions. In conclusion, joint modelling of non-time-to-event longitudinal data is a promising area of research, and we recommend further investigation in this field.

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-07

Quantifying associations between immune cell kinetics and (competing) allo-immunological events after allogeneic stem cell transplantation using joint models: Opportunities and challenges

Edouard Bonneville[1], Eva A. S. Koster[2], Hein Putter[1,3], Johan H. Frederik Falkenburg[2], Constantijn J. M. Halkes[2], Liesbeth C. de Wreede[1,4]
1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
2 Department of Hematology, Leiden University Medical Center, Leiden, Netherlands
3 Mathematical Institute, Leiden University, Leiden, Netherlands
4 DKMS Clinical Trials Unit, Dresden, Germany

Background/Introduction: The successful treatment of haematological malignancies with allogeneic stem cell transplantation (alloSCT) relies on the introduction of donor-derived T-cells. While these T-cells can prevent disease relapse by eliminating malignant cells, they may also target other patient cells, potentially leading to the development of Graft-versus-Host-Disease (GvHD). We sought to model the T-cell kinetics after T-cell depleted alloSCT, and quantify their impact on the development of GvHD or relapse, in a cohort of 166 acute leukaemia patients. The challenge was to build a model, with relatively small data, that accounted for heterogeneity in (measurement error-prone) non-linear individual T-cell trajectories, time-dependent interactions, and post-transplant interventions such as donor lymphocyte infusions (DLI).

Methods: We used joint modelling to investigate the impact of different T-cell subsets (CD4, CD8 or total CD3 counts; outcome of the longitudinal submodel) on the competing risks GvHD, relapse, and other failures (e.g. death) [1]. The effect of a scheduled DLI at 3 months after transplant on both the T-cell kinetics and competing outcomes proved particularly challenging to model, given the typically short interval between DLI and associated GvHD, with limited number of longitudinal measurements. Moreover, not all patients scheduled for 3-month DLI actually received it, therefore an intention-to-treat (ITT) analysis was done by means of a three-way interaction in the longitudinal submodel.

Results: The developed models were able to adequately capture patient-specific non-linear T-cell trajectories, and illustrate the significant opposing effects of T-cell counts on the competing outcomes. In particular, higher (current value) log CD4 counts increased the risk of GvHD (hazard ratio 2.44, 95% CI 1.45-4.12) and decreased the risk of relapse (hazard ratio 0.65, 95% CI 0.45-0.92). Furthermore, the ITT analysis suggested that a scheduled DLI can induce T-cell expansion.

Conclusion: The opposing effects of the CD4 counts on the cause-specific hazards for GvHD and relapse underline the importance of accounting for competing risks in alloSCT. In a setting where longitudinal T-cell measurements are typically analysed using simplistic approaches, joint models show great promise. Nevertheless, more complex extensions (e.g. multivariate longitudinal submodel, and/or multi-state time-to-event submodel) will require carefully navigating the balance between detailed data on relatively few patients, and larger numbers of patients with more limited longitudinal information.

References: *[1] Koster, Eva AS, et al. "Joint models quantify associations between immune cell kinetics and allo-immunological events after allogeneic stem cell transplantation and subsequent donor lymphocyte infusion." Frontiers in Immunology 14 (2023): 1208814.*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-08

Investigating a domain adaptation approach for integrating different measurement instruments in a longitudinal clinical registry

Maren Hackenberg[1,2], Michelle Pfaffenlehner[1,2], Max Behrens[1,2], Astrid Pechmann[3], Janbernd Kirschner[3], Harald Binder[1,2,4]

1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
2 Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany
3 Department of Neuropediatrics and Muscle Disorders, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
4 Centre for Integrative Biological Signalling Studies (CIBSS), University of Freiburg, Freiburg, Germany

Background: In a longitudinal clinical registry, different measurement instruments might have been used for assessing individuals, potentially at different time points. For combining these measurements, we investigate the use of deep learning techniques for obtaining a joint latent representation, to which the items of different measurement instruments are mapped. This corresponds to domain adaptation, an established concept in computer science for image data. Using our proposed approach as an example, we evaluate the potential of domain adaptation in a longitudinal cohort setting with a rather small number of time points. This is motivated by an application with different measurement instruments for assessing motor function skills in a registry of spinal muscular atrophy (SMA) patients.

Methods: We use variational autoencoders for mapping the items of each measurement instrument, at each time point where it is available, to a joint latent representation. In this representation, we model the instruments' embeddings as solutions of a shared ordinary differential equation (ODE), where patient-specific ODE parameters inferred from patients' baseline characteristics. The goodness of fit and complexity of the ODE solutions then allows to judge the measurement instrument mappings. We subsequently explore how alignment can be improved by incorporating corresponding penalty terms into model fitting. To systematically investigate the effect of differences between measurement instruments, we consider several scenarios based on modified SMA data, including scenarios where a mapping should be feasible in principle and scenarios where no perfect mapping is available.

Results: While the ODE fits are seen to be increasingly overshadowed by mis-alignment as the complexity of scenarios increases, some structure is still recovered, even if the availability of measurement instruments depends on patient state. A reasonable mapping is feasible also in the more complex real SMA dataset.

Conclusion: The proposed approach can help to address the issue of patients' measurements taken by different instruments at different times, while dealing with limited numbers of individuals and time points per individual, which are typical constraints of clinical registries. These results indicate that domain adaptation might be more generally useful in statistical modeling for longitudinal registry data.

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-09

Using latent profile analysis to understand burnout in a sample of Greek teachers

Eleni Kalamara[1], Clive Richardson[2]
1 European Union Agency for Asylum, Valetta, Malta
2 Panteion University, Athens, Greece

Background/Introduction: Burnout in teachers and other occupations has been extensively studied in the literature. Most studies measure burnout using the Maslach Burnout Inventory (MBI), an instrument with three dimensions (emotional exhaustion, depersonalisation, lack of personal accomplishment). However, traditional factor analysis or correlation methods have often failed to confirm this structure. Here we employ Latent Profile Analysis (LPA) in order to better study the burnout phenomenon.

Methods: Burnout, job satisfaction and job characteristics were studied in a convenience sample of 460 Greek secondary and high school teachers during the economic crisis in 2019. Explanatory and confirmatory factor analysis were used to validate MBI structure together with LPA which was applied for the first time in a Greek setting. The results were then correlated with job satisfaction and job characteristics.

Results: The three-dimensional structure was confirmed by factor analysis. LPA identified four profiles: burnout (all MBI dimensions have high values), overextended (high for emotional exhaustion only), inefficient (high for personal accomplishment only) and engaged (all MBI dimensions have low values). The most common profile among the teachers in this sample was overextended (50%). Teachers in each profile behaved differently with regard to job satisfaction and attitudes towards school-related sources of problems. In particular, Burnt-out teachers were more negative than the Engaged ones in almost all variables assessed, except educational policy, while differences between the two intermediate profiles were less marked. Engaged teachers showed the highest levels of job satisfaction, followed by Overextended and Ineffective, with Burnt-out teachers showing the lowest levels.

Conclusion: The derived profiles and their different attitudes showed the significance of all three dimensions of burnout syndrome. In practical terms, interventions appropriately targeted to each profile can be designed and implemented to prevent or reduce burnout.

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-10 Utilising screening cohorts to model DCIS and invasive breast cancer

Evripidis Kapanidis, Keith Humphreys
Karolinska Institutet, Stockholm, Sweden

Introduction: Ductal Carcinoma in Situ (DCIS), is considered a precursor of invasive ductal carcinoma. Essentially, a DCIS lesion is comprised of abnormal cells, that are restricted in the ducts of the breast. If left untreated, a portion of these lesions will, during the lifetime of an individual, expand outside of the breast duct, progressing into invasive breast cancer. The transition of DCIS to invasive cancer is poorly understood and there are growing concerns about its possible overdiagnosis and overtreatment.

Methods: In order to describe the natural history of DCIS and its invasive component we develop a continuous tumour growth model that describes the distributions of DCIS onset and progression to invasiveness, tumour growth rate and the type of detection (screened or symptomatically detected). The model is suitable for a cohort design setting where healthy women enter the study and are followed-up for a period of time, while attending screening examinations. For detected individuals, each likelihood contribution utilizes information on the type of carcinoma, its size and the age at detection. This information can be extracted from Swedish breast cancer registers, upon which we will base an observational study. Our approach is novel in that it incorporates more complex functions for the DCIS detection and progression to invasiveness, than those used classically for analysing breast cancer screening data [1]. We perform microsimulations to assess parameter identifiability when the parametric assumptions are met and to assess robustness under model misspecification. Because the growth pattern of DCIS is not well known, we use a flexible two-parameter power law growth function and test how well its parameters are estimated by our model. We simulate cohorts of 50000 women who are followed-up until age 74 and start attending screening from age 40. The estimation is based on likelihood inference.

Results: Overall, our simulations show that if parametric assumptions are reasonable, the parameters are identifiable and the growth characteristics of 1 DCIS as well as its potential to invade can be estimated well.

Conclusions: New natural history models of cancer, such as ours, can be used for assessing cancer control policies that cannot be evaluated in a clinical trial setting.

References: *[1] van Ravesteyn, N. T., et al (2018). Modeling ductal carcinoma in situ (DCIS): an overview of CISNET model approaches. Medical Decision Making, 38(1_suppl), 126S-139S.*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-11 A correlated joint frailty model for recurrent and terminal events: Application to breast cancer patientss

Elaheh Talebi-Ghane[1], AhmadReza Baghestani[2]
1 Department of Biostatistics, School of Public Health, Modeling Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
2 Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

In many epidemiological and clinical studies, the progression of disease such as recurrent or repeated events in cohort studies could be terminated by a dependent final outcome such as death. In this context, the survival time could be lengthened or shortened according to the time or the frequency of recurrent events which be evidence for the existence of dependence between time of relapse and death events. Moreover, some patients were more predisposed in encountering relapse earlier or more frequent than others due to some known or unknown prognostic factors which indicated the dependence between the occurrences of these events and the potential heterogeneity across subjects. Frailty and correlated frailty models are recently becoming increasingly applied for modeling dependencies in multivariate survival analysis[1,2]. Joint frailty models as the most popular model can perfectly assess the effect of covariates on the risk of recurrent and death events, simultaneously and can also consider both the inter-recurrences dependence and the dependence between the recurrences and the survival times [3,4]. Despite the great advantages of these models, when the correlation was existed between the recurrent events and death, these models were not able to include it and the performance of them was decreased and may lead to erroneous results. We introduced a correlated joint frailty model for analysis of the recurrent and terminal events and estimated the effect of covariates on them, concurrently. This model considered three different frailties: an independent frailty for recurrent events with gamma distribution and two frailties for recurrent and death events with bivariate log-normal distribution to consider variance parameters of the frailties in shared frailty model and contain additional parameters for modeling the correlation between frailties. To estimate the model parameters, the maximum likelihood estimation based on numerical integration was applied for a parametric and piecewise baseline hazard function. The performance of this proposed model was evaluated and compared with two previous reduced models using simulation studies. We detected positive, negative or independent correlation of frailties in the bivariate correlated frailty model and analyzed them in detail. In addition, the application of these proposed models was evaluated using a real dataset on patients with breast cancer who had undergone surgery. In the presence of positive, negative correlation and independency between recurrent events and death, the performance of proposed model was better than those two reduced models, however the percentage of coverage in the negative correlation was a little low.

References: *[1] Hens N, Wienke A, Aerts M, Molenberghs G. The correlated and shared gamma frailty model for bivariate current status data: an illustration for cross-sectional serological data. Statistics in Medicine. 2009 Sep 30;28(22):2785-800. [2] Dabade AD. Compound negative binomial multivariate correlated frailty model for long-term survivors. Communications in Statistics-Simulation and Computation. 2022 May 2:1-6. [3] Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. Biostatistics. 2007 Oct 1;8(4):708-21. [4] Mazroui Y, Mathoulin□Pelissier S, Soubeyran P, Rondeau V. General joint frailty model for recurrent event data with a dependent terminal event: application to follicular lymphoma data. Statistics in medicine. 2012 May 20;31(11-12):1162-76.*

POSTER SESSION A-07:
Joint and Latent Variables Models

P-A07-12 LB

A joint frailty model to assess the relationship between time to curative treatment and biochemical recurrence in Prostate Cancer patients

Abderrahim Oussama Batouche[1,2,3], Denis Rustand[4], Eugen Czeizler[2,3], Tuomas Mirtti[5,6], Antti Rannikko[7], Haavard Rue[4]

1 Department of Computer Science, University of Helsinki, Helsinki, Finland
2 Research Program in Systems Oncology, University of Helsinki, Helsinki, Finland
3 ICAN-Digital Precision Cancer Medicine Flagship, Helsinki, Finland
4 King Abdullah University of Science and Technology (KAUST), Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, Thuwal, Saudi Arabia
5 Helsinki University Hospital, Department of Pathology, Helsinki, Finland
6 Department of Biomedical Engineering, Georgia Tech and Emory University, Atlanta, United States
7 Helsinki University Hospital, Department of Urology, Helsinki, Finlandn

Background: Conflicting evidence exists for the effect of prostate cancer (PCa) treatment delay on outcomes after curative treatment. PCa is generally slow-growing, and treatment delays up to 6 months have not been associated with impaired outcomes. However, the effect of treatment delay on outcomes in intermediate PCa is more controversial, especially during the MRI era. In this study, we assess the effect of delaying or expediting curative treatment (surgery or radiation therapy) on the time to biochemical recurrence (BCR).

Methods: We used 'HUS' Helsinki University Hospital data lake (within "HUS Acamedic") for data mining and to categorise PCa patients by Gleason grade group (1-5), treatment (RP/RT), and outcome (BCR) for a final sample size of 11719 patients (1993–2019). A broader definition of biochemical recurrence was used considering secondary treatments as an additional indicator of relapse alongside traditional PSA cut-offs (PSA of 0.2 ug/l for surgery and PSA nadir +2 ug/l for radiation therapy) [1]. A shared frailty joint survival model (utilising the INLAJoint R package [2]) was employed to model the relationship between the time from diagnosis to curative treatment (treatment risk) and the time from diagnosis to BCR (relapse risk).

Results: Conditional on covariates, including age at diagnosis, MRI status, diagnostic PSA and Gleason grade group, our joint survival model revealed a negative association $\gamma = -1.32$ [-1.49, -1.13] between risk of treatment and risk of treatment recurrence. Regardless of the grade group, patients within the top 15% with the lowest risk of receiving treatment (i.e., the longest time to treatment) exhibited a 1.89 [1.05, 6.47] fold (increased) risk of recurrence compared to the average patient.

Conclusion: We found that time to curative treatment is associated with the risk of relapse. Patients of the same age, diagnostic PSA and same grade group who are treated early, are less likely to observe recurrence. These results could be incorporated into the update of PCa treatment guidelines to reflect the importance of treatment timing.

References: *[1] Batouche et al., Synergizing data imputation and electronic health records for advancing prostate cancer research: Challenges, and practical applications, in: Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - 2024. doi:10.5220/0012350300003657 [2] Rustand et al., Joint Modeling of Multivariate Longitudinal and Survival Outcomes with the R package INLAjoint, arXiv:2402.08335 - Apr. 2024. doi:10.48550/arXiv.2402.08335*

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-01

Multivariate survival analysis for high dimensional data: A sample splitting approach

Beryl Ang'iro[1], Roel Braekers[2]
1 Karatina University, Kenya
2 Hasselt University, Belgium

Introduction: In multivariate survival analysis, we are interested in estimating regression parameters. However, when the sample size is extraordinarily large or the estimation procedure is complex, it is difficult to finish the analysis within a reasonable time period. Furthermore, accounting for correlation in clustered data remains a challenge especially when the total number of clusters are large and/or there are large number of replications per cluster. Clustered multivariate failure times are frequently analysed using frailty and copula models. Here, the different response event times are grouped into clusters hence the dependence between the different event times in a cluster is of interest. In this talk, we focus our attention to the clayton copula, a member of the Archimedean copula family and propose a sample splitting methodology to analyze large-scale multivariate failure time data. Methods: We indroduce the method of sample splitting for multivariate survival data by combining estimates from each subgroup into global estimates. We develop both one- and two-stage estimation proceedures for the copula paramers in each sub-group and consider three choices of weights; minus second derivative of the log likelihood; the inverse of the estimated variance in each subgroup; and naive weights depending on the number of clusters per subgroup. Our method is illustrated on the Kenya Demographic health 1 Survey (KDHS, 2014) data.

Results: It is shown that the proposed global estimator is asymptotically exquivalent to the full data estimator. Given heterogeneous dataset, consistency and asymptotic normality of the global estimator still hold as well as same asymptotic variance given the three weights.

Conclusion: The combined results using each of the three weights provides a statistical inference which is similar to the one from analyzing the entire data set all at once.

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-02

Infusing structural assumptions into dimension reduction for single-cell RNA sequencing data to identify small gene sets

Niklas Brunn[1,2], Maren Hackenberg[1,2], Tanja Vogel[3], Harald Binder[1,2,4]

1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg im Breisgau, Germany
2 Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg im Breisgau, Germany
3 Institute of Anatomy and Cell Biology, Department of Molecular Embryology, Faculty of Biology, University of Freiburg, Freiburg im Breisgau, Germany
4 Centre for Integrative Biological Signaling Studies (CIBSS), University of Freiburg, Freiburg im Breisgau, Germany

Background: Dimension reduction techniques significantly enhance the exploration of cellular heterogeneity in single-cell RNA sequencing data. While these approaches are predominantly data-driven, it may still be useful to incorporate biologically plausible assumptions about the underlying structure or the experimental design. For instance, dimensions that help to distinguish between cell groups intuitively should be characterised by distinct small sets of genes. Additionally, the design in a time series experiment should be incorporated such that gradual changes in corresponding gene sets characterise temporal changes of cell states.

Methods: We propose the boosting autoencoder approach, which synergises unsupervised deep learning models for representation learning and component-wise boosting to formalise constraints. Specifically, the approach selects distinct small sets of genes by maximising a score function to explain cell state-related patterns in separate latent dimensions. In addition, the boosting component can couple dimensions across time corresponding to groups of temporally evolving cells via a transfer learning strategy.

Results: We demonstrate the functionality of our approach through applications on simulated data, accounting for different structural assumptions. There, our approach captures small sets of descriptive genes for different simulated cell stages in distinct dimensions. In addition, we explore the diversity of neural cell identities and temporal patterns of embryonic development in applications to gene expression data.

Conclusion: The boosting autoencoder represents a complementary dimension reduction strategy to learn a structured representation of single-cell RNA sequencing data and to identify cell stage-related genes without requiring predefined cluster memberships. The synergy of deep learning and component-wise boosting improves the interpretability and biological relevance of the identified patterns indifferent latent dimensions.

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-03 Utilising high-dimensional data in randomised clinical trials

Svetlana Cherlin, James M S Wason
Newcastle University, Newcastle upon Tyne, United Kingdom

Introduction: Even in effectively conducted randomised trials, the probability of a successful study remains relatively low. With recent advances in the next-generation sequencing technologies, there is a rapidly growing number of high-dimensional data, including genetic, molecular, and phenotypic information, that have improved our understanding of driver genes, drug targets, and drug mechanisms of action. The leveraging of high-dimensional data holds promise for increased success of clinical trials.

Methods: We conducted a review of recently published randomised clinical trials that utilise high-dimensional genetic data, to discuss the current practices and landscape. We summarised the clinical area, the type of high-dimensional data, the number of covariates used, number of treatment arms, and the purpose of collecting high-dimensional data. We provide an overview of available methods for utilising high-dimensional data in clinical trials, such as univariable and multivariable approaches, penalised approaches, machine learning, clustering methods, and adaptive signature design. We illustrate our recently developed method of an efficient use of high-dimensional data in clinical trials (the CVRS2 method by Cherlin & Wason 2023, doi:10.1093/biostatistics/kxaa055). The method is based on developing and testing the efficacy of a treatment in a high-efficacy patient group in two trial endpoints (the sensitive group). The design utilises bivariate risk scores divided into clusters using a nonparametric clustering procedure. The design is implemented in an R package "rapids".

Results: Our review of utilising high-dimensional data in clinical trials revealed that the most common method of analysis was univariable analysis. Articles that described multivariable analyses used standard statistical methods. While assessing the properties of the CVRS2 method on the simulated data, we showed that it was able to reliably identify the sensitive group. CVRS2 outperformed the univariable risk scores method that analysed each outcome separately. When applied to a real clinical trial, CVRS2 yielded a significant treatment effect for both outcomes, in contrast to the univariable risk scores method.

Conclusion: Utilising high-dimensional data in clinical trials can improve efficiency and increase patient benefit. New methodological approaches are required for more efficient analysis of the increasing amount of high-dimensional data collected in randomised clinical trials. For example, in many clinical trials, multiple outcomes are of interest, however the current methods have been proposed for a single endpoint. Our newly developed method for utilising high-dimensional data in trials with two end-point illustrates a more efficient use of high-dimensional data in clinical trials.

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-04

Prioritisation of differential methylation regions for the prediction of Coeliac disease: A Machine Learning approach

Paolo Dalena[1,2], Pio Adamo D'Adamo[1,2], Giulia Barbati[1]
1 University of Trieste, Trieste, Italy
2 IRCCS Burlo Garofolo, Trieste, Italy

Background: Different methylation patterns regulate the switching on and off of certain genes. When two groups of patients are analysed, it is possible to identify Differential Methylation Probes (DMP), i.e. points in the genome where there is a recurring significant difference in methylation between the two groups. From the DMPs, it is possible to identify Differential Methylation Regions (DMRs), i.e. entire DNA segments where significant differences are recorded.

Methods: The values for the methylation of the probes were normalised and, through the application of the ChAMP pipeline [1], the relevant DMPs were identified and categorised into DMRs. The DMRs were ordered by considering the average distances between the DMPs in each group. First, a Random Forest (RF) model was fitted, considering all DMPs and was considered for computing the variables importance. Secondly, a grouped variable selection was performed using Group Lasso (selecting the shrinkage coefficients considering a 10-folds cross-validation) and a second RF model was fitted including only the groups that resulted as non-null. Both the models were optimized considering the number of variables randomly sampled as candidates at each split, the minimum size of terminal nodes and the number of trees to grow and were evaluated through the accuracy, the Out-Of-Bag estimation for the error rate (OOBe) and the test set error rate (TSe).

Results: From the initial dataset (148 patients and 731239 probes), 1284 DMPs and 148 DMRs were identified. The model with all DMPs presented a OOBe of 17.86% and a TSe of 22.22%. Considering the resulting variable importance, a greater presence of the DMPs included in the first 24 DMRs (ordered according to the prioritization identified) was clear: 36.73%, 41.41% and 47.88% of the first 50, 100 and 500 most important variables were included in the first 24 DMRs. The Group Lasso identified only 11 groups of variables (including 9 DMRs) and 70 coefficients as non-zero and the reduced model considering only these groups had smaller errors: 16.96% for the OOBe and 11.11% for TSe.

Conclusions: DMRs are possible functional regions involved in gene transcriptional regulation, so a study of DMRs would allow a condition to be identified from the subjects' DNA. It is possible to obtain a prioritization for DMRs that resulted associated with the coeliac outcome. Furthermore, by considering DMRs in a group variable selection, a reduced model with very good prediction performance was identified.

Reference: *[1] Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics. 2014 Feb 1;30(3):428-30.*

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-05

Statistical methodologies for Olink data analysis: Different approaches for reliable evaluation of Omics Data

Arianna Galotta, Nicolò Capra, Francesco Maria Mattio, Marta Brambilla, Marina Camera, Alice Bonomi
IRCCS Centro Cardiologico Monzino, Milan, Italy

Introduction: Olink® data is a valuable and innovative resource for biomarker discovery, by a comprehensive protein profiles through highly sensitive detection technology. The application of various statistical methodologies is required for accurate data analysis and interpretation, aiming to extract meaningful information and identify patterns related to the measured biomolecules. We applied Olink® technology to identify plasma biomarkers predictive of cardiovascular mortality in coronary artery disease patients. The goal of the study was to compare statistical approaches to reliably assess Omics data.

Methods: Sixty-nine plasma samples were analysed. The outcome of interest was cardiovascular mortality (n=17, 24.6%). Four Olink Target 96 panels were generated for the analysis of protein biomarkers in the following areas: cardiometabolic, cardiovascular II and III, and inflammation. Various statistical analysis and data visualization methodologies were employed, including t-test and volcano plot, principal component analysis (PCA), Gene Set Enrichment Analysis (GSEA) and heatmap, Boruta analysis and multivariate analysis with stepwise selection. Statistical analyses were performed using SAS software v9.4 and R software v4.3.1 (Olink Analyze R package [Nevola, et al. 2023]).

Results: The dataset included 333 Olink NPX variables across 4 target panels. T-test results, visualized using adjusted p-values in a volcano plot, showed no significant differences in proteins levels between cases and controls. Through PCA plot, low variance explained by two principal components was observed in all panels. GSEA was performed, using the estimate from t-test analysis for a contrast for all proteins; however, by visualizing through a heatmap and using the adjusted p-values, no significance was shown for any enrichment in any panel. Boruta analysis identified crucial features for a multivariable logistic model. A stepwise selection applied to significant variables from univariate analysis (using unadjusted p-values) identified predictors to be used in a logistic model, observing an AUC of 0.89 (IC 95%: 0.81-0.97). Collaboration with clinical experts was pivotal for context-specific result identification and interpretation.

Conclusion: Integration of different statistical methodologies provides a comprehensive framework for Olink data interpretation in biomedical research. There is no a unique correct methodology, but the choice of approach depends on the nature of the data and the research objectives. In this context, the results obtained from each statistical approach differed, a subsequential adjustment of the analyses did not lead to satisfactory results. Collaboration between statistical experts and clinicians is crucial for the selection of the most appropriate statistical methodologies for a proper interpretation of Omics data.

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-06 How to benefit from high-dimensional gene expression data for dose-response modelling

Franziska Kappenberg, Julia Duda, Onur Gül, Ludger Sandig, Leonie Schürmeyer, Marieke Stolte, Kirsten Schorning, Jörg Rahnenführer
TU Dortmund University, Germany

Background/Introduction: High-throughput methods allow the simultaneous measurement of thousands of genes. However, due to cost, time and ethical constraints, sample sizes in dose- or concentration-response experiments, e.g. in early stages of clinical studies, are typically very small, leading to datasets with p>>n. In this talk, several research projects on the analysis of concentration-gene expression data are presented. All projects are part of the research of members of the Research Training Group "Biostatistical Methods for High-Dimensional Data in Toxicology".

Methods: For the classification of compounds into hepatotoxic and non-hepatotoxic, combining interpretable summaries of concentration-expression data is shown to lead to an improved accuracy in comparison to the consideration of viability information alone. The MCP-Mod procedure, combining a multiple comparison testing procedure and a modelling step, was originally developed for Phase II clinical studies. When simultaneously applying this procedure to high-dimensional gene expression datasets, favourable results are achieved. The first step in any experiment is to plan the experimental design, where corresponding statistical approaches often clearly improve analysis results. Especially in gene expression data, one challenge is the requirement that a common design must be determined for the entire set of genes. In the presented project, a simultaneous D-optimal design is found to lead to the best model fits. When fitting individual models for a gene expression dataset, estimation of many parameters is required. To reduce the number of parameters, common parameters are assumed within a group of genes. The use of a focused information criterion allows the selection of an appropriate model within this framework. A relaxation of the assumption of common parameters is information sharing between genes. A newly developed empirical Bayes method shrinks the estimates of the parameter corresponding to the half-maximal effect of a fitted curve towards the population mean, resulting in an improved estimation of this parameter. This idea is in a second step extended to a fully Bayesian hierarchical model.

Results: The feasibility or the benefit of applying specific statistical methods to high-dimensional gene expression datasets is explicitly demonstrated. For all newly developed methods, controlled simulation studies were conducted to assess the performance of the methods in comparison to established approaches.

Conclusions: All projects presented in this talk show that while high-dimensional gene expression data typically come with a p>>n challenge, many statistical approaches can be applied nonetheless, or even benefitting from thoughtful incorporation of gene expression data or sharing of information between genes is possible.

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-07

Spatial proteomics analysis for discriminating autoimmune liver diseases: Insights from MALDI-TOF mass spectrometry imaging

Elisa Merelli[1], Giulia Capitoli[1], Vanna Denti[2], Isabella Piga[2,3], Fabio Pagni[4], Pietro Invernizzi[5,6], Vincenzo L'Imperio[4], Marco Carbone[5,6], Alessio Gerussi[5,6]

1 Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy
2 Department of Medicine and Surgery, Clinical Proteomics and Metabolomics Unit, University of Milano - Bicocca, Monza, Italy
3 Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy
4 Department of Medicine and Surgery, Pathology, Fondazione IRCCS San Gerardo dei Tintori, University of Milan-Bicocca, Monza, Italy
5 Division of Gastroenterology and Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy
6 Division European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy

Introduction: Autoimmune liver diseases (AILDs) comprise a spectrum of conditions, including autoimmune hepatitis (AIH), primary biliary cholangitis (PBC), and variant syndrome (VS), which shows some features of both AIH and PBC. Traditional diagnostic approaches lack specificity. Spatial proteomics, which integrates spatially resolved mass spectrometry data, offers a promising avenue for identifying molecular signatures that can differentiate between these entities.

Methods: In this study, we employed spatial proteomics to differentiate between AIH, PBC, and VS. For each condition, biopsy samples collected from five patients underwent Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry imaging (MSI) to generate spatially resolved mass spectral data, reflecting a total sample size of 15 patients. The data were labeled by disease classification and annotated with liver region information, including portal tract and centrolobular region. Subsequently, a comprehensive data preparation and noise reduction phase was employed to enhance data quality and reduce variability. The processed data were then subjected to various unsupervised clustering techniques [1] and supervised algorithms, such as discriminant analysis, implemented in the R programming environment.

Results: Our analysis revealed distinct spatial expression patterns of peptides among the different autoimmune liver diseases, with significant variations observed across liver regions. Unsupervised clustering techniques identified three clusters of peptides with differential spatial expression patterns characteristic of each disease entity within specific liver regions. Furthermore, supervised algorithms successfully identified discriminatory features that accurately distinguished between AIH, PBC, and VS. The identified spatial molecular signatures exhibited high discriminatory power.

Conclusions: Our preliminary findings suggests that spatial proteomics can differentiate AIH, PBC and VS based on the spatial distribution of molecular signatures within distinct liver regions. If confirmed on larger cohorts, these data can represent the groundwork for the development of non-invasive diagnostic assays.

References: *[1] M. S. Nobile et al., 'Unsupervised neural networks as a support tool for pathology diagnosis in MALDI-MSI experiments: A case study on thyroid biopsies', Expert Syst. Appl., vol. 215, p. 119296, Apr. 2023, doi: 10.1016/j.eswa.2022.119296.*

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-08

Clustering high-dimensional spaces using a modified EM algorithm with fractional order assignment

Shivani Saxena, Ahsan Rizvi
Institute of Advanced Research, India

Background: Clustering algorithms often face challenges when dealing with high-dimensional data due to the curse of dimensionality. As the number of dimensions increases, conventional methods struggle to identify meaningful patterns and clusters, leading to reduced accuracy. Existing approaches have attempted to address these limitations, but still encounter issues such as sensitivity to noise, inefficiency in capturing complex relationships, and inability to handle uncertainties effectively.

Method: In this study, we propose a novel solution called EM-FASL (Expectation-Maximization with Fractional Assignment and Logistic Regression) to improve clustering accuracy in high-dimensional spaces. EM-FASL builds upon the Expectation-Maximization (EM) algorithm by incorporating fractional order assignment and logistic regression. The algorithm follows a systematic approach, initializing parameters for both Gaussian Mixture Model (GMM) and logistic regression models. In the Expectation (E) step, EMFASL fits the GMM, computes fractional assignments for data points, and evaluates fractional entropy. In the Maximization (M) step, it updates GMM parameters based on fractional assignments and adjusts logistic regression coefficients to refine mean vectors.

Result: Our experimental results demonstrate the effectiveness of EM-FASL in improving clustering accuracy in high-dimensional spaces. Compared to existing methods, EM-FASL converges to a solution with a lower fractional entropy, indicating better clustering performance. The algorithm successfully captures intricate data patterns that were previously challenging to identify, showcasing its superiority in handling high-dimensional data.

Conclusion: EM-FASL presents a promising solution to the challenges of clustering in high-dimensional spaces. By integrating fractional order assignment and logistic regression into the EM framework, EM-FASL offers improved accuracy and robustness in identifying clusters in complex datasets. The algorithm's ability to handle uncertainties and capture subtle relationships makes it a valuable tool for various applications, paving the way for further advancements in high-dimensional clustering techniques

POSTER SESSION A-08:
Methods for High Dimensional Data (including -omics)

P-A08-09 LB

Characterising the omics landscape based on 10,000+ datasets

Eva Brombacher[1,2,3,4], Oliver Schilling[5,6,7], Clemens Kreutz[1,2]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg, Germany
2 Centre for Integrative Biological Signaling Studies (CIBSS), University of Freiburg, Freiburg, Germany
3 Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Freiburg, Germany
4 Faculty of Biology, University of Freiburg, Freiburg, Germany
5 Institute for Surgical Pathology, Faculty of Medicine, Medical Center – University of Freiburg, Freiburg, Germany,
6 German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Heidelberg, Germany
7 BIOSS Centre for Biological Signaling Studies, University of Freiburg, Freiburg, Germany

Background/Introduction: The characteristics of data generated by omics technologies are crucial as they critically impact the effectiveness and feasibility of computational approaches used in downstream analyses, such as data harmonization and differential abundance analyses. Variability in these data characteristics across different datasets can lead to varying results in benchmarking studies, which are essential for selecting the appropriate analysis methods across all omics disciplines. Additionally, downstream analysis tools are developed and employed in specific omics communities due to presumed differences in data characteristics linked to each omics technology.

Methods: We analyzed over ten thousand datasets to examine how data characteristics differ among proteomics, metabolomics, transcriptomics, and microbiome data.

Results: Our study identified distinctive patterns in data characteristics for each omics type and introduced a tool that helps researchers evaluate how representative a given omics dataset is for its respective omics field. Moreover, we illustrate how data characteristics can impact analysis at the example of normalization in the presence of sample-dependent proportions of missing values.

Conclusion: Given the variability of these omics data characteristics, we recommend inspecting them in the context of benchmark studies and downstream analyses to avoid suboptimal method selection and unintended biases.

POSTER SESSION A-09:
Meta-analysis

P-A09-01 Comparative performance of different depression screening tools in primary care setting: A systematic review and network meta-analysis

Pattawan Pattanaatapong[2], Thunyarat Anothaisintawee[1], Assistant Kanokporn Sukhato[2], Itthipon Wongprom[2], Jirayus Siangfu[1], Ammarin Thakkinstian[1]

1 Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

2 Department of Clinical Department of Family Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Introduction: Depression is the most common mental disorder in primary care settings. Early identification and intervention can improve the clinical outcomes of depression. There are several screening tools available and commonly used in the primary care including Patient Health Question (PHQ)-9, PHQ-2, Hamilton Rating Scale for Depression (HAMS-D), Beck Depression Inventory (BDI), and Center for Epidemiological Studies Depression Scale (CES-D). These tools have different performance but there are a few studies that directedly compared the performance of these tests. Therefore, this systematic review aimed to compare the performance of PHQ-9, PHQ-2, BDI, and CESD and to identify the best tool for screening depression in primary care settings.

Methods: Relevant studies were identified from Medline, EMBASE, and CINAHL databases since their inceptions to October 2023. Diagnostic studies were eligible if they were conducted in the primary care settings, compared the performance of any of depression screening tools (i.e., PHQ-9, PHQ-2, HADS-D, BDI, or CESD) with structural psychiatric interview, and had the outcome as depressive disorder. One stage network meta-analysis by a multilevel logistic regression, was applied for pooling performance of screening tools across studies. Relative diagnostic odds ratios (DOR) of each tool relative to the standard test along with 95% confidence intervals (CI) were then estimated and compare the performance among all screening tools; an area under receive operating characteristic curve (ROC) was also estimated to identify the possible best screening test for diagnosis depressive disorders in primary care setting.

Results: One-hundred and eleven studies were included consisting of 63, 28, 12, 29, and 11 studies for PHQ-9, PHQ-2, HADS-D, CESD, and BDI, respectively. Pooled DOR was highest for PHQ-9, followed by BDI, CESD, PHQ-2 and HADS-D, with DORs (95% CI) of 23.24 (20.63-26.17), 22.43 (15.99-31.47), 17.07 (13.72, 21.23), 15.52 (13.31-18.09), and 14.57 (11.45-18.54). Pooled ROCs (95% CI) of these corresponding tools were 0.83 (082-0.84), 0.83 (0.80-0.85), 0.81 (0.79-0.82), 0.80 (0.78-0.81), and 0.79 (0.77-0.81). When compared among these tools, the diagnostic performance of PHQ-9 was significantly higher than those of PHQ-2, HADS-D, and CESD with relative DORs of 1.50 (1.26-1.78), 1.59 (1.23-2.07), and 1.36 (1.07, 1.74), respectively. In addition, the diagnostic performance of BDI was significantly better than PHQ-2 and HADS-D with relative DORs of 1.45 (1.00-2.09), and 1.54 (1.03-2.30), respectively.

Conclusions: Among all tests, PHQ-9 had the highest performance for diagnosis depressive disorders. Thus, PHQ-9 might be the most appropriate tool used for screening depressive disorders in primary care settings.

POSTER SESSION A-09:
Meta-analysis

P-A09-02

Sedentary lifestyle in childhood, adolescence, and young adulthood and its impact on health outcomes: An umbrella reviews

Maria Konstantina Bouka[1], Christos Chalitsios[2], Maria Efthymiou[2], Christina Pertsali[2], Elpiniki Laiou[4], Kostas Tsilidis[2,3], Georgios Markozannes[2,3]

1 Department of Mathematics, University of Tennessee-Knoxville, Knoxville, United States
2 Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece
3 Department of Biostatistics and Epidemiology, School of Public Health, Imperial College London, London, United Kingdom
4 Department of Nursing, University of Ioannina, Ioannina, Greece

Background: Sedentary behaviour is a distinct risk factor from physical activity, is defined as the energy expenditure of less than 1.5 metabolic equivalents of a task while one assumes a sitting or reclining posture. Sedentary behaviour has become increasingly prevalent in modern societies, and its rates are increasing at young ages. Mapping the diverse health outcomes associated with sedentary behaviour in early life is crucial for developing effective interventions and strategies.

Methods: We performed an umbrella review of meta-analyses of both observational and interventional studies examining sedentary behaviour in children or young populations (≤24 years) and any health outcome. We searched PubMed, EMBASE and CENTRAL till September 2023. Evidence was graded as strong, highly suggestive, suggestive, or weak based on the significance of the random effects model and the largest study in the meta-analysis, the number of cases, between-study heterogeneity, 95% prediction intervals, small study effects, and excess significance bias.

Results: We included 10 publications reporting 31 meta-analyses (16 of which included only cross-sectional studies and another 12 included both cross-sectional and prospective studies) with a median of 5 studies (range 3-17) and 35,082 participants (range 326-177,486) per meta-analysis. All meta-analyses included a mix of children, adolescents and young adults. Exposures included total sedentary behaviour (n=3), total (n=11) and specific screen time (TV, n=8; computer/video games, n=7; mobile, n=1), and two meta-analyses of interventional studies focused on reducing body mass index (BMI) though sedentary interventions resulting in weak and not significant evidence. Most studied outcomes pertained to anthropometry (n=10; BMI, obesity, waist circumference, fat %), depression (n=9), eye disorders (n=3; retinal arteriolar/venule equivalent, myopia), metabolic syndrome (n=3), attention deficit hyperactivity disorder (n=3), self-rated health (n=3), and neck pain (n=1). Only two associations, namely high internet/mobile phone use with depression (OR=1.90, 95%CI=1.62-2.21, 7 estimates, 3,996 cases) and computer/video gaming with poor self-rated health (OR=1.30, 95%CI: 1.20-1.41, 8 estimates, 25,131 cases) had strong evidence, and 20 associations (62.5%) received suggestive or weak evidence, all supporting the detrimental association of sedentary behaviours in examined health outcomes. Ten (31.3%) associations were not significant.

Conclusion: Despite that most meta-analyses provide weak or suggestive levels of evidence that might be due to relatively few available original studies of mostly cross-sectional design, it is an indisputable fact that a sedentary lifestyle has detrimental effects on a breadth of health outcomes in early life. Implementing effective measures to reduce sedentary behaviour during the early years of life can positively impact the long-term health and well-being of individuals across the life course.

POSTER SESSION A-09:
Meta-analysis

P-A09-03

Network meta-analysis of diagnostic test accuracy reported at multiple thresholds

Efthymia Derezea, Hayley E Jones
University of Bristol, United Kingdom

Introduction: Network meta-analysis of diagnostic test accuracy (NMA-DTA) is a relatively new field, involving combining evidence across studies to evaluate and compare the accuracy of different tests for a given disease or condition. However, the methods proposed to date cannot always capture complex aspects of the data. In fact, many commonly used diagnostic tests are continuous biomarkers, whose accuracy is evaluated at multiple thresholds within a study. Using current NMA-DTA methods we are feasibly able to include in our analysis only a few thresholds per study, discarding this way a big amount of data which could have provided us with useful information.

Methods: We propose a unification of an NMA-DTA model (Nyaga et al, 2018) and a model for meta-analysis of the accuracy of a single test across all thresholds (Jones et al, 2019). This incorporates multinomial likelihoods for studies reporting results across multiple thresholds and a parametric model structure for the relationship between the probability of testing positive and threshold within each disease class. Study-level random effects allow for between-study correlations arising from studies reporting on the accuracy of multiple tests, while additional random effects allow for variability across studies in the accuracy of each individual test. A hierarchical model structure can further be placed on the test-level variance components, allowing the inclusion of tests evaluated in very few studies in the model.

Results: We apply this model to data from a systematic review of the accuracy of tests for hepatocellular carcinoma in patients with liver cirrhosis. Across 161 included studies, a total of 42 different tests were evaluated. Many of the included tests are continuous biomarkers with some of them reporting data at as many as 48 different thresholds. Applying the new model allowed inclusion of a larger number of tests (42 compared to only 26 using Nyaga et al's NMA-DTA model) and estimation of sensitivity and specificity different thresholds with increased precision.

Conclusions: Our model provides a framework to jointly synthesise the accuracy of different tests reported at multiple thresholds. This allows more data and more tests to be included in the synthesis, and facilitates comparison of the accuracy of tests across the full range of thresholds.

POSTER SESSION A-09:
Meta-analysis

P-A09-04 A Bayesian network meta-analysis with mediation for length of stay and early mortality in ICU setting

Konstantina Thaleia Pilali, Maria Vittoria Chiaruttini, Cinzia Anna Maria Papappicco, Honoria Ocagli, Dario Gregori
University of Padova, Padova, Italy

Introduction: Network meta-analysis (NMA) extends standard meta-analysis by enabling the construction of an evidence network to compare interventions not directly compared in existing literature. Utilizing a Bayesian approach, it generates a posterior probability distribution of treatment effects, aiding in quantifying parameter estimate uncertainty and ranking treatments within the network.

Methods: This paper leverages the Bayesian NMA to compare the efficacy of different cardiac surgery techniques, including off-pump cardiac surgery (reference), extracorporeal membrane oxygenation (ECMO) as default treatment (ECMOd), ECMO as rescue treatment (ECMOr), and cardiopulmonary bypass (CPB) on benchmarking outcomes in Intensive Care Units (ICUs), notably mortality rates and length of stay. Twenty-three trials involving 4046 participants were included in the analysis. Data on protocol design, population characteristics and risk of bias were extracted and analysed for each included study. This information was used to conduct mediation analyses through meta-regression models, in order to evaluate the NMA results conditionally to the categories of potential effect-modifier. Mean difference and Odds Ratio (OR) along with corresponding credibility intervals were used as summary statistics to compare the treatment effects for continuous and binary outcomes, respectively. Furthermore, the surface under the cumulative ranking (SUCRA) and key metrics such as the deviance contribution of each trial-arm, were performed.

Results: The network comprised 6 possible comparisons among interventions, with direct comparisons ranging from 1 to 13 trials. In preliminary results, OFFPUMP emerged as the treatment with the highest posterior probability of being favoured over all the other interventions in the network, particularly 99.2% and 96.4% for Length of Stay (LOS) and early mortality, respectively. ECMOr had the second-highest posterior probability of being the favoured intervention in the network (64.9% for LOS and 50.8% for early mortality). Specifically, the CPB and ECMOd arms resulted in an increased odd of early mortality by 2.39 and 2.51 times greater, respectively, compared to the reference treatment. Age stands out as a significant mediator in ICU outcomes. For instance, regarding the LOS, the mean difference for CPB versus OFFPUMP and ECMOr versus OFFPUMP are equal to 9.1 (4.5-13.7) and 5.9 (4.3-7.4) for patients aged 50+ and 6.1 (2.3-9.2) and 0.23 (-0.6-1.1) for patients aged < 50, respectively.

Conclusion: Bayesian NMA is a comprehensive approach that allows for the estimation of both direct and indirect effects providing a fuller picture of how treatments work in diverse mediator categories, within the ICU setting.

POSTER SESSION A-09:
Meta-analysis

P-A09-05

Efficacy of targeted therapy or immunotherapy as adjuvant treatment for non-small cell lung cancer: A systematic review and network meta-analysis

Sasivimol Rattanasiri[1], Sunatee Sanguansai[1,2], Prapaporn Pornsuriyasak[1], Gareth J McKay[3], John Attia[4], Ammarin Thakkinstian[1]

1 Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
2 Oncology Unit, Department of Medicine, Rajavithi Hospital, College of Medicine, Rangsit University, Bangkok, Thailand
3 Centre for Public Health, Queen's University Belfast, Ireland, United Kingdom,
4 School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

Background: The integration of targeted and immunotherapy into adjuvant treatments for early-stage non-small cell lung cancer (NSCLC) remains challenging. This systematic review and network meta-analysis (NMA) aims to compare the efficacy and safety of available adjuvant options.

Methods: Relevant studies were identified from MEDLINE, Scopus, and Cochrane CENTRAL through August 2023. Randomized controlled trials comparing the effects of any pairs of adjuvant therapy (i.e., epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs), vascular endothelial growth factor (VEGF) inhibitor, immune checkpoint inhibitor (ICI), and non-ICI immunotherapy, and chemotherapy/placebo) on disease-free survival (DFS), overall survival (OS), and severe adverse event (SAE) were selected. Individual patient data were generated from Kaplan-Meier curves. A parametric survival model was used to estimate the median time and hazard ratio (HR) for DFS and OS. A two-stage NMA was applied to estimate the risk ratio (RR) of SAE.

Results: Nineteen RCTs (n = 9,438) were included in this NMA. The median DFS with EGFR-TKIs and ICI immunotherapy were longer than chemotherapy/placebo with the median DFS (95% CI) of 69.24 (44.52, 93.97), 53.47 (-7.35, 114.28), and 39.49 (29.32, 49.67) months, respectively. The HRs (95% CI) for DFS were estimated and it was found that patients who received EGFR-TKIs and ICI immunotherapy were 0.64 (0.52, 0.79; p-value < 0.001) and 0.73 (0.25, 2.15; p-value =0.572) times less likely to experience recurrence compared with chemotherapy/placebo. The estimated median OS (95% CI) for chemotherapy/placebo, EGFR-TKIs, VEGF inhibitor, ICI immunotherapy, and non-ICI immunotherapy were 103.24 (82.21, 124.26), 116.83 (91.36, 142.31), 103.81 (76.75, 130.87), 105.72 (79.49, 131.95), and 101.85 (76.99, 126.72) months, respectively. Compared with chemotherapy/placebo, EGFR-TKIs and ICI immunotherapy could reduce the risk of death by about 9% [HR = 0.91 (0.80, 1.04)] and 6% [HR = 0.94 (0.79, 1.13)], respectively. However, ICI immunotherapy, EGFR-TKIs, VEGF inhibitor, and non-ICI immunotherapy were more likely to develop SAEs than chemotherapy/placebo with the corresponding RRs (95% CI) of 1.60 (0.29, 8.91), 1.24 (0.52, 2.91), 1.23 (0.11, 13.72), and 1.08 (0.19, 6.19), respectively.

Conclusion: This study indicates that EGFR-TKIs should be the first-line choice for managing early-stage NSCLC and may represent the best adjuvant treatment regimen with an intermediate risk of SAE.

POSTER SESSION A-09:
Meta-analysis

P-A09-06

The effectiveness of delivery modalities of diabetes prevention programs: A systematic review and component network meta-analysis

Ellesha Smith, Stephanie Hubbard, Suzanne Freeman, Daniel March, Molly Wells, Elnaz Saeedi, Louise Haddon, Kamlesh Khunti, Laura Gray
University of Leicester, Leicester, United Kingdom

Background: and aims: Type 2 diabetes prevention programs have been shown to be effective but they are also costly and intensive, which makes translation into routine primary care and community settings challenging. Identifying drivers of intervention effectiveness can inform pragmatic future implementation whilst maintaining the effectiveness of the programs. Translational studies have demonstrated that delivery modalities impact their effectiveness. This systematic review and component network meta-analysis assessed which delivery modality components of diabetes prevention programs are associated with reductions in type 2 diabetes incidence for individuals at high risk of type 2 diabetes (or pre-diabetes).

Methods: We searched MEDLINE, The Cochrane Library, Opengrey and clinicaltrials.gov from inception to 17th November 2022 for translational studies comparing lifestyle interventions with a minimum 12-month follow-up. Data were extracted from each trial by two investigators. Interventions were classified according to their delivery modalities. Random effects network meta-analyses and component network meta-analyses estimated the intervention effects.

Results: We identified 50 eligible studies involving 29,286 participants. Thirty-six (72.0%) randomized controlled trials, 10 (20.0%) cluster randomized controlled trials, and four (8.0%) observational studies were included. The network meta-analyses demonstrated that interventions delivered in person, both individually and in a group, and by telephone (individually) showed the greatest reduction in the incidence of type 2 diabetes and showed the greatest probability of being the most effective intervention (hazard ratio: 0.15 (95% credible interval: 0.02, 0.84), probability: 39%). Component network meta-analyses found in-person (individually) delivery was associated with greater reduction in incidence of type 2 diabetes (hazard ratio: 0.66, 95% credible interval: 0.41, 0.96) and in-person (group-based) delivery was associated with greater reductions in weight (mean difference: -1.53 kg, 95% credible interval: -2.18, -0.85) and HbA1c (mean difference: -0.74 mmol/mol, 95% credible interval: -1.28, -0.17), relative to usual care.

Conclusions: Although the interest in digital diabetes prevention programs is increasing, these results suggest that in-person delivery modalities are the most effective components for diabetes prevention. Future research needs to focus on how to improve the effectiveness of digital programs and ensuring preferential delivery for target populations to reduce health inequalities.

POSTER SESSION A-09:
Meta-analysis

P-A09-07 LB

Network meta-analysis with dose-response relationships

Maria Petropoulou, Gerta Rücker, Guido Schwarzer
Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany

Background: Network meta-analysis (NMA) has become a widely used methodology for synthesizing evidence from multiple interventions in a given medical condition. However, NMA applications typically overlook the crucial role that drug dosage plays in shaping intervention effects. Traditional NMAs usually treat each intervention dose as unrelated, using lumping or splitting approaches to define treatment nodes. Evaluating interventions with respect to dose-effect relationships can overcome the limitations of these approaches, significantly impacting decision-making, guiding future study designs, and aiding drug development. Therefore, we introduce a novel frequentist NMA model that evaluates the effects of several interventions by accounting for dose-response relationships (termed dose-response network meta-analysis).

Methods: We propose a dose-response NMA approach that models both linear and several nonlinear dose-response relationships. We considered exponential, quadratic, fractional polynomials and restricted cubic splines functions to capture the non-linear relationships. All parameters are estimated using weighted least-squares regression. Model fit is assessed using the Akaike Information Criterion and heterogeneity metrics such as a Chi-square test. We discuss the methodological properties and challenges of the proposed model. One of the advantages is that the model can also be applied to disconnected networks, if shared intervention doses are found in different subnetworks. An R function, netdose, was developed to make the proposed dose-response NMA model easily available to the scientific community.

Results: We illustrate the novel frequentist model with an NMA of interventions for preventing postoperative vomiting in adults after general anaesthesia. Standard NMA, linear and several nonlinear dose-response functions are applied. The results of the different models are presented with illustrative methods, and a model comparison is provided. We conclude that the results from dose-response NMA models differ from those of the standard NMA model, and the choice of the dose-response function is crucial for model fit and final conclusions.

Conclusion: The proposed dose-response NMA model provides valuable insights into how intervention effects vary with dosage. By incorporating dose-response relationships, this model significantly enhances informed decision-making processes in healthcare, offering insights that traditional methods cannot provide.

POSTER SESSION A-09:
Meta-analysis

P-A09-08

Implications of diagnostic meta-analysis studies for the design of future diagnostic studies

Nana-Adjoa Kwarteng, Adriani Nikolakopoulou
University of Freiburg Medical Center, Germany

Background/Introduction: In an age of ever-increasing existing studies, meta-analytic methods enable the synthesis of existing knowledge. However, existing evidence may lack clinical conclusiveness, and clinical conclusiveness is also relevant for meta-analyses of diagnostic test accuracy (DTA) studies. Therefore, when planning new DTA studies, considering findings from existing meta-analyses is valuable. We explore how to determine the need for additional evidence in a meta-analyses of DTA studies and how to estimate the required sample size to achieve conclusive results in an updated meta-analysis.

Methods: When synthesizing diagnostic test accuracy evidence, sensitivity and specificity estimates can be obtained simultaneously, necessitating bivariate methods for synthesis and sample size determination. To assess precision in existing sensitivity and specificity estimates, the combined elliptical area of standard errors from the variance–covariance matrix is used. Then, based on a desirable improvement in precision, the required sample size for one or more subsequent, conclusive studies is calculated. We propose two approaches for estimating the required sample size: an algebraic approach, targeted for common effect meta-analysis and low expected heterogeneity, and an iterative approach, targeted for scenarios with high expected heterogeneity.

Results: A published meta-analysis of diagnostic tests for plague is used to illustrate the suggested sample size calculation. In the example scenario with an original specificity value of 0.6998 [0.6441 - 0.7502], the iterative approach suggests that adding a single study with at least 83 participants to the published meta-analysis data would exclude the target value of 0.65 from the specificity confidence region. However, considering the addition of multiple studies, four additional studies with at least 28 participants in total would achieve equivalent precision.

Conclusion: The suggested approaches facilitate sample size estimation for planning subsequent DTA studies and inclusion in an updated meta-analysis. In some cases, the sample size estimation may suggest that multiple studies with low heterogeneity may be more efficient than a single additional study. Heterogeneity between studies impacts the number of studies needed to efficiently reach the target level of precision. The presented sample size calculation can aid investigators in making evidence-driven decisions for planning new DTA studies.

POSTER SESSION A-09:
Meta-analysis

P-A09-09

Diagnostic performance of WatchPAT peripheral arterial tonometry in detecting obstructive sleep apnea: Systematic review and meta-analysis

Ninlaya Rattanamongkolsak[1], Pawin Numthavaj[2], Visasiri Tantrakul[1], Sasivimol Rattanasiri[2]
1 Division of Sleep Medicine, Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
2 Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Introduction: Obstructive sleep apnea (OSA) is common in the general population and is considered as one of the risk factors for cardiovascular disease. In-laboratory specialist-monitored polysomnograph (PSG) is a gold standard for the diagnosis of OSA but requires extensive equipments and resources to conduct. WatchPAT peripheral arterial tonometry tool, one of the proposed screening tools incorporating measurements of peripheral arterial tone, pulse oximetry, heart rate, and actigraphy that could be easily administered at home, has been evaluated against PSG for the measurement of apnea-hypopnea index (AHI). We systematically reviewed the performance and applied meta-analysis to the sensitivity, specificity, and differences between WatchPAT and PSG. (Prospero registration CRD42023480556)

Methods: MEDLINE and SCOPUS were searched for relevant studies comparing WatchPAT and PSG and reported diagnostic performances of WatchPAT. We extracted the numbers of true positive, false positive, true negative, and false negative patients detected by WatchPAT for any OSA (defined as AHI>5), moderate to severe OSA (AHI>15), and severe OSA (AHI>30). In studies that reported correlation plots between AHI from WatchPAT and PSG, individual patient data points were also extracted for AHI by WebPlotDigitizer software. Diagnostic performances were pooled using metadata command in Stata software. We also assessed the quality of studies by QUADAS-2 risk of bias tool.

Results: A total of 16 studies were retrieved. Pooled sensitivity from 8 studies that measure the diagnostic performance of WatchPAT in detecting any OSA was 0.93 (95% CI 0.80 - 0.98, I2 53.89, ?2 1.67) while pooled specificity was 0.72 (95% CI 0.52 – 0.86, I2 71.73, ?2 1.23). In detecting moderate OSA, pooled sensitivity was slightly lower at 0.86 (95% CI 0.80 - 0.90, I2 9.87, ?2 0.04; N=8) with pooled specificity of 0.75 (95% CI 0.67 – 0.81, I2 22.98, ?2 0.09; N=8). In detecting severe OSA, pooled sensitivity was at 0.89 (95% CI 0.80 – 0.94, I2 0.07, ?2 53.89; N=5) and pooled specificity was 0.85 (95% CI 0.71 - 0.93, I2 53.89, ?2 0.38; N=5). Twelve studies reported correlation plot between AHI from two methods. The quality of studies is moderate as only a small number of studies use appropriate consecutive patient selection.

Conclusion: WatchPAT can be used as a screening tool for the diagnosis of OSA with high sensitivity. However, clinicians should note the lower specificity of the WatchPAT tool as the sole diagnosis of OSA.

POSTER SESSION A-09:
Meta-analysis

P-A09-10 Diagnostic accuracy of D-dimer for acute aortic syndromes: Systematic review and meta-analysis

Munira Essat, Steve Goodacre, Abdullah Pandor, Sarah Ren, Shijie (Kate) Ren, Mark Clowes
University of Sheffield, Sheffield, United Kingdom

Background: Acute aortic syndrome (AAS) is a life-threatening emergency condition. Previous systematic reviews of D-dimer diagnostic accuracy for AAS have been contradictory and based on limited data, but recently published studies offer potential for a more definitive overview. We aimed to perform a systematic review and meta-analysis to determine the diagnostic accuracy of D-dimer for diagnosing AAS.

Methods: We searched MEDLINE, EMBASE, and the Cochrane Library from inception to January 2023. Additionally, the reference lists of included studies and other systematic reviews were thoroughly searched. All diagnostic cohort studies (prospective or retrospective) that assessed the use of D-dimer for diagnosing AAS compared with a reference standard test (e.g. computer tomographic angiography (CTA), ECG-gated CTA, echocardiography, magnetic resonance angiography, operation, or autopsy) were included. Two independent reviewers completed study selection, data extractions and quality assessment using the QUADAS-2 tool. Data were synthesised using a bivariate random effect meta-analysis model.

Results: Of 1805 potentially relevant citations, 25 cohort studies met the inclusion criteria and 18 reporting the 500ng/mL threshold were included in the primary meta-analysis. The summary sensitivity was 96.5% (95% credible interval [CrI]: 94.8%, 98%) and summary specificity was 56.2% (95% CrI: 48.3%, 63.9%). Study specificity varied markedly from 33% to 86%. Sensitivity analysis including the 7 studies reporting other thresholds showed summary sensitivity of 95.7% (95% CrI: 93.2%, 97.5%) and summary specificity of 57.5% (95%CrI: 50.1%, 64.6%).

Conclusions: D-dimer has high sensitivity (96.5%) and moderate specificity (56.2%) for AAS. Previous meta-analysis reporting higher specificity may be explained by inclusion of case-control studies that may overestimate accuracy.

POSTER SESSION A-09:
Meta-analysis

P-A09-11 Analysis of aggregated data via generation of pseudo Individual Participant Data at Danone Nutricia Research

Luca Carlini, Mehdi Gholam
Danone Nutricia Research, Utrecht, Netherlands

Background/Introduction: In clinical research, most of the meta analyses are based on pooled/aggregated data (two-stage approach), while Individual Participant Data (IPD) has always been considered the gold standard. Sometimes this is not feasible due to lack of available and complete data, e.g. owing to privacy reasons. Nowadays, especially in industry, it is crucial to process data coming from different business units efficiently, trying to avoid as much as possible delays or obstacles due to legal regulations, while keeping high quality standards of data collection and analysis. In this work, we present a technique used to effectively generate pseudo-IPD based on summary statistics and we apply it to one of the studies undertaken at Danone Nutricia Research, where our aim is to obtain estimates of some parameters of interest, and to evaluate how the product effect differs across sites, via the application of mixed models on the generated raw pseudo-IPD.

Methods: The method was developed in Evidence synthesis methods for continuous outcomes. It provides a simple algorithm to generate pseudo-IPDs that have the same sufficient statistics as the true IPD; consequently, any likelihood-based analysis produces the same results as we would obtain from the true data. The technique can be employed either for a single continuous outcome, or when continuous outcomes are available both at baseline and follow-up. Once the IPD is generated by following the steps of the algorithm, mixed models can be applied (for example, to test the overall treatment effect and to construct confidence intervals).

Results: We apply this method to perform a pooled analysis for one of our studies, where we have continuous outcomes regarding symptoms improvement after product intake, Quality of Life (QoL) and product appreciation parameters, expressed in terms of summary statistics coming from sites in different countries. Due to the sensitive nature of the study and its results, simulated data similar to the real ones will be presented. All the parameters showed considerable improvements after product intake.

Conclusion: Thanks to the great advantages of this method, we managed to obtain quick access to raw pseudo-IPD, to employ the flexibility of mixed models, to take advantage of the possibility of modelling the within-study residual variances and to produce reliable estimates of the treatment effect of our product in a time-efficient manner, avoiding complications related to privacy and lack of availability of complete data.

POSTER SESSION A-09:
Meta-analysis

P-A09-12

Single and dual hormone artificial pancreas systems in patients with type 1 diabetes: An overview of systematic reviews

Stavros Kanakis[1], Anna Bettina Haidich[1], Thomas Karagiannis[2], Eirini Pagkalidou[1], Naoum Kalfopoulos[1], Persefoni Talimtzi[1]

1 Department of Hygiene, Social & Preventive Medicine and Medical Statistics, Aristotle University of Thessaloniki, Thessaloniki, Greece
2 Clinical Research and Evidence-Based Medicine Unit, Second Medical Department, Aristotle University of Thessaloniki & Diabetes Centre, Second Medical Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

Background: Type 1 diabetes is a chronic condition where the pancreas produces little to no insulin which is essential for regulating blood sugar (glucose) levels in the body. Artificial pancreas represents an advancement in the management of type 1 diabetes, combining an insulin pump and continuous glucose monitoring with a control algorithm to deliver insulin in a glucose responsive manner as a single (insulin only) or dual hormone (insulin and glucagon) artificial pancreas system. Several artificial pancreas systems have been developed, and their safety and efficacy have been evaluated in many studies, and even in systematic reviews. The purpose of this overview is to identify, critically appraise and summarize systematic reviews (with or without a meta-analysis) that assess artificial pancreas systems (either single or dual hormone).

Methods: An extensive literature search will be conducted in PUBMED, the Cochrane Database of Systematic Reviews (CDSR), Epistemonikos, as well as reference lists of other systematic reviews until the end of March 2024. Systematic reviews of randomized controlled trials or observational studies evaluating a single or dual hormone artificial pancreas system in people with type 1 diabetes will be included. The primary study overlap among all eligible systematic reviews/meta-analyses will be assessed by producing a citation matrix and by calculating the corrected covered (CCA) area, via the 'ccaR' package in R. Eligible systematic reviews will be critically appraised using the AMSTAR 2 risk of bias tool with the 'amstar2Vis' package in R.

Results: A descriptive table will summarize findings extracted from eligible systematic reviews, including primary and secondary outcomes, statistical model used for the meta-analysis, effect sizes with 95% confidence interval (CI) for each outcome, p-value, statistics for heterogeneity and additional analyses (e.g., subgroup or sensitivity analysis, meta-regression); instrument used for quality assessment of the primary studies and rating; whether the GRADE approach was used per outcome and the rating and methods of detecting publication bias. Furthermore, transparency characteristics will be recorded, including protocol availability, searched databases, search strategy availability, reporting guidelines checklist, study selection process, data collection process, publicity of detailed statistical code and data, authors' conflicts of interest, funding and data sharing policy.

Conclusion: The present overview will provide high-quality evidence to support informed decision-making and may contribute to future guideline development in type 1 diabetes and use of artificial pancreas.

POSTER SESSION A-09:
Meta-analysis

P-A09-13

Meta-Analysis models with group structure for pleiotropy detection at gene and variant level by using summary statistics from multiple datasets

Benoit Liquet[1,2]
1 Macquarie University, Sydney, Australia
2 University of Pau et Pays de L'Adour, Pau, France

Background/Introduction: Genome-wide association studies (GWAS) have highlighted the importance of pleiotropy in human diseases, wherein a single gene can influence multiple unrelated traits. Exploring shared genetic risk factors across diverse diseases can enrich our comprehension of these conditions by identifying novel genes and biological pathways involved. Moreover, with an expanding array of GWAS summary statistics accessible to the scientific community, harnessing these discoveries across multiple phenotypes could reveal novel pleiotropic associations.

Methods: Current selection methods examine pleiotropic associations individually at either the genetic variant or gene scale, thus unable to consider all genetic information simultaneously. To overcome this limitation, we propose a novel approach termed Meta-PSG (Meta-analysis model adapted for Pleiotropy with Sparse Groups). This method employs a penalized multivariate meta-analysis approach tailored for pleiotropy and incorporates the group structure information inherent in the data to select relevant variants and genes (or pathways) from the entire genetic dataset. To achieve this, we implemented an alternating direction method of multipliers (ADMM) algorithm.

Results: We evaluated the method's performance against other benchmark meta-analysis approaches such as GCPBayes, PLACO, and ASSET using various types of summary data as inputs. Simulations demonstrate that our method, Meta-PSG, surpasses PLACO and ASSET in retrieving pleiotropic associations at both SNP and gene levels. We achieve comparable performance to GCPBayes, but our method exhibits significantly faster computation. To exemplify the application of our method, we explore the shared genetic effects between thyroid cancer and breast cancer within candidate pathways.

Conclusions: The proposed meta-analysis approaches offer flexibility and computational efficiency. They enable the incorporation of biological structures in genetic data to enhance the detection power of pleiotropic associations. The design and flexibility of these methods are likely to be valuable across various genomics applications (such as gene-expression studies) and other fields for detecting non-zero effects of potentially different directions for response variables involving structured covariates.

POSTER SESSION A-09:
Meta-analysis

P-A09-14

Meta-analysis models relaxing the random effects normality assumption: A simulation study

Kanella Panagiotopoulou[1], Theodoros Evrenoglou[2], Christopher H Schmid[3], Silvia Metelli[1], Anna Chaimani[1]

1 Université Paris Cité, Center of Research in Epidemiology and Statistics, Inserm, Paris, Paris, France
2 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg im Breisgau, Germany
3 Department of Biostatistics, School of Public Health, Brown University, Providence, United States

Background: Random effects meta-analysis is widely used for synthesizing the studies of a systematic review under the assumption that the underlying effects come from a common normal distribution. However, the presence of large heterogeneity or outlying studies may violate this assumption and invalidate the results. Alternative options have been suggested but never used in published meta-analyses.

Methods: We first conducted a methodological systematic review to identify articles that proposed alternative meta-analysis models assuming non-normal distributions for the random effects, such as skewed or semi-parametric distributions. Subsequently, we performed a simulation study to evaluate the performance of the identified models and to compare them with the normal model. We considered 22 scenarios varying the amount of heterogeneity between studies, the number of included studies, and the shape of the true distribution: normal, skew-normal, and mixture of two normal distributions. For each scenario, we generated 1000 meta-analyses datasets. Additionally, the different models were applied to an exemplar meta-analysis of 23 heterogeneous observational studies assessing the relationship between metabolic syndrome and psoriasis.

Results: We identified in total 13 articles suggesting 10 alternative models that can be classified in three broad categories: models based on long-tail and skewed distributions, on mixtures of (normal) distributions, and on Dirichlet process priors. We compared 7 models in our simulation study implemented in the Frequentist or Bayesian framework. Results revealed that the bias of both the mean treatment effect and heterogeneity is substantial in presence of high heterogeneity regardless of the model used. However, as the true distribution moved away from the normal, the model based on the Dirichlet process resulted in coverage probabilities closer to the nominal level than the other evaluated models. In such scenarios, the same model yielded on average less biased estimates of the underlying effects of the studies and predictive distributions closer to the true distributions. In the real example, all applied models gave similar estimates of the mean treatment effect but the Dirichlet process model better captured the heterogeneity across studies and automatically identified clusters of studies with similar characteristics.

Conclusion: In meta-analyses, when substantial heterogeneity among studies is suspected or outlying studies are present, focusing on the mean treatment effect may lead to spurious conclusions. In such cases, the plausibility of the normality assumption should be assessed thoroughly and alternative, more flexible models should be used.

POSTER SESSION A-09:
Meta-analysis

P-A09-15 Population-adjusted indirect comparisons: A simulation study

Arnaud Serret-Larmande[1,2], Jérôme Lambert[3,4], David Hajage[1,2]
1 Sorbonne University, INSERM, UMR-1136 Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France
2 AP-HP Pitié-Salpêtrière Hospital – Public Health Department, Pharmacoepidemiology Center, Paris, France
3 Université Paris Cité, INSERM, UMR-1153 ECSTRRA Team, Paris, France
4 AP-HP Saint Louis Hospital – Biostatistics and Medical Informatics Department, Paris, France

Introduction: Population-Adjusted Indirect Comparisons (PAICs) allow for the comparison of treatments initially evaluated in different clinical trials when access to individual data is impossible for one of the treatments of interest. Their practical applications increased significantly recently 1. PAICs are divided into "anchored" and "unanchored" forms (with and without a control arm, respectively). Despite their appeal, these methods have limitations, and their use is subject to a large number of methodological choices leading to significant results' variability: type of PAIC (based on regression model or propensity score), population to include, identification of confounding factors taken into account. While previous simulation studies compared various aspects of these methodologies 2,3, current literature lacks a systematic comparison of anchored versus unanchored PAICs and of their performance against methods using complete individual data.

Methods: We conducted Monte Carlo simulations to assess these methods under diverse confounding scenarios in treatment-outcome associations, varying by confounding factors and their distributions, treatment effects, and sample sizes. For each simulation, we compared the treatment effect estimators evaluated by: - two PAIC methods, Matching-Adjusted Indirect Comparisons and Simulated Treatment Comparison - their individual data counterparts: propensity score weighting and multivariable outcome regression model Performance was evaluated using bias, variability ratio, RMSE, and coverage rate. Results Under perfect theoretical conditions (correct specification of the adjustment model), PAICs offer similar performances to models with individual data. Anchored comparisons show a reduction in bias in the presence of confounding factors not taken into account in the analysis, at the cost of an increase in the variability of the estimation. The impact of non-normal confounding distributions varied by estimator.

Conclusion: Our results show that unanchored adjusted comparisons present interesting theoretical performances but suffer from notable bias when the model is misspecified, i.e., in the absence of one or more confounding factors. As a common situation in practice 1, these results underline the importance identifying these factors when implementing these methodologies.

References: *[1] Serret-Larmande A, Zenati B, Dechartres A, Lambert J, Hajage D. A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. J Clin Epidemiol. 2023;163:1-10. doi:10.1016/j.jclinepi.2023.09.004 [2] Remiro Azócar A. Population-Adjusted Indirect Treatment Comparisons with Limited Access to Patient-Level Data. PhD thesis. University College London; 2021. [3] Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Stat Med. 2020;39(30):4885-4911. doi:10.1002/sim.8759*

POSTER SESSION A-09:
Meta-analysis

P-A09-16

A versatile software package for combining dependent or independent p-values

Evgenia Nikolitsa[1], Panagiota Kontou[2], Pantelis Bagos[1]
1 Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece
2 Department of Mathematics, University of Thessaly, Lamia, Greece

Introduction: Combining P value s is a n important technique in various fields. For example, in genome wide association studies (GWAS), different Single Nucleotide Polymorphisms (SNPs) within a gene are combined in the so-called gene-based testing methods for integrating data from rare variants. Other similar approaches are encountered in multiple traits analysis or in more general multi omics methods. In this work, we present a software tool that implements a wide range of methods for combining p values, both independent and dependent ones.

Methods: We incorporate standard methods which in general are applicable for combining independent p values such as the methods of Fisher, Stouffer Edgington the inverse chi squared method, the logit method and the binomial test. Regarding dependent p values we implement the Cauchy Combination Test (CCT), the MinP method, the combined MinP CCT MinP (MCM) and CCT MinP CCT (CMC) methods, the Brown's method along with its modifications, and the Harmonic Mean P value (HMP) test. We also implement several well-known methods that adjust for the number of effective tests in the methods that assume independence The tool is implemented in Python.

Results: The program takes as input a text file of n rows and m columns, plus the headers, which contain the p values or z scores to be combined. Different rows can correspond to different SNPs or genes from a GWAS or probes/genes from transcriptomics studies. The different columns can correspond to different statistical tests, different traits, different SNPs within a gene and so on, depending on the setting. In case the same SNP gene appears in multiple rows a meta-analysis can be performed prior to combining the different p values from different columns.

Conclusion: The tool presented in this work is very fast and easy to use, requiring only minimal input. It offers a useful resource for combining both independent and dependent p values, responding to diverse analytical needs for researchers in a variety of fields Depending on the input data it can be used for gene-based testing or for analysis of multiple traits in GWAS, or for combining diverse multi omics data such as those of a TWAS, a co-localization or an RNAseq study. Compared to other similar packages (like poolr or metap) offers the largest possible collection of analytical methods. The source code will be available in www.github.com/pbagos

POSTER SESSION A-10:
Causal Inference

P-A10-01

Rest-and-jump-and then? – Identifying changes in gene expression

Lucia Ameis, Kathrin Möllenhoff
University of Cologne, Cologne, Germany

Background/Introduction: Evaluating the change in gene expression is a common goal in many research areas, such as in toxicological studies as well as in clinical trials. In practice, the analysis is often based on multiple t-tests evaluated at the observed time points. This severely limits the accuracy of determining the time points at which the gene changes in expression. Even if a parametric approach is chosen, the analysis is often restricted to identifying the onset of an effect.

Methods: In this talk, we propose a parametric method to identify the time frame where the gene expression significantly changes. Precisely, we achieve this by fitting a parametric model (e.g. a monotonic sigmoidal 4pLL or a non-monotonic beta model) to the time-response data and constructing a confidence band for its first derivative. The confidence band is derived by a flexible two step bootstrap approach, which can be applied to a wide variety of possible curves. Our method focuses on the first derivative, since it provides an easy to compute and reliable measure for the change in response. It is summarised in terms of a hypothesis test, such that rejecting the null hypothesis means detecting a significant/relevant change in gene expression.

Results: We demonstrate the validity of our approach by means of a simulation study and present an application to gene expression data of mice from a study investigating the effect of a western diet on the progression of non-alcoholic liver fatty disease. Precisely, we show how our approach allows for the identification of a rest-and- jump gene, that is a gene that has a delayed response to the diet and does not change again after the first jump in expression.

Conclusion: The suggested approach is widely applicable to many parametric models in a variety of situations. Due to being formulated as a hypotheses test and constructed from a parametric model, changes in gene expression can be interpreted based on statistical significance and over the whole study period without limiting to the time points at which measurements were taken.

POSTER SESSION A-10:
Causal Inference

P-A10-02

Stepwise, forward, backward selection and cross-validation: How to choose the best predictors?

Alice Bonomi, Arianna Galotta, Nicolò Capra, Francesco Mattio, Giancarlo Marenzi
IRCCS Centro Cardiologico Monzino, Milan, Italy

Background/Introduction: In the world of clinical research, it is often necessary to implement statistical models to identify independent predictors. In the literature, there are different models to use and different criteria to identify the most important predictors. Assuming the outcome is a dichotomous variable, the most commonly used models are stepwise, forward, and backward with logistic regression. Do the three approaches identify the same predictors? Are the identified predictors really the most important ones? To answer these questions, we implemented several logistic regression and cross-validation models.

Methods: Using a cohort of 2370 patients with myocardial infarction and in-hospital mortality as the event (4%), we implemented 3 logistic models to select the most important predictors: stepwise, backward, and forward, after imputation of missing data with proc MI. We also implemented k-fold cross-validation with k=10. The dataset was split into 10 equal parts; the model was trained and validated on different combinations of data (split 50% vs 50% and 70% vs 30%), each time repeated 200 times. Finally, the different models were compared using AIC and AUC values. The p-value level for selection is 0.10, while the significance level is 0.05. All analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA).

Results: From 12 possible predictors, 6 predictors (age, FE, glycaemia, smoking, EGFR, hypercholestemia) were identified by logistic regression, independent of the selection technique. After cross-validation, age, FE, glycaemia, and hypercholestemia were selected when the database was split 50% vs 50% and we used stepwise, backward as techniques of selection. In other scenarios, age, FE, glycaemia were selected. The models obtained with the different models have slightly different performances, identifying as the best model (lowest AIC (1193) and highest AUC (0.85)) the model selected without cross-validation, followed, without statistically significant difference, by the model selected with 50 vs 50 cross-validation (AIC value: 1206 and AUC value: 0.85), while the AUC of the model selected with 70 vs 30 cross-validation is significantly lower (0.84), and the AIC value was 1232.

Conclusion: After imputation of missing data, there were no significant differences between the different selection techniques. In terms of AIC and AUC, selection without cross-validation was comparable to selection with 50 vs 50 cross-validation, resulting in the most accurate models for selecting independent predictors.

POSTER SESSION A-10:
Causal Inference

P-A10-03 Simple approaches for portfolio quantitative decision-making

Hugo Hadjur[1], Véronique Robert[2], Jean-François Collin[3], Gaëlle Saint-Hilary[1]
1 Saryga, France
2 Daiichi-Sankyo Oncology, France
3 Ceva, France

Background/Introduction: An effective pharmaceutical portfolio management is essential, but complex as it contains many assets in various development stages, with different success probabilities, timelines, costs, and sales potential. Statisticians can play a pivotal role aiding management teams to make more informed decisions, by integrating risk assessment and probabilistic analyses. In this presentation, we focus on two simple, yet powerful, quantitative approaches for portfolio decision-making.

Methods: First, we present predictions of marketing authorizations over time. Typically, an average yearly count of marketing authorizations is derived by summing the success probabilities of individual projects. However, this simplistic approach overlooks uncertainties. Instead, using Monte Carlo methods to simulate potential portfolio realisations enables the calculation of probabilities of reaching some target number of marketing authorisations over time. It permits to identify the portfolio's strengths and weaknesses, and to trigger strategic actions such as licensing-in or partnerships. The second approach integrates financial considerations by deriving the portfolio's risk-value profile. The Net Present Value (NPV) is a common metric used to assess a project's profitability. By assessing the NPV distribution of all projects within the portfolio, and employing again simulations, we estimate the portfolio's NPV distribution and calculate probabilities of attaining predefined target values. This approach empowers decision-makers to comprehensively evaluate a portfolio's value and risk, facilitating informed investment decisions.

Results/Conclusion: These methods illustrate that even with basic data, simple predictions, and summary statistics that account for uncertainties, a more meaningful portfolio assessment is possible. This aids decision-makers in making better-informed choices, and the simplicity of these quantitative approaches makes them particularly appealing to top managements.

P-A10-04 Quantitative decision making in the vaccines world: A new paradigm for risk assessment

Leroy Jide Ovbude, Francesca Solmi, Valentino Conti, Fabian Tibaldi, Olivier Collignon, Tapiwa Ganyani, Matt Carli
GSK, Wavre, Belgium

Decision making in vaccine development is crucial for numerous reasons. Effective decision making not only ensures that vaccines are developed quickly to address health concerns and protect individuals from diseases but also helps ensure that trial designs are appropriate to answer the scientific questions, data analysis is robust, and conclusions obtained from the clinical study results are valid and reliable.

Drug and vaccine development is known for being risky, lengthy, and expensive. The average cost for a new vaccine exceeds $1-2 billion, and development typically spans more than 10 years. Attrition rates can reach 90% from Phase 1 to Phase 3. To get the delivery of efficient medicines and vaccines to patients and enhance return on investment, the pharmaceutical industry has increasingly adopted quantitative methodologies. These approaches make data driven decisions routine and improve the accuracy of risk prediction.

GlaxoSmithKline (GSK) has implemented the Quantitative Decision Making (QDM) framework designed evaluate the risk across the different clinical development plans and to optimize vaccine development.

In this poster, the main concepts used in QDM assessment within the pharmaceutical industry will be discussed focusing on vaccine development and their specificities, such as reactogenicity, immunogenicity, correlate of protection etc. We will share how we are using innovative approaches in the application of QDM framework to support vaccine research within GSK.

POSTER SESSION A-10:
Causal Inference

P-A10-05

Multiple testing in statistical inference: A historical overview

Thomas Forstner
Johannes Kepler University, Linz, Austria

Introduction: Multiple testing is a common issue in medical research and occurs when researchers perform multiple statistical comparisons simultaneously, leading to an increased type I error. Over the years, various methods have been proposed to address this issue. This poster traces the historical milestones and key contributors in the field of multiple testing.

Methods and Results: The earliest attempts to address the challenges of multiple testing can be traced back to the year 1935. Bonferroni correction, named after Carlo Emilio Bonferroni, was among the first methods proposed to control the familywise error rate by adjusting the type I error based on the number of tests performed. This was a fundamental step in acknowledging the importance of multiple testing in statistical inference. The Bonferroni correction corrects for type I inflation, but at the cost of increasing the type II error, which results in a loss of statistical power. In the year 1979, Sture Holm introduced a version of the Bonferroni correction that provides a more powerful alternative, the Holm-Bonferroni correction The Holm-Bonferroni correction prioritises the adjustment of p-values in a sequential order, resulting in an increased power compared to the original Bonferroni correction.

The development of multiple testing procedures continued in the 1950s with the introduction of more methods aimed at controlling the familywise error rate. Henry Scheffé and John Wilder Tukey published groundbreaking work addressing the problem of multiple testing. Tukey's Honestly Significant Difference test and Scheffé's method emerged as prominent techniques, offering alternatives to Bonferroni-based corrections and providing more tailored control over the type I error.
As the focus shifted towards addressing the limitations of familywise error rate control, the concept of the false discovery rate (FDR) gained prominence. In 1995, Yoav Benjamini and Yosef Hochberg published a groundbreaking approach to control the FDR, enabling researchers to balance the trade-off between type I error and type II error more effectively. A year later, in 1996 the first international conference on multiple comparisons was held in Tel Aviv, emphasising the growing significance of this topic.
In recent years, Frank Bretz and Carl Fredrik Burman have developed graphical approaches for multiple comparisons, allowing more flexible designs

Conclusion: From the early Bonferroni correction to more sophisticated methods, researchers have strived to strike a balance between controlling type I error and maintaining statistical power. Despite the advancements in multiple testing procedures, challenges persist, particularly in the context of high-dimensional data and complex study designs.

POSTER SESSION A-10:
Causal Inference

P-A10-06

Overcoming model uncertainty - How equivalence tests can benefit from model averaging

Niklas Hagemann[1,2], Kathrin Möllenhoff[1]
1 Institute of Medical Statistics and Computational Biology, Faculty of Medicine, University of Cologne, Cologne, Germany
2 Mathematical Institute, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Background/Introduction: A common problem in numerous research areas, particularly in clinical trials, is to test whether the effect of an explanatory variable on an outcome variable is equivalent across different patient groups, e.g. based on gender, age, or treatments. In this regard, equivalence is usually assessed by testing whether the difference between the groups does not exceed a pre-specified equivalence threshold. Such tests are classically based on single quantities, e.g. the mean, the AUC or other values of interest, or, if differences depending on a particular covariate are observed, on some distance of regression curves over an entire covariate range. However, fitting a regression model to the data bases on the assumption that the true underlying regression model is known. Hence, these approaches might not be robust with regard to model misspecification and, consequently, suffer from problems like inflated type I errors or reduced power.

Methods: In this talk, we propose a flexible extension of such methodology that uses model averaging in order to overcome this assumption and make the test applicable under model uncertainty. Model averaging is introduced based on smooth AIC weights. In order to ensure numerical stability, we propose a testing procedure which makes use of the duality between confidence intervals and hypothesis testing.

Results: We demonstrate the validity of our approach by means of a simulation study. A dose-response case study demonstrates the practical relevance of the approach.

Conclusion: Our suggested approach is widely applicable and, therefore, overcomes the requirement of knowing the true underlying regression model. Simulations show its validity as well as its sufficiently high power.

POSTER SESSION A-10:
Causal Inference

P-A10-07

A simple-to-use R package for mimicking study data by simulations

Georgios Koliopanos[1], Francisco Ojeda[2], Andreas Ziegler[1,2,3]
1 Cardio-CARE, Medizincampus Davos, Switzerland
2 Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
3 School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Background: Data protection policies might prohibit the transfer of existing study data to interested research groups. To overcome legal restrictions, simulated data can be transferred that mimic the structure but are different from the existing study data.

Objectives: The aim of this work is to introduce the simple-to-use R package Mock Data Generation (modgo) that may be used for simulating data from existing study data for continuous, ordinal categorical, and dichotomous variables. Methods The core is to combine rank inverse normal transformation with the calculation of a correlation matrix for all variables. Data can then be simulated from a multivariate normal and transferred back to the original scale of the variables. Unique features of modgo are that it allows to change the correlation between variables, to perform perturbation analysis, to handle multicenter data, and to change inclusion/exclusion criteria by selecting specific values of one or a set of variables. Simulation studies on real data demonstrate the validity and flexibility of modgo.

Results: modgo mimicked the structure of the original study data. Results of modgo were similar with those from two other existing packages in standard simulation scenarios. modgo's flexibility was demonstrated on several expansions.

Conclusion: The R package modgo is useful when existing study data may not be shared. Its perturbation expansion permits to simulate truly anonymized subjects. The expansion to multicenter studies can be used for validating prediction models. Additional expansions can support the unraveling of associations even in large study data and can be useful in power calculations.

POSTER SESSION A-10:
Causal Inference

P-A10-08

Three questions about ordinal outcomes in neurological trials

Yongxi Long, Ewout Steyerberg, Erik van Zwet
Leiden University Medical Center, Netherlands

Background: Ordinal scales such as the modified Rankin Scale are widely used as outcome measures in neurological clinical trials. The proportional odds model, which is sometimes called shift analysis, is often used to analyze such outcomes. There has been debate whether violation of the proportional odds assumption should preclude its use. Other statistical issues (such as power) have also received much attention. Here, we take a step back. We suggest that the analyst should start by asking three basic questions before reaching for the statistical toolbox.

Methods & Results: The first question is: Is a single measure of the treatment effect sensible? Qualitative violation of the proportional odds assumption happens when for example there is an increase in favorable outcomes at the expense of higher mortality. In such cases, summarizing the effect of the treatment into a single number will be misleading. If the answer to the first question is "yes", then one should ask: Is it possible to assign numerical values (utilities) to the categories? This is an uncomfortable question because utilities will always be subjective to some extent. However, if the "steps" between adjacent categories have different clinical importance, it would be an oversimplification and a loss of information to ignore that. A possible remedy is to lump similar categories together, although that will reduce statistical power. If the steps are roughly equally spaced, then we may ask: Which summary measure should be preferred? The main choices are the common odds ratio (from a proportional odds model) and the Win statistics (associated with Mann-Whitney U test, e.g., Win odds, Win ratio). We illustrate these three questions with a large stroke trial which showed benefit of intraarterial treatment.

Conclusion: Ordinal scales capture neurological functionality of patients in a more detailed way than a dichotomy such as good versus bad. We suggest that three key questions need to be addressed to decide on the most appropriate statistical analysis. In particular, we want to challenge the idea that an ordinal analysis obviates the need to think about the relative clinical importance of the categories. Therefore, we recommend that researchers and trialists make an effort to attach interpretable numbers to ordinal categories whenever possible.

POSTER SESSION A-10:
Causal Inference

P-A10-09

An extension of the win odds for ordinal repeated measurements

Yongxi Long, Ewout Steyerberg, Erik van Zwet
Leiden University Medical Center, Netherlands

Background: The win odds is an effect size measure closely related to the Mann-Whitney U test. It considers all possible pairwise comparisons between the treated and the control patients, with ties split evenly. Initially developed for analyzing composite endpoints, the win odds has recently received increasing interest for the analysis of ordinal outcomes in randomized controlled trials. One reason for this is that many consider it to be more intuitive than the common odds ratio from the proportional odds model. The win odds can be adjusted for covariates by the probabilistic index model [1]. Here, we aim to extend this model for repeated measurements.

Method & Results: We modified the estimation equations of the probabilistic index model to account for temporal correlation. Parameter estimation can be conveniently done via some data re-structuring and the R package geepack. We implemented a sandwich-type estimator to estimate the variance-covariance matrix separately. Simulations showed that the estimation for win odds is consistent and the coverage of confidence intervals is close to nominal. The sandwich estimator is relatively robust to the misspecification of the working correlation structure.

Conclusion: This extension of the win odds offers an attractive summary measure to capture the effect of treatment during follow-up. Implemetation can be readily done in existing R packages with some minor additonal scripting.

Reference: *[1] Thas, O., et al., Probabilistic index models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2012. 74(4): p. 623-671.*

POSTER SESSION A-10:
Causal Inference

P-A10-10

Incorporation of patient and public involvement in statistical methodology research: Developing resources for both researchers and public contributors

Rachael Stannard, Aiden Smith, Laura Gray, PPI-SMART
University of Leicester, United Kingdom

Background: Monitoring trends over time of cancer incidence, mortality and survival is vital for the planning and delivery of health services, and in the evaluation of diagnostics and treatment at the population level. Comparisons are often made between populations and over calendar time to explore inequalities in outcomes. The disruption to the routine delivery of the health service due to the COVID-19 pandemic has largely impacted cancer diagnosis and treatment pathways. We must consider how to fairly compare cancer metrics over time to continue monitoring the progress against cancer.

Methods: We use the Pohar-Perme estimator to compare age-standardised 1-year relative survival in April-December 2016-2019 with the same period in 2020 and 2021 across a range of cancer sites in the UK. We consider survival for all patients combined and stage-specific survival. Given the observed shift to more advanced stage at diagnosis due to the pandemic, especially among screened cancers, we expect the decline in stage-specific survival to be less than the decline in combined survival. To enhance the understanding of survival differences driven by changes in stage distribution due to the pandemic, we use stage-standardisation to make fair comparisons over calendar time. There are also changes in the proportion of missing stage information. The missing data mechanisms may have changed and should be accounted for in the analysis. There are changes in the distribution of other key variables affecting cancer survival such as age, which may impact survival estimates and are associated with COVID-19 outcomes.

Results: Comparing April-December 2016-2019 with the same period in 2020, we see a shift towards more advanced stage at diagnosis, and changes in the proportion of missing stage at diagnosis. There is a large increase in stage IV colorectal cancer in 2020 compared to previous years. These changes in stage distribution are partially responsible for the decline in 1-year survival that we see across several cancer sites. Combined survival is more affected than stage-specific survival.

Conclusion: Understanding the contribution of differences in stage at diagnosis to differences in observed survival due the COVID-19 pandemic is key to making fair comparisons over calendar time. Investigating both stage at diagnosis and survival simultaneously is essential. Consideration of stage-specific 1-year survival, combined 1-year survival and stage-standardised 1-year survival can be used to achieve this goal. We expect these same approaches will continue to aid the interpretation of trends in survival over time as longer-term data becomes available.

POSTER SESSION A-10:
Causal Inference

P-A10-11

Robust estimation of overdispersion parameter for outlier detection in clinical audits

Anqi Sui, Gareth Ambler, Rumana Omar, Menelaos Pavlou
University College London, London, United Kingdom

Introduction: Monitoring clinical performance of health-care units is nowadays an essential aspect of national audits. Interest often lies in units whose performance (e.g. in-hospital mortality) diverges substantially from an expected performance. A commonly used method is the common mean model (CMM) [1], which assumes a single true performance for all units and any variability in the observed performance is due to chance. This is often not a tenable assumption, because the observed variability is often higher than that assumed (overdispersion) due to e.g. clustering of patients within units or imperfect risk adjustment. An overdispersion correction can be applied; ideally this should be applied after the removal of potential outliers, otherwise the allowable variability will be too high, hindering the detection of true outliers. To minimise the effect of potential outliers on the estimation of overdispersion parameter, winsorisation, which 'shrinks in' most extreme values based on a given winsorisation percentage, can be used; however, choosing an appropriate winsorisation percentage can be challenging. To overcome this challenge, we propose a robust overdispersion correction, which is based on the effect of sequential removal of units and assumptions for the degree of the clustering of patients.

Methods: Simulation is used to evaluate the performance of the proposed approach and compare it with the existing methods (the simple CMM, the overdispersion correction, and winsorisation with different winsorisation percentages), in scenarios with varying number of units, intra-class correlation (ICC), and proportion and extremeness of outliers. A higher ICC corresponds to a stronger violation of the CMM assumption. Performance measures include false positive rate (FPR) (proportion of non-outliers incorrectly identified as outliers) and sensitivity (proportion of true outliers correctly identified).

Results: The FPR of the simple CMM is too high, indicating overdispersion correction is necessary. When there are outliers, winsorisation is needed, otherwise the FPR Word count: 400 1 Abstract for ISCB 2024 and sensitivity tend to be low. The FPR and sensitivity are sensitive to winsorisation percentage. An appropriate winsorisation percentage depends on the actual proportion of outliers, which is typically unknown. The proposed robust overdispersion correction shows better and more stable FPR and sensitivity than the other methods in the scenarios investigated.

Conclusions: Overdispersion correction is necessary when applying the CMM. When outliers are present, winsorisation is needed in the overdispersion correction, but its performance is sensitive to the unknown proportion of outliers. Alternatively, the proposed robust overdispersion correction can be used.

References: *[1] Spiegelhalter, David J. "Funnel plots for comparing institutional performance." Statistics in medicine 24.8 (2005): 1185-1202.*

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-01 Epilepsy survival analysis: The integration of machine learning with genetic data from the SANADII Dataset

Bushra Alrashdi, Anna Fowler, Andrea Jorgensen, Tony Marson, Tsegay Gebru
Univeristy of Liverpool, Liverpool, United Kingdom

Introduction: Epilepsy is defined by the occurrence of recurrent seizures due to certain abnormalities in the brain function, impacting various aspects of health and day-to-day life. Studies suggest that genetic factors may impact the frequency and the severity of seizures. The SANADII dataset contains clinical details of 479 patients, along with whole exome sequencing data and seizure timings. Seizure timings provide a time-to-event outcome. However, the traditional survival analysis methods can not efficiently enable complex analysis within high-dimensional genetic data, therefore, this study utilises machine learning techniques to improve the predictive capabilities of survival models in order to improve understanding of factors influencing the intervals between epileptic seizures.

Methods: The methodology encompasses several key stages: data preprocessing, feature selection, and the application of cox model and machine learning models tailored for survival analysis. The core of the analysis involves applying machine learning models, including the Cox proportional hazards model augmented with LASSO regularisation, Random Survival Forest (RSF), Gradient Boosting Machine (GBM), and Recursive Partitioning and Regression Trees (RPART). These models are chosen for their ability to handle the complexities of high-dimensional genetic data. To assess the performance of these models the concordance index (C-index) and Brier scores are the main model evaluation measures in this study that would provide insights into the models' predictive accuracy and reliability.

Results: This study demonstrates the applicability of machine learning approaches for genetic datasets with time-to-event outcomes and identifies genetic factors which may be associated with improved treatment response in epilepsy. The machine learning models have shown variability in their ability to handle genetic profiles associated with epilepsy. Specifically, certain models have identified some genetic markers that could impact the time to the next seizure. This highlights the role of machine learning to enhance predictive modelling. In addition, accuracy measurements have shown varying results across the models, indicating the different capabilities of each model in capturing the complexities of epilepsy. This provides insights into the strengths and limitations of the models.

Conclusion: The integration of machine learning techniques with the SANADII genetic dataset provides a potential advancement in epilepsy research. Models such as the Cox proportional hazards model with LASSO regularisation, RSF, GBM, and Recursive Partitioning and Regression trees, address some challenges of high-dimensional genomic data. Enhanced predictive accuracy and interpretability highlight the effectiveness of these computational approaches. The identification of prognostic markers through this methodology offers insights into the genetic foundations of epilepsy. Finally, the findings advocate for a tailored approach and more personalised medical treatment, emphasising the need for personalised therapeutic strategies based on individual genetic profiles.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-02

Comparative analysis of ChatGPT and human performance in post classification and sentiment analysis

Malamatenia Panagiotou[4], Vasiliki Bountziouka[1,2,3], Konstantinos Gkatzionis[4], Efstathios Kaloudis[1]
1 Computer Simulation, Genomics and Data Analysis Laboratory, Depart. of Food Science and Nutrition, University of the Aegean, Lemnos, Greece
2 Population, Policy and Practice Research and Teaching Dept, GOS Institute of Child Health, UCL, London, United Kingdom
3 Dept. of Cardiovascular Science, College of Life Science, University of Leicester, Cardiovascular Research Centre, Leicester, United Kingdom
4 Laboratory of Consumer and Sensory Perception of Food & Drinks, Dept. of Food Science and Nutrition, University of the Aegean, Lemnos, Greece

Background/Introduction: Artificial intelligence (AI) is a rapidly evolving field that offers unparalleled opportunities for progress in many healthcare fields. One of AI's most exciting advances is the emergence of Large Language Models (LLMs), sophisticated AI systems trained on vast amounts of text data, capable of generating human-like text and understanding language paterns with contextual relevance. ChatGPT is a LLM based on the Generative Pre-trained Transformer (GPT) architecture, developed by OpenAI, that was received with great enthusiasm following its release to the public in November 2022. If prompted appropriately, ChatGPT can understand and respond to multiple languages generating refined and highly sophisticated responses in various scientific fields such as healthcare and food science. In this study we aimed to test ChatGPT's performance, against human responses, as a natural language processing and sentiment analysis tool for food-related social media posts.

Methods: Instagram was selected as the social media platform due to its widespread use by marketers globally, particularly in Greece where users engage with companies frequently. Utilising the Apify platform for web scraping, posts with snail-related hashtags were collected, resulting in a dataset of 1,769 posts, after removing duplicates and irrelevant posts, and further refined by eliminating non-Greek and non-English content. The food-relatedness criterion was applied using ChatGPT 3.5 API, and the process was automated with Python. Subsequently, manual extraction was performed. Sentiment analysis to categorize posts into positive, neutral, and negative sentiments was also conducted using ChatGPT and manually. Kappa statistic was used to assess the agreement between ChatGPT and human.

Results: Both ChatGPT and human identified 44% (n=773) food-related and 836 (47%) not food-related posts, thus reaching an overall correct classification rate of 91%. Beyond chance agreement between human and ChatGPT classification was substantial (kappa=0.82, 95% CI: 0.79-0.85). Additionally, sentiment analysis revealed high agreement between ChatGPT and human evaluations, with 61% (n=1,075) of posts identified as positive and 37% (n=654) as neutral, resulting in an overall correct classification rate of 98% and near-perfect agreement (kappa=0.95, 95% CI: 0.94-0.97). However, ChatGPT's performance in identifying "Greeklish" was inconsistent.

Conclusion: This study highlights the remarkable performance of ChatGPT compared to humans in post classification and sentiment analysis tasks, demonstrating high agreement. However, there is a need for further refinement and adaptation to address specific linguistic challenges. This study is an exemplar of ChatGPT's capabilities, that are also transferable in other settings, although areas for improvement in natural language processing systems still exist.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-03

Embedding models for multimorbidity detection: A deep learning framework using healthcare administrative databases

Emanuele Koumantakis[1,2,3], Alessia Visconti[2], Sara Urru[4], Paola Berchialla[2]

1 Department of Public Health and Pediatrics, University of Torino, Torino, Italy
2 Department of Clinical and Biological Sciences, University of Torino, Orbassano, Italy
3 Post Graduate School of Medical Statistics, University of Torino, Torino, Italy
4 Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac,Thoracic, Vascular Sciences and Public Health, University of Padova, Padova, Italy

Introduction: Traditional machine learning and statistical models often struggle with high-dimensional data and varying time intervals between events, common characteristics of Healthcare Administrative Databases (HADs). Deep Learning (DL) approaches are well-suited to handle the vast amounts of data contained in HADs and can learn concise representations of diagnosis sequences over time with minimal pre-processing. This study aims to train a natural language processing (NLP) model on structured HADs to obtain embeddings to be used for multimorbidity detection.

Methods: Patients aged 65 or older from the Piedmont Longitudinal Study were included in the analyses. Patient longitudinal observations included diagnosis and medication sequences. To capture the inherent structure of the observations, the original Google BERT code was pre-trained for two self-supervised tasks [1]: (i) the Masked Language Model (MLM) task, wherein words were randomly masked, BERT was trained to predict the original words from their bidirectional context (left and right words); (ii) the Next Sentence Prediction (NSP) task, which involves predicting whether one sentence follows another, such as subsequent visits of a patient. A t-SNE analysis evaluated embedding quality [2].

Results: BERT pre-trained model achieved high accuracy (96.58%, SE = 0.02 for the MLM task and 99.75%, SE = 0.005 for the NSP task), indicating its ability to learn the structure of HADs' data. Embeddings potentially allow identifying co-occurring pathologies and polypharmacy conditions. In particular, were recognized as patterns of multimorbidity like: (i) rehabilitation, osteoarthritis, and connective tissue diseases, (ii) respiratory failure with respiratory and dysrhythmia diagnosis, (iii) anti-inflammatory/anti-rheumatic and anti-thrombotic drugs with reflux drugs, and (iv) beta-lactam/penicillin antibiotics, anti-inflammatory/anti-rheumatic, quinolone antibacterial drugs and other bacterial infections diagnoses.

Conclusion: The accuracy of the pre-trained model suggests that BERT can efficiently learn the data structure of HADs. Training on MLM and NSP reached very satisfying performances with at least 97% of accuracy. t-SNE analysis showed that co-occurring medical prescriptions and hospitalization diagnoses share a similar medical meaning (for example thyroid preparations are indeed linked to thyroid disorders).

References: *[1] Abadi M, Agarwal A, Barham P, et al. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [Internet]. http://tensorflow.org/ [2] van der Maaten LJP, Hinton GE. 2008. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research. 9(nov):2579–2605.*

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-04

A machine learning model for malaria prediction in an endemic area of Nigeria

Oluwole Nuga, Adekunle Sangotola
Bells University of Technology, Nigeria

Introduction: The challenges of accurate malaria prognosis in rural endemic areas of Nigeria where parasitological tests for malaria are not readily accessible is well known. Related febrile diseases have been misdiagnosed as malaria in suspected patients who present themselves for examination while many others commence treatment of the disease without any proper diagnosis. The incorrect prognosis can increase the use of unnecessary antimalarial medicines thereby increasing the spread of drug resistance. Therefore, it is important to identify significant malaria symptoms that can be used for the proper diagnosis of malaria in Nigeria.

Methods: Medical data of 116 patients diagnose with malaria and 221 patients diagnose with other febrile diseases were collected from a hospital in an endemic region of southwestern Nigeria. The data consist of the age of patients (15-65 years) and the absence or presence of 11 symptoms (cold, fatigue, diarrhea, black urine, hypoglycemia, rigor, headache, bitter tongue, convulsion, jaundice, prostration). Binary Logistic Regression (BLR) model was fitted using 70% of the dataset as the training data, ten-fold cross validation was performed and the remaining 30% test dataset was used to evaluate the performance of the model. To improve the overall performance of the model, the SMOTE technique was used to perform oversampling on the malaria class.

Results: The trained BLR model had three significant odds ratios for hypoglycemia, headache and prostration at 5% level of significance. The odds ratio of. 2.15, 2.18 and 0.55 for hypoglycemia, headache and prostration respectively indicates a high likelihood of a patient with hypoglycemia and headache to have malaria and a lower likelihood for a patient with prostration to have malaria. The overall Model Accuracy MA=69.3% while the sensitivity and specificity were 21.9% and 91.3% respectively using the test dataset. The low sensitivity value reveals the weak ability to detect malaria as it was only able to detect 7 out of the 32 malaria patients in the data. However, after resampling with SMOTE technique, the sensitivity improved further to 36.6% indicating that 19 out of 51 malaria patients were correctly identified.

Conclusion: The Binary Logistic regression model obtained from this work can assist in classifying malaria from other febrile diseases in endemic regions of Nigeria.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-05

Machine learning for HIV prediction in people who inject drugs: Addressing challenges of imbalanced data

Sotirios Roussos, Vana Sypsa
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Background/Introduction: Machine learning (ML) algorithms offer promising tools for predicting HIV infection among people who inject drugs, a key population for the prevention and care of this infection. However, imbalanced data, where the number of HIV-negative cases are outweighed by the number of positive ones, poses significant challenges. This imbalance can lead to biased models that miss diagnoses and may hinder public health efforts. Prior research has demonstrated the potential of ML for HIV prediction in this population, but addressing imbalanced data is essential for reliable and generalizable models. This study investigated the impact of imbalanced data and explored sampling techniques to mitigate its effects.

Methods: Data on HIV testing and questionnaires (including sociodemographic characteristics, injection and sexual practices) from 3,320 people who inject drugs enrolled in an intervention in Greece (2012-2013) were analyzed. After data cleaning and feature selection based on domain knowledge and statistical methods, five classification algorithms (Logistic Regression, Random Forests, Support Vector Machines, k-Nearest Neighbors and Decision Tree) were applied with four sampling techniques (no resampling, undersampling, random oversampling, Synthetic Minority Oversampling Technique-SMOTE). Model performance was evaluated using accuracy, precision, recall, and AUC-ROC on an 80/20 training-validation split.

Results: The prevalence of HIV infection was 15.2%. Among the algorithms tested, the Random Forest algorithm, applied with random oversampling on the entire set of available variables, demonstrated the best performance. Sensitivity, accuracy, and area under the curve were 0.9929, 0.9805, and 0.9967, respectively. Selecting 34 out of the 112 variables yielded sensitivity, accuracy, and AUC-score of 0.9929, 0.9751, and 0.9967, respectively. Variables contributing most to the predictions included the individual's age, frequency of injecting drug use, use of cocaine as the main substance of injecting drug use, frequency of injecting drug use, speedball, and the number of sterile syringes received by the individual in the last 12 months.

Conclusion: This study highlights the importance of addressing imbalanced data for accurate ML-based HIV prediction in people who inject drugs. The findings suggest that ML can be a valuable tool for early identification of HIV infection, enabling timely linkage to care, reduced transmission risks, and targeted interventions. Limitations include the use of data from a single program, and further research is needed for broader validation and real-world implementation. Future efforts should focus on integrating these models into healthcare systems and ensuring effective prevention strategies reach high-risk people who inject drugs.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-06

A novel simple method for evaluating the contribution of explanatory variables in deep learning: an application to pathogenicity prediction results for breast cancer-related variants of unknown significance

Asuka Suzuki[1], Kazuma Takeuchi[1], Wataru Urasaki[2], Shigeaki Sunada[3,4], Yoshio Miki[5], Masaaki Matsuura[1]
1 Teikyo University Graduate School of Public Health, Tokyo, Japan
2 Department of Information Sciences, Tokyo University of Science, Chiba, Japan
3 Juntendo Advanced Research Institute for Health Science, Juntendo University, Tokyo, Japan
4 Department of Oncology, School of Medicine, Juntendo University School of Medicine, Tokyo, Japan
5 Department of Molecular Genetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Introduction: Prediction models developed using deep learning (DL) are black boxes, and eXplainable Artificial Intelligence (XAI) is becoming increasingly important. We propose a novel and simple method to evaluate the contribution of input features to prediction results. We applied the proposed method to the DL model developed to predict the clinical significance of variants of unknown significance within the hereditary breast and ovarian cancer syndrome-related gene BRCA1/2, and compared it with SHAP, a representative XAI.

Methods: As an approach to the explainability of a black-box DL model, the dataset including the input to and output of the trained DL model was created and was analysed using regression models, which allows us to have a formula to obtain mean prediction values and to evaluate the contribution with the partial regression coefficient. As an application example, the predictive DL model was developed using the training data of a total of 1265 BRCA1 gene mutations with the pathological/ benign results published by NCBI in the USA, and a set of the results of pathological/ benign predictions by six different prediction tools that predict the effect of mutations on protein function using the gene sequence as input features.

Results: In the application example, the final model including the main effect and the interaction term was obtained, as a result of searching for multiple regression models for the goodness of fit. The model was also an approximation of the DL model, and the estimates of the partial regression coefficients allowed the magnitude and direction of the contribution to the prediction of the main effect and the interaction term to be determined. The ranking of the contribution of the main effects was revealed, and the interaction terms allowed interpretation of the relationship between explanatory and objective variables under the specific condition. The top 5 features of the proposed method and SHAP were the same, and their rankings were close.

Conclusion: A method was proposed to interpret the entire inference process performed by a black box, DL predictive model. In the application example, it was shown that the contribution of explanatory variables can be easily assessed by the proposed method and it contributes to the interpretability of the predictions. The method is applicable to DL models for supervised learned prediction based on numerical data and is expected to be widely applied.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-07 CASc winner

Machine learning methods in dynamic survival prediction under standard and mixed-model landmarking framework

Nabil Ahmed Uthso[1,2], Md Hasinur Rahaman Khan[2]
1 BRAC James P Grant School of Public Health (JPGSPH), BRAC University, Dhaka, Bangladesh
2 Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka, Bangladesh

Introduction: In medicine, typically predictive models for time-to-event responses are frequently used to point out people who are at high risk, guide treatment plans, and advise patients about their prognosis. It is only reasonable for persons with diseases that decrease their lifespan to wish to know their chances of longterm and short-term survival as time passes. Biomarker values, including medical lab results, are usually assessed over time, but traditional prediction models only rely on baseline information. Instead, implementing machine learning methods including random survival forest and extended approaches of gradient boosting in standard and mixed-model landmarking frameworks is expected naturally to overcome the problem.

Methods: This study proposes a novel methodology for dynamic prediction by utilizing random survival forest (RSF) and gradient boosting (XGBoost, LightGBM) within a standard and mixed-model landmarking framework. This approach was compared to landmarking (standard and mixed-model) and joint modeling through simulation tests across multiple scenarios where longitudinal markers have intricate associations with survival outcomes. Our models for the 'Aortic Valve' dataset served as the foundation for our simulation work under three scenarios. Two types of model misspecification are used- misspecification of the longitudinal model's temporal effect's functional form and misspecification of the functional structure that establishes the connection between longitudinal and survival outcomes.

Results: The results reveal that the RSF algorithm outperformed all other counterpart techniques when used with mixed-model landmarking. In some cases, LightGBM and XGBoost under mixed-model landmarking framework perform similarly to RSF under a mixed-model framework. Furthermore, the RSF using both standard landmarking and mixed-model landmarking shows strong performance in various circumstances for the implementation of method to a medical data.

Conclusion: This research holds significant promise for this type of developments in personalized risk prediction within the medical field by the integration of machine learning algorithms. Besides, extending the research to evaluate the feasibility of incorporating more advanced machine learning techniques into existing clinical software systems would facilitate their adoption in medical practice, making personalized dynamic predictions more accessible and efficient.

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-08

Comparison of machine learning models for classification of diabetes mellitus using body composition indicators in men

Mi Hong Yim
Korea Institute of Oriental Medicine, Daejeon, Republic of Korea

Introduction: Diabetes mellitus is a chronic metabolic disorder caused by defects in insulin secretion or insulin action [1]. Several studies have been carried out to predict or classify diabetes mellitus using noninvasive measurements. However, there are still few studies using body composition indicators to predict or classify diabetes mellitus. The purpose of this study was to build diabetes mellitus classification models in men using body composition indicators through several machine learning approaches and compare the predictive powers of the models.

Methods: This study included 820 men who were recruited at five hospitals in the Republic of Korea between April 2022 and December 2022. Six machine learning techniques were used to construct the diabetes mellitus classification models utilizing body composition indicators: elastic nets, extreme gradient boosting, neural networks, support vector machines, k-nearest neighbor, and random forests. To determine the contribution of the individual body composition indicators selected from the models, relative variable importance was computed. Nested cross-validation with five outer and five inner splits was used to assess each model's performance [2]. The results included accuracy, Kappa, precision, F1 score, sensitivity, specificity and areas under the receiver operating characteristic curve (AUC) value along with individual 95% confidence intervals (CIs). The Youden index was used for the optimal threshold, and 2000 bootstrap repetitions were used to calculate the confidence interval.

Results: The model with the elastic nets technique had the greatest F1 score, Kappa, accuracy, and AUC value (F1 score = 0.303 [95% CI, 0.220-0.384], Kappa = 0.243 [0.166- 0.322], accuracy = 0.833 [0.807- 0.857], and AUC = 0.839 [0.776-0.891]). The model that employed support vector machine displayed comparatively low AUC values of 0.708 [0.617-0.793]. Age, 50kHz phase angle of body trunk, percent skeletal muscle, and extracellular water ratio of body trunk were selected as important indicator for diabetes mellitus.

Conclusion: These findings demonstrated the potential of body composition indicators to classify diabetes mellitus noninvasively. Furthermore, classification via machine learning algorithms aids in the decision-making and insight-gaining of clinicians.

References: *[1] American Diabetes Association. "Diagnosis and classification of diabetes mellitus." Diabetes care 37.Supplement_1 (2014): S81-S90. [2] Bates, Stephen, Trevor Hastie, and Robert Tibshirani. "Cross-validation: what does it estimate and how well does it do it?." Journal of the American Statistical Association (2023): 1-12.*

POSTER SESSION B-01:
Artificial Intelligence and Machine Learning Methods

P-B01-09 LB

LLPowershap: Logistic loss-based automated Shapley values feature selection method

Iqbal Madakkatel[1,2], **Elina Hypponen**[1,2]
1 University of South Australia, Adelaide, Australia
2 South Australian Health and Medical Research Institute, Adelaide, Australia

Background/Introduction: Shapley values have been extensively used in machine learning to interpret black-box models, debug models, conduct sensitivity and fairness analyses, and for dimensionality reduction via feature selection, enabling further follow-up analyses using domain expertise and well-established statistical procedures. Shapley values satisfy certain axioms that ensure fair distribution of feature contributions towards model predictions or the mismatch between prediction and the truth, even in the presence of non-linearity and interactions in black-box machine learning models. Several methods that incorporate statistical methods in Shapley value-based feature selection have been recently introduced (i.e.Borutashap, Shapicant, and Powershap).

Methods: We modify the Powershap method and introduce LLpowershap, a novel wrapper feature selection method for classification tasks, leveraging XGBoost and Shapley values explaining the mismatch between prediction and the truth, as calculated by the logistic loss function on truly unseen data. Additionally, LLpowershap calculates p-values for identifying important features by augmenting training data with noise features of different characteristics. It also automates the process of feature selection using enhanced statistical power calculations, eliminating the need to specify the number of rounds of model training and calculating Shapley values. We evaluated LLpowershap along six other features selection methods on simulation datasets created using scikit-learn package in Python. We also benchmarked predictive performance of selected features on four real-world datasets using area under the receiver operating characteristics curve (AUROC) by two independent state-of-the-art gradient boosting algorithms, CatBoost and LightGBM.

Results: Simulation results on a total of two hundred datasets with the number of features ranging from 20 to 500 and the proportion of informative features ranging from 3% to 90%, with 5,000 and 10,000 samples, show LLpowershap identifies a greater number of informative features and fewer noisy features compared to other leading feature selection methods including Powershap, in most scenarios. Benchmarking on the real-world datasets, including the UK Biobank cancer prediction data, shows features selected using LLpowershap have higher or comparable predictive performance, when performance is assessed using AUROC, using 10-fold cross validation on the training sets and 1,000 bootstrap datasets on the test sets.

Conclusion: LLpowershap enhances the robustness of feature selection by integrating logistic loss-based Shapley values calculated on unseen test sets and enhanced p-value and statistical power calculations, offering a reliable method for identifying relevant features with reduced noise. The method holds promise for better feature selection in large biomedical datasets and other settings, improving model performance and interpretation.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-01 Comparison of random forests modelling strategies for EHR data in the presence of competing risks – A case study on central line-associated bloodstream infections

Elena Albu[1], Shan Gao[1], Pieter Stijnen[2], Frank Rademakers[3], Veerle Cossey[1,4], Christel Janssens[5], Yves Debaveye[6], Laure Wynants[1,7,8], Ben Van Calster[1,7,9]
1 Department of Development and Regeneration, KU Leuven, Belgium
2 Management Information Reporting Department, University Hospitals Leuven, Belgium
3 Faculty of Medicine, KU Leuven, Belgium
4 Department of Infection Control and Prevention, University Hospitals Leuven, Belgium
5 Nursing PICC team, University Hospitals Leuven, Belgium
6 Department of Cellular and Molecular Medicine, University Hospitals Leuven, Belgium
7 Leuven Unit for Health Technology Assessment Research (LUHTAR), KU Leuven, Belgium
8 School for Public Health and Primary Care, Maastricht University, Netherlands
9 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands

Objective: Prognostic outcomes related to hospital admissions often do not suffer from censoring, and can therefore be modelled categorically. On the other hand, competing events are common but often ignored. We compared the performance of random forest (RF) models to estimate the risk of central line-associated bloodstream infections (CLABSI) using different outcome operationalizations.

Materials and Methods: We included data from 27478 admissions to the University Hospitals Leuven, covering 30862 catheter episodes (970 CLABSI, 1466 deaths and 28426 discharges) to build static and dynamic RF models for binary (CLABSI vs no CLABSI), multinomial (CLABSI, discharge, or death), survival (time to CLABSI) and competing risks (time to CLABSI, discharge or death) outcomes to estimate the 7-day CLABSI risk. We used 21 features, comprising catheter type and location, comorbidities, laboratory results and vital signs. We trained baseline and dynamic RF models using randomforestSRC [1] and tuned the number of variables selected at each split, node size and subsample size using mlrMBO [2]. We evaluated model performance across 100 train/test splits.

Results: Model performance was similar for all models except survival models. AUROC reached 0.74 for baseline models, 0.78 for predictions made 5 days after catheter placement, but decreased thereafter. Models with survival outcome strongly overestimated the risk of CLABSI (O:E ratios between 1.2 and 1.6), and had AUROCs about 0.01 lower than other models; however, if censoring for the competing events is applied at the time horizon of interest (7 days) instead of at their event time, similar to Fine-Gray models [3], survival models performance was comparable to the other models. Models for categorical outcomes (binary and multinomial) had lowest computation times for model tuning. Despite small differences in overall performance, models including multiple levels of the outcome (multinomial and competing risks) display a different internal structure compared to binary and survival models, prioritizing different variables for early splits in the trees of the RFs.

Conclusion: In the absence of censoring, complex modelling choices do not considerably improve the predictive performance compared to a binary model for CLABSI prediction in our studied settings.

References: *[1] Ishwaran, H., Kogalur, U. B., & Kogalur, M. U. B. (2022). Package 'randomForestSRC'. breast, 6(1). [2] Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., & Lang, M. (2017). mlrMBO: A modular framework for model-based optimization of expensive black-box functions. arXiv preprint arXiv:1703.03373. [3] Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association, 94(446), 496-509.*

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-02 Methodological considerations for target trial emulation: A systematic review

Faye Baldwin[1], Rukun Khalaf[2], Ruwanthi Kolamunnage-Dona[1], Andrea Jorgensen[1]
1 Department of Health Data Science, University of Liverpool, Liverpool, United Kingdom
2 Department of Public Health, Policy and Systems, University of Liverpool, Liverpool, United Kingdom

Background: Target trial emulation is an emerging methodological framework used within health research, and involves the application of design principles from randomized controlled trials (RCT) to observational data. Target trial emulation is particularly applicable in situations where an RCT would be unethical or unfeasible. However, the methodology used within target trials can vary based upon the research question proposed and the outcomes of interest. This systematic review is the first of its kind to critically appraise the methodologies proposed for target trial emulation when comparing treatment effectiveness using observational data.

Methods: A comprehensive search strategy was used to identify studies detailing methods for emulation of a target trial comparing the effectiveness of treatments using observational data in the following bibliographic databases: PubMed, Scopus, Web of Science and Ovid MEDLINE. Data was extracted from eligible studies pertaining to study design, outcomes measured, and statistical methods used to control for confounding. A summary and critical appraisal of the various identified statistical methods was developed.

Results: 59 papers were included in the systematic review, with 47 papers emulating a target trial comparing treatment effectiveness (application papers) and 12 papers proposing and/or utilizing methods to emulate a target trial (methods papers). All application and methods papers reported use of one or more methods to control for baseline confounding, this was most commonly achieved by adjustment for covariates within a regression model (N=25, 53% application papers; N=5, 42% methods papers). However, only a subset of papers reported use of one or more methods to control for residual or unmeasured confounding. Further to this, we also identified that some of the application papers did not fully report on the methods utilized.

Conclusion: In this review, we identified a number of different methods proposed for adjusting for confounding and other biases within target trials, and our summary and critical appraisal of those methods will guide those embarking on target trial emulation in the future. Given the incomplete reporting of methods in some papers, we recommend that researchers follow a target trial protocol, and that future studies emulating a target trial adhere to the TARGET reporting guideline currently in development [1].

Reference: *[1]. Hansford HJ, Cashin AG, Jones MD, Swanson SA, Islam N, Dahabreh IJ, et al. Development of the TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline. BMJ Open. 2023;13(9):e074626.*

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-03

Assessment of frailty in the general population based on electronic health data

Giuseppe Occhino[1,2], Carlo Alberto Scirè[2], Pasquale Dolce[3], Maria Grazia Valsecchi[2], Paola Rebora[2]
1 Strategic Regional Agency for Health and Social Care of Apulia (AReSS Puglia), Bari, Italy
2 Bicocca Center of Bioinformatics, Biostatistics and Bioimaging (B4 Center), School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy
3 Department of Translational Medical Science, University of Naples Federico II, Naples, Italy

Introduction: Frailty is a clinical syndrome resulting from interaction of the age-related decline in physiologic systems with chronic diseases and can be useful from a public health point of view to timely identify which people require major health care. Some tools to measure individual's frailty from electronic health data have been proposed and one of most widespread is the cumulative frailty index (FI), defined as proportion of deficits in a large number of domains. This is easy to interpret, but it does not consider the consistency and redundancy of the deficits and their predictive value. We aimed to develop and validate a FI using electronic regional health databases (e-RHD) by statistical regression approaches, in which deficits with higher predictive values get a higher weight, and Artificial Neural Networks (ANN), which account for the interrelationships among predictors and are especially beneficial for prediction purposes when large samples are available.

Methods: By the end of 2020, a 40-item FI (e-RHD-FI) was created using data from the Lombardy e-RHD for adult beneficiaries of the Regional Health System. Subjects were followed up to death, emigration or last available follow-up time. We randomly split the cohort into a training (70%) and a validation set (20%), leaving 10% of subjects as an internal testing cohort. The association between the e-RHD-FI and 1-year mortality was evaluated by Cox proportional hazards regression model. Then we will use a model-based approach and ANNs for time-to-event prediction. The performance of the frailty indices in predicting mortality will be assessed on the internal test set by the area under the Receiver-Operating-Characteristic (ROC) curve (AUC), adjusted for censoring.

Results: We calculated the e-RHD-FI (median 0.03 [first-third quartiles, 0-0.05]) in 8,347,517 adults (51.7% females, median age 52 years). 5,845,669 adults, of whom 74,934 (1.3%) died within 1 year, were included in the training set. 1,667,158 adults, with 21,583 (1.3%) deaths within 1 year, were included in the validation set. The AUC for 1-year mortality was 0.866 (99% CI: 0.863-0.869) for e-RHD-FI. In the multivariable Cox model, adjusted for sex and age, each 0.1-point increment of e-RHD-FI was associated with increased 1-year mortality (HR: 2.04, 99% CI: 2.01-2.08).

Conclusion: In the general population, the e-RHD-FI was associated with 1-year mortality after accounting for age and sex and showed a good discrimination ability. ANNs for time-to-event end-points will provide high prediction accuracy and enable the modeling of complex relationships between the frailty items and overall mortality.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-04

A new strategy to automate the high-volume integration of near real-time data streams in the international multicentre CRICKET (Critical Events in Anaesthetised Kids Undergoing Tracheal Intubation) study

Cinzia Anna Maria Papappicco[1], Giulia Lorenzoni[1], Thomas Riva[2], Nicola Disma[3], Dario Gregori[1]

1 Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy, Padua, Italy
2 Department of Anaesthesiology and Pain Medicine, Bern University Hospital, Inselspital, University of Bern, Bern, Switzerland, Bern, Switzerland
3 Unit for Research in Anaesthesia, IRCCS Istituto Giannina Gaslini, Genoa, Italy, Genoa, Italy

Introduction: International multicenter prospective study has several advantages over single-centre or limited to a single geographical area study. By involving worldwide participants, it allows issues of global relevance to be addressed and provides a more representative overview of genetic, environmental, and cultural diversity, thus improving generalizability and robustness of results. Despite these advantages, multicenter study presents the dichotomous challenge of management complexity, coordination between different institutions and managing sensitive data stream integration, including privacy concerns arising from varied legislative frameworks in different countries. In real-time data management, latency in data transmission and interoperability must be minimized to ensure data integrity. It is crucial to identify a strategy to automate the process of extracting high volume of real-time data from different sources and transforming it into a single standard format that ensures data integrity, and accurate results.

Methods: CRICKET is a prospective, international multicenter study with the aim of analyzing critical events associated with airway management during tracheal intubation in children. The study includes a minimum of 105.000 patients recruited from more than four hundred anaesthesia departments worldwide, involving North and South-Central America, Australia, Asia, Europe, and Africa. Study data are collected and managed using REDCap (Research Electronic Data Capture). Given the worldwide volume of study centers and to cope with different regulatory frameworks that hinder direct centralized data collection, procedures for data integration and interoperability between multiple sources are performed using REDCap-ETL (Extract, Transform, Load) application. In ETL process, we specify multiple ETL tasks, so a given workflow may extract data from multiple data sources using REDCap's API (Application Programming Interface), transform extracted data based on transformation rules, and load transformed data into a centralized database. To enable this data pipeline, which goes from capturing from different sources to rapid transformation into a database ready for analysis, we built a dedicated cloud computing infrastructure.

Results: Since validation, the preliminary implementation showed a promising ability to avoid security breaches for a large volume of patient records, data completeness, inconsistent data handling, incongruent data correction, and increase in data integrity with a centralized database available for advanced statistical analysis.

Conclusion: This articulated strategy poses a stimulating challenge of finding a balance between multidimensional computing approach, to synchronize, store and process large amounts of data from heterogenous sources, ethical-legal solutions to enable privacy-protected data flows, and machine learning techniques to transform near real-time data into real high-quality evidence that supports personalized patient care.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-05

Comparative analysis of methods for identifying multimorbidity patterns among people with opioid use disorder in Ontario, Canada

Myanca Rodrigues[1], Glenda Babe[2], Tea Rosic[1], Brittany B. Dennis[1], Richard Perez[2], Claire de Oliveira[2], Sameer Parpia[1], Lehana Thabane[1], Zainab Samaan[1]
1 McMaster University, Hamilton, Canada
2 ICES McMaster, Hamilton, Canada

Background: Multimorbidity is defined as the co-occurrence of two or more (2+) chronic conditions. As the prevalence of chronic conditions increases, identification of patterns of co-occurring illnesses is crucial to health system planning. Cluster analysis is a commonly used method to identify multimorbidity patterns. However, to-date, studies have focussed on hierarchical cluster analysis (HCA), with a paucity of research examining non-hierarchical cluster analytic methods, such as K-means analysis. Our primary aim was to compare multimorbidity patterns using two methods - HCA and K-means clustering - in our cohort of people with OUD in Ontario, Canada.

Methods: We linked observational cohort data collected from 3,430 people receiving treatment for OUD between 2011 and 2021 in Ontario, Canada to provincial health administrative databases. We identified 18 chronic conditions, commonly used in multimorbidity studies in Ontario, using ICD-10-CA diagnostic codes and the diagnostic codes of physician billing claims, and followed the cohort over an eight-year period in the data holdings. We used HCA and K-means clustering to identify multimorbidity patterns. Analyses were stratified by sex, with results compared for each method.

Results: HCA identified 4 clusters for males, and 2 for females. K-means identified 3 multimorbidity patterns for each sex. Although there were some differences by sex and method of analysis, two combinations of disease were observed consistently across both methods: (i) diabetes, hypertension and stroke for males, and (ii) asthma and chronic obstructive pulmonary disease for females.

Conclusion: Our study findings illustrate that males and females experience different clusters of chronic conditions, and that ascertainment of multimorbidity patterns varies depending on the method of analysis used (HCA vs. K-means). HCA may be better used to determine multimorbidity patterns in large, health care administrative datasets to provide an in-depth examination of multimorbidity, whereas K-means may be better suited for identification of disease clusters found in clinical practice.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-06

Jointly estimating treatment dosage and number of mature oocytes for oocyte cryopreservation using a national clinical outcome database

Lutfiyya Muhammad[1], Bahar Yilmaz[1], Mrs. Chen Yeh[2], Jennifer Bakkensen[1], Eve Feinberg[1]
1 Northwestern University, Chicago, United States
2 Rush University Medical Center, Chicago, United States

Planned oocyte cryopreservation (OC) is common in the United States for deferred reproduction. Oocyte yield is impacted by multiple clinical factors including female age, total dosage of gonadotropins and hormone levels. Currently, clinicians determine medication dosages for ovarian stimulation based on their personal experience. Administering dosages that are too low may result in poor ovarian response or decrease the number of oocytes. The objective of this study is to construct a prediction model that jointly estimates the total dose of gonadotropin and the amount of retrieved mature oocytes in planned OC cycles.

For this study, 15,806 individual OC cycles were obtained from the 2013-2018 Society for Assisted Reproductive Technology Clinic Outcome Reporting System Database. To jointly predict the total dose of gonadotropin and the number of retrieved mature oocytes, a bivariate copula additive model was constructed. 70% of OC cycles were utilized as model training to construct the prediction model, and the remaining 30% were utilized to test the performance of the model. A Cullen and Frey graph was first used to determine the distributions of the two outcomes. Candidate covariates in the model included demographic and clinical factors such as age, race, diminished ovarian reserve, and hormone levels. Graphical displays and Akaike Information Criterion (AIC) values were used to decide the appropriate copula function. The final prediction model had the smallest AIC value and a Clayton copula rotation of 270 degrees. Model performance measures such as model accuracy and mean absolute error were computed based on the tolerance levels for each outcome (number of days on gonadotropin and the number of mature oocytes).

In our final model, covariates associated with gonadotropin total dose were age, race, smoking history, follicle-stimulating hormone (FSH), Anti-Müllerian hormone (AMH), diminished ovarian reserve, and anovulatory disorder. Similar covariates were associated with the number of retrieved mature oocytes, except for smoking history and anovulatory disorder. Model performance statistics were comparable for the training and test datasets. Model accuracy ranged between 70% - 90% for the training and test datasets at various tolerance levels. For instance, the predictive model for both datasets achieved 70% accuracy at a gonadotropin dosage of 450 IU per day for 12 ± 3 days and a tolerance level of 6 mature oocytes.

The findings from this prediction model would be helpful in counseling patients during the OC planning process to frame expectations on both anticipated cost and realistic expectations.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-07

Propensity score matching to compare the healthcare resource utilisation of patients with multiple sclerosis using the French national health data system

Xavier Moisset[4], Sophie Larrieu[3], Alexandre Civet[1], Clarisse Marchal[2], Manon Belhassen[2], Floriane Deygas[2], Guillaume Bourel[1], Laure Rolland[1], Alexandre Civet[1], Gregoire Mercier[5]
1 Roche, Boulogne Billancourt, France
2 PELyon, Lyon, France
3 Horiana, Bordeaux, France
4 CHU Montpellier, Montpellier, France
5 CHU Clermont-Ferrand, Clermont-Ferrand, France

Background:/Introduction: In order to compare the healthcare resource utilisation of patients with multiple sclerosis (PwMS) regarding their treatments in a real world setting, a burden of illness study was set-up using the French claims database (Système National des Données de Santé: SNDS). This abstract describes how matching between different subgroups of PwMS was performed.

Methods: A retrospective cohort study was conducted through the SNDS, identifying all PwMS who initiated MS treatment of interest (ocrelizumab; first-line disease modifying treatment (DMT); second-line DMT) between 2019 and 2020, with a history period starting from 2014. Index date was defined as the date of initiation of ocrelizumab (or treatment of interest). PwMS were classified as naïve (no MS-treatment reimbursement before index date) or non-naïve in each subgroup. Four 1:1 propensity score (PS) matching were performed: Naïve and non-naïve patients using ocrelizumab with respectively naïve and non-naïve patients using first-line DMT / same groups for second-line DMT. To control for potential channelling bias in subgroups comparisons, fourteen covariates were considered. PS conventional method was used: probability to belong to a given subgroup was estimated using a logistic regression model. Matching was performed using a sequential pairwise nearest neighbour approach without replacement.

Results: From 142,210 PwMS identified in the SNDS database, 19,853 received at least one dispensation of MS treatment of interest in the inclusion period. 18,912 PwMS had sufficient history data: 5,571, 10,322 and 3,019 who respectively initiated ocrelizumab (1,046 naïve, 4,525 non-naïve), a first-line DMT (5,943 naïve and 4,379 non-naïve), a second-line DMT (748 naïve and 2,271 non-naïve).

Non-naïve subgroups: 70.8% and 48.3% of the PwMS treated with ocrelizumab were respectively matched with 73.3% of the PwMS treated with a first-line DMT and 96.3% with a second-line DMT, representing respectively 6,350 (3,175 pairs) and 4,330 (2,165 pairs) matched PwMS.

Naïve subgroups: 94.5% and 59.3% of the PwMS treated with ocrelizumab were respectively matched with 16.7% of the PwMS treated with a first-line DMT and 84.1% with a second-line DMT, representing respectively 1,952 (976 pairs) and 1,226 (613 pairs) matched PwMS.

Conclusion: In all 4 matchings performed, PS were well balanced between subgroups regarding all covariates. Some important clinical covariates were not available in the SNDS leading to potential biases in the estimation of the treatment effect.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-08 Integration of historical data into the design and analysis of clinical trials for rare diseases

Emilie Jounot, Lucie Truffaut-Chalet, Xiangmin Zhang, Audrey D'Abrigeon
Servier, Paris, France

Background: In the rare neurological disease field, demonstrating treatment efficacy in clinical trials (CT) for new therapies is a daunting task due to limited patient availability. Recruiting an adequate number of patients for a randomized clinical trial (RCT) becomes unfeasible in some cases. Recognizing the growing importance of incorporating historical data in CT planning, especially in expanding the control group of a new trial, is crucial. However, this introduces the risk of potential bias, given the disparities between the historical control population and the RCT.

Methods: To address population differences and mitigate bias, various methods can be employed. These include frequentist approaches such as propensity score matching (PSM) and inverse probability of treatment weighting (IPTW), or a combination of frequentist and Bayesian (PSM or IPTW combined with power prior or commensurate prior).The primary objective is to compare the operational characteristics of these methods that enable the borrowing of historical data from an external real-word cohort in the context of a new trial in rare neurodegenerative disease using a continuous endpoint. Simulated scenarios, accounting for imbalances in patient group characteristics, are implemented.

Results: The results indicate that borrowing historical data methods can enhance power without significantly introducing bias or inflating type I error. The actual benefit depends on the hypotheses and specific settings. A trade-off between power increase and mitigation bias or inflating type I error, and discussion with authorities will finalize the choice of method for a new trial.

Conclusion: Augmented concurrent controls with a combination of propensity score-matched methods and dynamic Bayesian borrowing can improve the efficiency of clinical trials in rare disease and enable the trial to be conducted with smaller overall sample size, while maintaining covariate balance and study power and minimizing bias in response estimation. This approach does not fully eliminate the concern that introducing non-randomized historical controls in a trial may lead to bias in estimating treatment effects, and should be carefully considered on a case-by-case basis.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-09

The use of healthcare systems data for RCTs

Alice-Maria Toader, Carrol Gamble, Susanna Dodd, Paula Williamson
University of Liverpool, Liverpool, United Kingdom

Background: Healthcare systems data (HSD) has the potential to optimise the efficiency of randomised controlled trials (RCTs), by decreasing trial-specific data demands. Therefore, the use of HSD in trials is expected to increase. In 2019, it was estimated that 47% of NIHR-funded trials were planning to use HSD. We aim to understand the extent and nature of its current use and its evolution over time.

Methods: We identified a cohort of RCTs within the NIHR Journals Library that commenced after 2019 and were described as being in progress on 6 June 2022. Details on the source and use of HSD were extracted from eligible RCTs. The use of HSD was categorised according to whether it was used as the sole data source for outcomes and whether the outcomes were primary or secondary. HSD is often insufficient for patient-reported outcomes (PROs). We aimed to determine methods used by trialists for collecting PRO data alongside HSD.

Results: Of the 84 eligible studies, 52 (62%) planned to use HSD and 79 (94%) planned to collect PROs. The number of RCTs planning to use HSD for at least one outcome was 28 (54%) with 24 of these planning to use HSD as the sole data source for at least one outcome. The number of studies planning to use HSD for primary and secondary outcomes was 10 (20%) and 21 (40%) respectively. The sources of HSD were National Health Service (NHS) Digital (n = 37, 79%), patient registries (n = 7, 29%), primary care (n = 5, 21%), The Office for National Statistics (ONS) (n = 3, 13%) and other (n = 2, 8%). PROs were collected for 92% of the trials planning to use HSD. Methods for collection of PROs included in-person (n = 26, 54%), online (n = 22, 46%), postal (n = 18, 38%), phone (n = 14, 29%) and app (n = 2, 4%).

Conclusions: HSD is being used in around two thirds of the studies but cannot yet be used to support PRO data collection within the cohort we examined. Comparison with an earlier cohort demonstrates an increase in the number of RCTs planning to use HSD.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-10

Social support as a key determinant of resilience and well-being of mothers: A network analysis of data from a general population survey

Rossana Di Staso[1], Paola Rucci[2], Valeria Calcaterra[3,4], Gian Vincenzo Zuccotti[3,5], Bernardo Dell'Osso[5,6], Gilla Cauli[6]

1 Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
2 Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
3 Pediatric Department, "Vittore Buzzi" Children's Hospital, Milan, Italy
4 Department of Internal Medicine, University of Pavia, Pavia, Italy
5 Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
6 Department of Mental Health and Addictions, ASST Fatebenefratelli-Sacco, Bologna, Italy

Background: Family environment has been reported to influence neurodevelopmental and psychiatric outcomes in children, affecting motor and sensory development, temperament, cognitive abilities, behavioural and emotional responses. Individual resilience of mothers depends on the attachment and parental bonding experience, on the social support from partner and relatives and on the capacity to develop a positive family climate for children growth. Early identification of family frailty may inform decisions about appropriate targeted interventions.

Methods: Data are based on an online survey for women with child ren aged ≤12 years, referred to Vittore Buzzi" Children's Hospital, Milan (for an auxological evaluation. The survey consisted of 8 sections: 1) socio demographic characteristics, 2) pregnancy and postpartum variables, 3) child health (viral infections in the last 6 months, chronic diseases, 4) perceived social support from partner and family, 5) stressful experiences in the last months, 6) attachment style, 7) resilience and well-being, 8) perceived parental care. The relationship among conceptual domains was investigated using network analysis, that represents graphically variables as nodes and correlations as edges. Correlations between variables were estimated using a LASSO procedure that assigns penalties to partial correlations between variables to make small correlations automatically shrink to 0. Network display was based on Fruchterman and Reingold's algorithm, that places strongly associated node s at the center of the graph. To quantify the importance of each node in the adaptive LASSO network, we computed four centrality indices: betweenness, closeness, and strength, and expected influence. For each index, high values reflect great centrality i n the network, but high strength may also derive from very strong correlations between peripheral nodes belonging to the same domain. Centrality plots were created to represent these indices.

Results: Survey participants were 205 women with a mean age of 39 years (SD=5.6). The four parental bonding variables were strongly associated with one another and with social support, that was in turn directly related with well-being. High parental levels of care were correlated with low overprotection levels. Anxious attachment was associated with low social support and low well-being. Notably, well-being and social support had the highest expected influence in the network and were the most central in terms of strength of relationships with the other variables and number of direct and indirect connections.

Conclusions: Results of network analysis indicate that interventions targeted to improving social support may have the potential to increase the well-being and resilience of mothers.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-11 Using simulations to compare cohort and case-control study designs for evaluating education outcomes of children with congenital or rare diseases

Joachim Tan[1,2], Milagros Ruiz Nishiki[1], Ania Zylbersztejn[1,2], Pia Hardelid[1,2]
1 UCL Great Ormond Street Institute of Child Health, London, United Kingdom
2 NIHR GOSH Biomedical Research Centre, London, United Kingdom

Background: Linked administrative data on whole populations are key for comparing the educational outcomes of children with rare conditions with those of their peers. Due to patient confidentiality and disclosure risks, however, some ethics and data sharing committees only allow case-control study designs with smaller sample sizes.

Objectives: To assess bias, precision and coverage of estimates of educational attainment for children with an exemplar rare disease (congenital hypothyroidism; CHT), comparing full cohort (FC) and case-control study designs.

Methods: We followed the ADEMP (Aims, Data-generating mechanisms, Methods, Estimands, Performance measures) structure described by Morris et al. (Statistics in Medicine. 2019; 38: 2074–2102. https://doi.org/10.1002/sim.8086). We simulated 1000 Monte Carlo datasets, each comprising 100,000 children with and without CHT representative of 27 North London and surrounding local authorities (LA) in England, using information from: the North Thames newborn screening database; the Great Ormond Street Hospital database of CHT cases; official population estimates and deprivation indices. The outcome, a standardised z-score for reading attainment at age 11 in national tests, was modelled with linear predictors CHT (binary variable), sex and deprivation quintiles with parameters -0.3, -0.2 (male) and +0.1 (per quintile above lowest) respectively. Parameters were extrapolated from published statistics and literature on children with other conditions assumed to have similar achievement. Regression coefficients and standard errors (SEs) for the effect of CHT on reading attainment adjusted for sex and deprivation using mixed models with random intercepts were obtained, varying factorially: number of controls (1;2;5;10;25;100 per case); disease prevalence (1 in 5,000=low; 1 in 1,000=reference; 1 in 333=high); adjustment for deprivation.

Results: Relative bias with FC was always lowest (range: 0.3%-1.6%) compared to 1/2/5/10/25 controls per case (median: 6.7%/5.3%/4.5%/4.1%/3.0% respectively). For reference and high prevalence, increasing controls from 1-10 per case did not consistently decrease bias. Relative precision (measured by empirical SE) compared to FC ranged from 60% poorer (one control) to similar (25 controls). Coverage remained within ±1% of nominal for most scenarios but was least stable for low prevalence and few controls (range: 90.2%-96.4%). Omitting deprivation increased relative bias by up to 4-fold compared to FC but little affected precision or coverage (range: 93.7%-94.8%).

Conclusion: We suggest 25+ (0.025% of cohort) controls per case so as to minimise bias and achieve similar precision as FC. Varying the effect of CHT and developing more complex yet plausible models would help inform data negotiations for studying other diseases and outcomes.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-12 Missing data in routinely collected electronic health records: An approach to characterize different levels of missing data

Georgy Gomon[1], Janet Kist[1], Rachel Knevel[1,2], Saskia Le Cessie[1]
1 Leiden Univerisity Medical Center, Leiden, Netherlands
2 Newcastle University Translational and Clinical Research Institute, Newcastle Upon Tyne, United Kingdom

Background/Introduction: Electronic health records (EHRs) provide valuable information, yet the considerable amount of missing data poses a significant challenge. Conventional methods to determine the source of missing data, such as visual inspection, often fall short for routinely collected EHRs because of the large sample size and different levels of missing data. These include missing measurements within visits, missing visits (visits are not recorded) and missing patients. The last two types are especially challenging, as it makes it impossible for researchers to distinguish between visits not taking place and visits not being recorded. Important sources of missing data within routinely collected EHRs include variations in recording patterns among medical professionals, the EHR information system used, and evolving habits of medical professionals over time. We propose a method to characterize the different levels of missing data.

Methods: We analyzed a subset of the ELAN (Extramuraal LUMC Academisch Netwerk) primary care dataset, originating from the The Hague/Leiden region in the Netherlands, containing 430,000 patients with musculoskeletal complaints with over 7,000,000 patient years. To identify the source of missingness at different levels, GLMM's (Generalized Linear Mixed Model) were used. Missing measurements were modelled using a binomial GLMM with the EHR information system, the decade of record, and the primary care provider as random effects. Missing visits and missing patients were modelled using a count process and fitted as a Poisson GLMM, incorporating patient age as fixed effect and the EHR information system, the decade, and the primary care provider as random effects.

Results: We found that the main source for missing measurements within visits is the EHR information system. In total 90-95% of all missingness is attributable to the information system, the primary care provider and the decade in which the visit was recorded. For missing visits we found that the information system and the decade contribute equally to the missingness, together explaining 80-90% of all missingness.

Conclusions: We show a simple yet elegant method to quantify the percentage of missingness attributed to different factors in large EHRs with multiple levels of missingness. Applying this method to the ELAN dataset revealed the dominant influence of the information system at the level of missing measurements within visits, while the information system and the decade are the main factors on the level of missing visits and missing patients. With this we can assume that the missing data mechanism is likely missing at random.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-13 Care pathway heterogeneity in Amyotrophic Lateral Sclerosis: Effects of gender, age and onset

Marc Dibling, Juliette Ortholand, Sophie Tezenas du Montcel
Aramis, Sorbonne Universite, Institut Du Cerveau - Paris Brain Institute - ICM, CNRS, INRIA, INSERM, AP-HP, Hopital De La Pitie Salpetriere, Paris, France

Background: Amyotrophic lateral sclerosis (ALS) is a rare neurodegenerative disease characterized by progressive motor neuron death resulting in loss of muscle function and accompanied with heterogeneous clinical manifestations. Care management is restricted to palliative strategies. However, benefit and timing assessment of ALS major clinical interventions remains challenging with different and unspecific time-to-events estimates reported in literature. Consequently, we proposed a care pathway investigation of strategies diversity and temporality, considering 3 patient characteristics reported to play a role in disease progression: age, gender, and onset site.

Methods: We developed an identification algorithm designed to specifically identify incident ALS patients in the French medico-administrative hospital database, based on age, hospital stays characteristics and diagnoses, and assessed its quality through comparison with literature. We compared patient profile-specific median time-to-clinical event and clinical interventions occurrence dynamics, through an Accelerated Failure Time model and a time-inhomogeneous multi-state model, defined by 15 transitions and 7 states, encompassing intervention history status. Twelve different patient profiles were defined according to sex, age group and the presence of symptoms indicative of the site of onset site, to model care pathway trajectories and estimate acceleration factors considering interactions.

Results: We identified 21,153 incident patients with ALS between 2013 and 2022 with a mean age of 67.7±13.1 years, male/female and spinal/bulbar ratios of 1.2 and 1.9, respectively. Non-invasive ventilation (NIV), gastrostomy, tracheostomy or death at hospital was recorded for 55.24% of the study population. We demonstrated that gender, age class and onset site influence care pathway strategies. Notably, older age and bulbar onset site and being a women prompt gastrostomy use and spinal onset site was associated with delayed NIV initiation while tracheostomy, mainly considered for younger patients (<64 years), is rarely indicated in ALS care management. Alongside investigation of time-to-event speed, we report extensively the patient profile-specific estimated median delay before clinical event start.

Conclusion: Care pathway strategy investigation is crucial in the context of rare diseases restricted to life-support care management to anticipate patient needs and foresee healthcare strategies. Health claims databases are powerful resources which, at the expense of rigorous data extraction methodology, offer a large sample size, growing temporal coverage and can be applied to such analyses. In conjunction with multi-state models, natural history data can provide a detailed analysis of care pathways which reveals disparities in ALS care management strategies depending on patient profiles and contributes to improved support strategy anticipation.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-14

Investigating the potential anti-viral effects of Proton Pump Inhibitors on influenza and COVID-19: Intention-to-treat trial emulation using electronic health records

Yat Yi Fan[1], Caroline Dale[1], Rohan Takhar[1], Michail Katsoulis[2], Rui Providencia[3], John Overington[4,5], Reecha Sofat[1,6]

1 University of Liverpool, United Kingdom
2 MRC unit for Lifelong Health and Ageing, Institute of Cardiovascular Sciences, University College London, United Kingdom
3 Institute of Health Informatics Research, University College London, United Kingdom
4 Exscientia plc, United Kingdom
5 Institute of Cardiovascular Science, University College London, United Kingdom
6 Health Data Research UK, United Kingdom

Background: Respiratory viruses pose a global public health threat, prompting the exploration of novel treatment and repurposing existing medication. Proton pump inhibitors (PPIs), widely used for gastric acid-related conditions, have been suggested to exhibit anti-viral effects against respiratory viruses by inhibiting the egress of enveloped viruses from the infected host cell. We hypothesised their potential prophylactic role against COVID-19 and influenza.

Methods: Utilising a comprehensive electronic health record of approximately 57 million individuals in England, we established nested sequential non-randomised trials. Two analyses were conducted to investigate the outcomes of influenza events (either a diagnosis by general practitioner or hospital admission due to influenza) and COVID-19 events (either a positive PCR test, general practitioner diagnosis, hospital admission or death caused by COVID-19) in adults over 18 years old and were newly diagnosed with Gastroesophageal Reflux Disease (GERD). Individuals who had prior use of PPI in the past 6 months, had the outcome in the past 3 months, received more than one PPIs were excluded. COVID-19 analyses spanned from February 2020 to March 2022, while influenza analyses covered September 2022 to January 2023. We emulated a series of sequential trials starting at each month of the study period comparing initiators to non-initiators of PPIs (omeprazole, esomeprazole, rabeprazole, pantoprazole or lansoprazole). Pooled logistic regression was performed to pool data for all emulated trials and estimate the intention-to-treat effect.

Results: Between February 2020 and March 2022, a total of 35,524 eligible individuals were included in the analysis, comprising 50% males, with a median age of 55 (IQR, 40-68). During this period, 151,438 trial cases were generated, with 13% of them initiating PPIs. 91 initiators and 634 non-initiators experienced COVID-19 events, resulting in an intention-to-treat hazard ratio (HR) of 0.90 (95% CI 0.72-1.12). During the 2022 flu season, 153,104 trial cases were evaluated consisting of 35,028 individuals (50.2% male) with a median age of 55 (IQR, 40-69). Of these cases, 12.8% involved the initiation of PPIs. 28 initiators and 196 non-initiators experienced influenza events, yielding an intention-to-treat HR of 0.89 (95% CI 0.60-1.34).

Conclusion: Our study did not find evidence of PPI's conferring protection against flu or COVID-19, however, it demonstrates an approach to begin to get non-biased estimates from observational data, on which randomised controlled trials can be prioritised.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-15 Congenital heart defects and educational outcomes: Findings from a South Australian data linkage study

Lynne Giles[1,3], Mumtaz Begum[2,3], Vivienne Moore[1,3], Alice Rumbold[3,4], Renae Fernandez[1,2,3], Michael Davies[2,3]
1 School of Public Health, The University of Adelaide, Adelaide, Australia
2 Adelaide Medical School, The University of Adelaide, Adelaide, Australia
3 Robinson Research Institute, The University of Adelaide, Adelaide, Australia
4 South Australian Health and Medical Research Institute Women and Kids Theme, Adelaide, Australia

Background: Congenital heart disease (CHD) occurs in around 1% of live births. Children with CHD may have multiple developmental deficits. Few studies have followed children beyond pre-school years and it is possible that the prevalence of neurodevelopmental problems in children with CHD increases over time. We investigated educational outcomes in children with CHD from ages 8 to 14 and compared these to outcomes for children without CHD.

Methods: The study cohort comprised linked perinatal and childhood administrative data from South Australia. Data linkage was performed by SANT DataLink, a third-party data linkage unit that is a consortium comprising state government agencies and South Australian universities. Participants included infants born in South Australia between 1/1/95 and 31/12/15 who had educational outcomes as measured by the National Assessment Program Literacy and Numeracy from 2008 onwards (NAPLAN) when children were in school grades 3, 5, 7 or 9 (corresponding to approximate ages of 8, 10, 12, or 14 years, respectively).NAPLAN assesses five domains of reading, spelling, writing, grammar, and numeracy. For each domain, a child's score is classified as absent, exempt due to special needs, not meeting minimum standard, meeting minimum standard, or high achievement. Children could contribute up to four serial scores for each domain. Children with CHDs were identified from British Paediatric Association codes using data from the South Australian Birth Defects Register spanning the study period. The exposure was CHD compared with no CHD. Mixed effects GLMs were fit for each NAPLAN domain, with random effects for child and grade. Sociodemographic variables were considered in adjusted models.

Results: In the study period, a total of 203,717 of the children had at least one NAPLAN domain subsequently recorded. Children with CHDs were at increased risk of not meeting minimum standard/exempt in each of the domains compared to children without CHDs, with that risk varying between 18% (95%CI 14-21%) for writing and 24% (95%CI 19-28%) for numeracy domains. Adjustment for a range of sociodemographic variables made negligible change to these results.

Conclusion: Understanding educational outcomes for children with CHD is important to provide guidance for parents, health care providers, and educators. Our findings may help to ensure that children with CHD are better supported to achieve their educational potential.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-16 The relationship between acute kidney injury and chronic kidney disease in patients with Type 2 Diabetes: An observational cohort study

Simona Hapca[1], Samira Bell[2], Colin Palmer[2]
1 Computing Science and Mathematics, University of Stirling, Stirling, United Kingdom
2 School of Medicine, University of Dundee, Dundee, United Kingdom

Background: Type 2 diabetes (T2D) is one of the leading causes of chronic kidney disease (CKD) and an independent risk factor for Acute Kidney Injury (AKI). The unprecedented access to routine healthcare individual-level data offers great promise to improve understanding of disease mechanism, but requires methods tailored to the complexity of the data. The proposed study aims to apply longitudinal methods to evaluate rates of AKI and how this relates to CKD status and renal function decline in patients with and without T2D using electronic healthcare records.

Methods: Study design was a retrospective cohort study. An algorithm was developed to identify AKI episodes from longitudinal serum creatinine measures based on the KDIGO definition. The negative-binomial model was used to evaluate AKI rates in people with and without T2D depending on CKD status. A mixed effect linear model adjusted for demographic characteristics and co-morbidities was developed to evaluate decline in glomerular filtration rate (GFR) before and after an AKI event depending on diabetes and CKD status.

Results: The cohort was formed of 16700 participants with a median follow-up of 8.2 years. 9417 of these had T2D and 7283 had no diabetes. 48.6% (N=4580) of participants with diabetes developed AKI compared to 17.2% (N=1257) of controls. 46.3% (N=4359) of those with diabetes had CKD vs 17.1% (N=1251) of controls. In the absence of CKD, AKI rate was five times higher in people with diabetes than controls (121.5 vs 24.6 per 1000 person-years, Rate Ratio RR=4.9, 95% CI 4.4-5.5), whereas for people with CKD, rate of AKI was twice higher in people with diabetes than controls (384.8 vs 180.0 per 1000 person-years, RR=2.1, 95% CI 1.9-2.4 after CKD date and 109.3 vs 47.4 per 1000 person-years, RR=2.3, 95% CI 1.8-3.0 prior to CKD). Analysis of eGFR longitudinal data showed that people with diabetes had a significant decrease in eGFR slope prior to AKI (compared to those without a prior AKI event). In turn, there was no significant change in eGFR slope post AKI in people with diabetes, whereas further decline in eGFR slope was observed in the control group post AKI.

Conclusion: Rates of AKI are significantly higher in patients with diabetes, however, the long-term effects of AKI on renal function remains poorly understood. Development of methods for analysis of electronic medical records is critical to improve understanding of the underlying mechanism of disease development and progression to help improve prevention and treatment.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-17 Impact of clinic community ART referral on service delivery for HIV patients newly initiated on antiretroviral therapy

Lara Lewis[1,2], Francesca Little[2], Reshma Kassanjee[2,3], Nigel Garrett[1], Jienchi Dorward[1,4]
1 Caprisa, South Africa
2 University of Cape Town, South Africa
3 Centre for Infectious Disease Epidemiology and Research, School of Public Health, University of Cape Town, South Africa
4 Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom

Introduction: Differentiated service delivery (DSD) programmes allow HIV-positive patients who are stable on antiretroviral therapy (ART) to collect their medication from pick-up points outside of healthcare facilities. The decanting of patients to communities should reduce clinic burden whilst increasing staff capacity, potentially enhancing care for acute patients (<12 months since ART initiation) who are ineligible for the programme. We assess whether the proportion of patients referred to DSD in a clinic impacts retention of acute patients.

Methods: De-identified, routinely collected data from 55 clinics in South Africa from January 2018-January 2020 was aggregated by month and clinic. Marginal structural models (MSM) were used to estimate the causal effect of the time-dependent exposure, monthly clinic DSD referral, on the proportion of patients who were retained-in-care 12 months after ART initiation. The proportion of the clinic referred to DSD was measured for each of the 12 clinic-months before the outcome and was categorized into 2 "dose" groups, high (>10%) and low (<10%). We use a weighted linear regression to model the relationship between the proportion of acute patients retained at 12 months and the cumulative "dose" of DSD through the 12-month period. We use stabilized inverse probability of treatment weights, where the denominator of the weights estimated the conditional probability that a cohort of acute patients had their observed DSD "dose" history in the first 12 months on ART. The model of the probability of receiving a high DSD dose at time k is estimated using a pooled logistic regression treating each clinic-month as an observation and includes clinic growth at time k, dose value at k-1 and clinic size at time 0 as covariates.

Results: The mean monthly number of patients visiting a clinic was 1,366, and a median (interquartile range [IQR]) of 10.87% (5.55%-17.30%) were referred to DSD. The mean monthly number of patients initiated on ART in a clinic was 40 and the median (IQR) 12-month retention proportion among them was 64.18% (57.35%-71.05%). Stabilized weights had a mean (range) of 1.00 (0.57-1.64). The weighted estimate (95% confidence interval) of the effect of DSD referral on retention was -0.005(-0.0053;0.0043).

Conclusion: Our data do not suggest that higher DSD referrals facilitate greater retention among acute HIV-positive patients. Future work will incorporate multi-level correlation structure when fitting the MSM, consider a continuous exposure variable and compare the above results to those from a linear mixed model.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-18

Re-examination of the threshold value of HbA1C in therapeutic guideline of type 2 diabetic patients taking circannual rhythms into account using large registry dataset

Masako Nishikawa[1], Masahiro Yuki[2,3], Masaya Sakamoto[2]
1 The Jikei University School of Medicine, Minato-ku, Japan
2 International University of Health and Welfare Mita Hospital, Minato-ku, Japan
3 International University of Health and Welfare Graduate School, Japan

Background: Current therapeutic guideline of type 2 diabetic patients set the threshold value of HbA1c to prevent cardiovascular event or renal deficiencies. Recently, it has been revealed by Sakamoto et al. (2019) that HbA1c, blood pressure, and LDLC may have certain circannual rhythms. The threshold value in the guideline did not take circannual rhythms into account. We tried to quantify the effect of circannual rhythms on the risk of renal event by analyzing a large registry dataset of JDDM Study Group.

Methods: The dataset were accumulated since 2014 until 2019. We translated above medical question to a more concrete statistical expression taking clinical practice into consideration. We formulized as follows: If there should be a seasonal effect, the magnitude of the effect of the threshold value of HbA1c on the events would be different depending on which time point in a year a value more than the threshold was observed for the first time in a patient. We treat such time point as seasonal time-dependent variables in a Cox proportional hazard model. If ignoring the seasonal effect, these time-dependent variables are suppressed into one categorical variable. In the Cox model, age, sex, and duration of diabetics were modelled as ordinary explanatory variables. In this study, we defined renal event as eGFR < 30 ml/min/1.73m2. Patients on treatment more than 6 months and whose first observed HbA1c in 2014 less than the threshold value were included for the analysis set. To mimic real clinical practice, measurement time points will be every 2 to 3 months with some irregular intervals by random number. If the HbA1c are missing on the selected time points of a patient in the registry dataset, single point imputation was performed by a non-linear population PD model for circannual rhythms of HbA1c developed by us in the previous year. Sensitivity analyses were performed.

Results: According to the examined threshold values of HbA1c 7.5, 8, and 8.5, numbers of the included patients for analysis set were 6920 (DatasetA), 8238 (DatasetB), and 8912 (DatasetC), respectively, from the original dataset (N=9949). And 3951(57.1%), 3229(39.2%), and 2285(25.6%) experienced the respective threshold values (HbA1c) until December 2019 before imputation. The goodness of fit of non-linear population PD model was reasonable. Other results will be shown on the Conference.

Conclusion: The concept of HbA1c variation might become more important in near future.

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-19

Predictive monitoring of depression relapse using administrative data

Paulina Von Stackelberg[1], Rob Goedhart[1], Leo Huberts[2], Ş. İlker Birbil[1]
1 University of Amsterdam, Amsterdam, Netherlands
2 University of New South Wales, Sydney, Australia

Major Depressive Disorder (MDD) is a mental disorder with a high lifetime prevalence and one of the leading causes of disability worldwide. Its effects span across several domains, influencing mood, cognition, and social relationships. If left untreated, MDD can lead to a number of complications, including an increased risk of suicide. As many patients experience another depressive episode after being treated [1], predictive monitoring for the risk of relapse is essential for healthcare professionals to be able to follow up on patients and intervene early. Several influencing factors have been identified in the literature, such as comorbid anxiety and the number of previous depressive episodes [2]. Given the large prevalence of MDD in the population, it is important to be able to identify the individuals who have the highest risk of relapse in the group of patients. However, automatically monitoring these large groups requires additional considerations going beyond predictive performance such as data availability, data updating, and interpretability.
In the present study, we investigate how individuals most at risk can be identified using information previously established in existing databases. More specifically, we study the suitability of readily available administrative data for this problem, using a unique dataset of more than 10.000 patients from the Netherlands. In this context, we focus on three problems: 1) Next to the previously suggested influence factors, we investigate to what extent the addition of more detailed available individual information improves predictions. 2) Moreover, we show the effect of the frequency of data updating on model performance when monitoring throughout the year. 3) Lastly, we illustrate differences in predictive performance when choosing between regression- and machine learning-based models when approaching depression relapse monitoring (logistic regression, Lasso/Ridge regression, Random Forest, XGBoost).

Based on our findings, we derive a set of considerations for practical implementations when constructing monitoring methods for mental health care.

References: *[1] Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W., & Dobson, K. (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. Clinical Psychology Review, 41, 16-26. [2] Moriarty, A. S., Meader, N., Snell, K. I., Riley, R. D., Paton, L. W., Dawson, S.,... & McMillan, D. (2022). Predicting relapse or recurrence of depression: systematic review of prognostic models. The British Journal of Psychiatry, 221(2), 448-458.*

POSTER SESSION B-02:
Electronic Health Records and Real-World Data for Observational and Interventional Studies

P-B02-20 LB

Comparison of analysis methods for nested case-control studies with risk set sampling in the presence of tied data

Isao Yokota
Hokkaido University, Sapporo, Japan

Background: Nested case-control designs are increasingly applied in both clinical epidemiology using large medical records and molecular epidemiological studies integrated with biobanks. This is due to their efficiency in handling large datasets and their ability to reduce computational intensity and measurement resources. Risk set sampling is employed to match controls each time an event occurs. In practice, event occurrence is often measured at certain intervals, resulting in tied data. While several analysis methods exist for nested case-control designs, performance comparisons in finite samples, especially in the presence of heavily tied data, remain limited.

Methods: We conducted Monte Carlo simulations to evaluate the performance of analysis methods for nested case-control designs: the conditional logistic model, the unconditional logistic model, and the weighted partial likelihood (WPL) method (Støer and Samuelsen, Stat Med (2013)). Evaluations were carried out under several scenarios, particularly altering sample size, matching ratio, and the event-evaluation intervals (reflecting the heaviness of ties). The performance was assessed using bias and statistical efficiency.

Results: As the matching ratio decreases, resulting in increased sparsity of the strata, both the conditional logistic model and the WPL method show increased bias, with the bias being especially significant in the WPL method. When the amount of tied data increases (resulting in larger stratum sizes), the bias decreases. It was found that the bias caused by the sparsity of the strata was much more significant than the ability to appropriately account for tied data. Additionally, when the event rate is high, the bias in the WPL method becomes more noticeable. Despite these biases, the WPL method demonstrates greater efficiency compared to the conditional and unconditional logistic models.

Conclusion: The performance of the analysis methods varied depending on sample size, event rate, and matching ratio. No single method consistently outperformed the others in all scenarios. Increasing the number of ties led to larger stratum sizes, which improved the bias in both the conditional logistic model and the WPL method. A practical approach for addressing sparse data bias might be to group event times and enlarge stratum sizes.

POSTER SESSION B-03:
Survival Analysis

P-B03-01

Identification and validation of survival-associated hub genes in the VAV1 gene network in acute myeloid leukemia

Gizem Ayna Duran, Cansu Ergun, Yağmur Kiraz Durmaz
Izmir University of Economics, Izmir, Turkey

Background/Introduction: Acute myeloid leukemia (AML) is a bone marrow malignancy and is characterized by the proliferation of immature bone marrow cells. There are many gene mutations and genes with varying expression rates that cause AML disease. In addition, activated signalling pathways also cause the poor course of the disease. PI3K/AKT signalling pathway is one of these pathways. It is known that this pathway is effective in processes such as cell growth, proliferation, transcription, and regulation of metabolic events. In addition, the PI3K/AKT signalling pathway is active in AML, as in many types of cancer. Due to the importance of the pathway in the disease, in a previous study we conducted, the effect of the signalling pathway on survival in AML disease and the expression levels of the genes in the pathway in the disease state were investigated. As a result of research, there is a negative effect of guanine nucleotide factor (VAV1) and related genes on survival, and this gene has become an important focus with the increase in gene expression in AML. VAV1 gene is a signal transducer gene and literature research has shown that it has high expression in many types of malignancies.

Methods: Gene intensity values of AML cancer patients (GSE36642, GSE12417) and healthy individuals (GSE5900) were used from the Gene Expression Omnibus (GEO) database [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. By including the demographic data in the data set, DEG and survival were performed using the "LIMMA" and "survival" packages in the R program. Gene enrichment analyzes will be performed and important upregulated genes in survival will be identified and drugs that have not been used in AML in the clinic and literature will be identified for the purpose of drug repurposing.

Results: In our study, up to 13 hub genes, including the VAV1 gene, were identified among the genes that were found to be upregulated in AML patients, causing poor survival. Drug docking analyses will be performed using these genes and candidate drugs and agents will be identified.

Conclusion: In our study, it is planned to use novel drug administration approaches targeting VAV1 and related hub genes. In this context, it is primarily aimed to identify drugs that have not been widely used in AML and have been tested in non-cancer diseases using drug repurposing methods.

POSTER SESSION B-03:
Survival Analysis

P-B03-02 A maximum penalised likelihood approach for semiparametric accelerated failure time models with time-varying covariates and partly interval censoring

Aishwarya Bhaskaran[1], Ding Ma[1], Benoit Liquet-Weiland[1], Serigne Lo[2], Stephane Heritier[3], Jun Ma[1]
1 Macquarie University, Sydney, Australia
2 Melanoma Institute, The University of Sydney, Sydney, Australia
3 Monash University, Melbourne, Australian

Accelerated failure time (AFT) models are frequently used for modelling survival data. This approach is attractive as it establishes a direct relationship between the time until an event occurs and various covariates. It asserts that the failure times experience either acceleration or deceleration through a multiplicative factor when these covariates are present. Semiparametric AFT models are obtained when the error distribution, or the corresponding baseline hazard function, is unspecified. While existing literature provides numerous methods for fitting semiparametric AFT models with time-fixed covariates, adapting these approaches to scenarios involving both time-varying covariates and partly interval-censored data remains challenging. In this paper, we introduce a maximum penalised likelihood approach to fit a semiparametric AFT model. Firstly, we utilise Gaussian basis functions to construct a smooth approximation of the nonparametric baseline hazard. We then alternate between implementing the Newton-Raphson algorithm and the multiplicative iterative (MI) algorithm to carry out maximum penalised likelihood (MPL) estimation via a constrained optimisation approach. Here, imposing a penalty term allows one to impose extra smoothness in the approximation of the baseline hazard function. It also provides greater flexibility in the estimation of the baseline hazard as the solutions become less dependent on the number of basis functions and the locations of their knots. The results of a comprehensive simulation study indicate that the biases of the regression coefficients for both time-fixed and time-varying covariates are generally small across sample sizes. In addition, the approach yields satisfactory estimates of the baseline hazard function. We also present an implementation of our approach on a data set on melanoma brain metastases. In conclusion, our method allows for the modelling of survival data with partly interval-censored failure times and accommodates both time-fixed and time-varying covariates, with significant findings from the application on melanoma brain metastases.

P-B03-03 Efficient nonparametric estimators of discrimination measures with censored survival data

Marie Skov Breum, Torben Martinussen
Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark

An important objective of many medical studies is to evaluate the predictive accuracy of a multivariable prediction model (scoring rule), which relates a set of prognostic markers to the risk of experiencing a particular event. In a survival setting the predictive accuracy of a scoring rule is typically evaluated using discrimination measures such as the condordance index (c-index) [1, 2] or the cumulative-dynamic time-dependent area under the ROC-curve (AUC) [3]. Often scoring rules are based on a survival regression model such as the Cox proportional hazards model. This has the undesirable feature that the scoring rule depends on the censoring distribution when the model is misspecified. In this research we focus on non-parametric scoring rules defined in the setting where there is no censoring. We view the c-index and the AUC as model-free estimands and use semiparametric theory to propose novel estimators based on the efficient influence functions. The proposed estimators are non-parametric and locally efficient, and allow for covariate-dependent censoring. We verify the theoretical properties of the estimators in a simulation study, and we demonstrate the use of the estimators by analyzing data from a brain cancer study where it is of importance to judge whether a cheaper tumor growth biomarker has a predictive power not worse than a more expensive tumor growth marker.

References: *[1] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, Evaluating the yield of medical tests, Journal of the American Medical Association, 247(18): 2543-2546, 1982. [2] M. Gönen and G. Heller, Condordance probability and discriminatory power in proportional hazards regression, Biometrika, 92(4):965-970, 2005. [3] P. J. Heagerty, T. Lumley, and M. S. Pepe, Time-dependent roc curves for censored survival data and a diagnostic marker, Biometrics, 56(2):337-344, 2000. 1*

POSTER SESSION B-03:
Survival Analysis

P-B03-04

Evaluation of the Survival-inferred fragility index to assess the robustness of the estimated treatment effect on survival endpoints

Roxane Couturier, Sylvie Chevret, Jerome Lambert
INSERM UMR 53 CRESS, Paris, France

Background/Aims: Phase III randomized clinical trials (RCTs) aim to evaluate the benefits of a new treatment on a primary outcome of interest. When conducted in patients with malignancies, most RCTs use a right censored endpoint, with conclusions regarding efficacy based on the p value of the logrank test. Recently, the survival inferred frailty index (SIFI) has been proposed as a measure of the robustness of the treatment effect of such trials. Applied to real RCTs, the reported SIFI values were very low. We hypothesized that such values relied on the contamination of the trial by individuals with either very good or poor prognoses.

Methods: We performed a Monte Carlo simulation study of individuals enrolled in an RCT, generating survival times under several realistic scenarios differing in treatment effects, sample sizes, proportionality of hazards (PH), and amount of censoring. Contamination of the sample by individuals with a very specific prognosis was studied.

Results: Surprisingly, under the null of no treatment effect, the standard SIFI exhibited very low values, poorly sensitive to the sample size. However, under both the null and the alternative hypotheses, the p value and the amount of censoring influenced the value. Under the alternative, violation of the PH assumption also impacted the SIFI. By contrast, randomly selected patients, rather than those selected at the survival tails, widely modified the results, notably with an impossibility of finding value under the null in a large proportion of cases. When contaminating the sample with individuals with very poor or good outcomes, results were close to those of the standard.

Conclusions: The SIFI should be interpreted with caution, given it addresses a very specific setting of individuals. Random selection of patients could otherwise be used in its computation.

POSTER SESSION B-03:
Survival Analysis

P-B03-06

Comprehensive analysis of parametric and semiparametric mixture cure models from low- to high-dimensional covariates settings

Fatih Kizilaslan[1], David Michael Swanson[2], Valeria Vitelli[1]
1 Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Oslo, Norway
2 Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Background: When dealing with survival data, the presence of substantial censoring following a prolonged follow-up period often indicates the presence of cured individuals who may never experience the outcome of interest. This phenomenon can be observed in certain clinical investigations, such as cancer studies, and becomes more noticeable among patients diagnosed in the early stages of disease progression. Furthermore, advancements in cancer treatments have led to a substantial number of patients being classified as cured. In such circumstances, it is reasonable to consider a mixture cure model that combines cured and uncured fractions, rather than using traditional methods in survival analysis, which assume that all individuals in the sample will eventually experience the outcome of interest. In the mixture cure model, the overall population lifetime is defined by weighting the survival time of susceptible and cured patients with the uncured and cured rates, respectively.

Methods: In this study, our aim is to investigate parametric and semiparametric mixture cure models under different settings. First, we develop a novel alternative approach to deal with low-dimensional covariates by incorporating a frailty variable into the mixture cure model, thus developing the enhanced mixture cure frailty model. Subsequently, we explore the extension of this model to the high-dimensional covariates setup. Here, we propose a novel approach that combines the use of a penalization with an adaptive expectation expectation-maximization algorithm. The proposed inferential framework offers the flexibility to accommodate both parametric and non-parametric estimation for the distribution of susceptible patients.

Results: We conduct comprehensive simulation studies to assess the performance of the models, in both the low- and the high-dimensional covariates scenarios. We undertake a comparative analysis with existing methods, including popular R packages such as smcure, to assess the method respective performance particularly in the low-dimensional case. We present the findings from these simulation studies alongside the analysis of real data obtained from a breast cancer study from the Oslo University Hospital.

Conclusion: In conclusion, our study highlights the importance of considering mixture cure models in survival analysis, particularly in scenarios with substantial censoring and a significant proportion of cured individuals, as often observed in cancer studies. By investigating parametric and semiparametric approaches, we provide robust methods that accommodate both low-dimensional and high-dimensional covariates settings, offering a versatile framework for analyzing complex survival data and contributing to improved understanding and management of diseases like breast cancer.

POSTER SESSION B-03:
Survival Analysis

P-B03-05

Assessing different matching methods to evaluate the effectiveness of colchicine in treating myopericarditis

Maria De Martino[1], Valentino Collini[2], Massimo Imazio[1,2], Miriam Isola[1]
1 Department of Medicine, University of Udine, Udine, Italy
2 Cardiothoracic Department, Santa Maria della Misericordia University Hospital, Udine, Italy

Introduction: Colchicine has proven to be effective and safe in treating acute and recurrent pericarditis, but there is a lack of data regarding its efficacy and safety in cases where pericarditis coexists with myocarditis. This gap in knowledge exists because such patients were excluded from previous studies. The objective of this study is to investigate the potential effectiveness of colchicine in reducing recurrences in patients experiencing their first episode of pericarditis with concomitant myocarditis, using and comparing two distinct approaches to address the imbalance in groups (with and without colchicine).

Methods: This retrospective observational cohort study included all patients referred for their initial occurrence of pericarditis with myocardial involvement to two tertiary centres in Northern Italy from January 2016 to June 2021. The first matching method used to generate balanced cohorts was propensity score matching (PSM) single nearest neighbour, without replacement. A region of common support was considered; therefore, observations in the treatment group with propensity scores higher than the maximum or lower than the minimum propensity scores of the controls were excluded. The second matching method used was coarsened exact matching (CEM). The variables considered for matching were cardiovascular risk factors, STsegment elevation, pericardial effusion (PE) and C reactive protein (CRP) values. Balance was measured using the standardized mean differences (SMD). Event-free survival was defined as freedom from recurrence and univariable and multivariable Cox regression analyses were performed.

Results: The mean of the SMDs before matching was 0,430. After matching the mean of the SMDs was 0,221 in PSM case while 0,287 in the CEM case. The choice of the matching technique had minimal influence on the results of the analyses. In fact in both cases in multivariable Cox regression analysis, women and corticosteroid use were independent risk factors for time to recurrence, and colchicine use prevented recurrences.

Conclusions: PSM performed better than CEM, having a lower mean SMD. However the results of the independent risk factors did not differ using the two different analysis.

References: *[1] Collini V et al., Efficacy and safety of colchicine for the treatment of myopericarditis. Heart. 2024*

POSTER SESSION B-03:
Survival Analysis

P-B03-07

Penalised power-generalised Weibull distributional regression

Shirin Moghaddam, Laura McQuaid, Kevin Burke
Department of Mathematics and Statistics (MACSI), University of Limerick, Ireland

Methods: We introduce a flexible penalized multi-parameter modelling framework and investigate its performance through simulation studies and real data analysis. Multi-parameter regression (MPR) modelling refers to the approach whereby covariates enter parametric model through multiple distributional parameters simultaneously (e.g.,scale and shape parameters), allowing more complex covariate effects to be captured; this activity is also known as "distributional regression". We particularly focus on three-parameter (one scale, two shapes) Power Generalized Weibull (PGW) distribution. The PGW distribution encompasses key shapes of hazard function (constant, increasing, decreasing, up-then-down, down-then-up) and a variety of common survival distributions (Weibull, log-logistic, Gompertz). This allows for a highly flexible approach for modelling survival data.

Results: The PGW is applied to a lung cancer dataset where the response variable is time to death. In this dataset, we observed non-proportional hazard curves, and as such the hazard ratios exhibited time-varying behaviour. This attribute leads a proportional hazards model to be inappropriate. Also, the radiation hazard curves were non-monotonic and consequently many distributions like Weibull would be unsuitable. Our finding shows that our proposed model only selects variables in the shape of the model if it is required and does not overcomplicate the model if a simpler model is more appropriate.

Conclusion: We develop a flexible survival modelling framework that can handle a wide variety of scenarios that arise in real-world applications. This is achieved using PGW distribution where multiple distributional parameters will have the ability to depend on covariates.

References: *[1] McQuaid L (2023). A Penalised Power-Generalised Weibull Distributional Regression Model (2023). MSc Thesis, University of Limerick. [2] Jaouimaa F-Z, Ha I-D, Burke K (2023). Penalized variable selection in multi-parameter regression survival modeling. Statistical Methods in Medical Research. [3] Burke, K and Jones, MC and Noufaily, A (2020).A flexible parametric modelling framework for survival analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics). [4] Burke, K and MacKenzie, G (2017). Multi-parameter regression survival modeling: An alternative to proportional hazards. Biometrics.*

POSTER SESSION B-03:
Survival Analysis

P-B03-08

Cox regression for comparing a mixture of personalized treatment and standard of care to standard of care

Marilena Müller, Dominic Edelmann
Dkfz Heidelberg, Heidelberg, Germany

Background/Introduction: We consider a two-arm clinical trial in precision oncology with time-to-event endpoint, e.g. progression-free-survival. The control arm consists of standard of care (SOC) whereas patients in the treatment arm are offered molecularly targeted treatment. However, some patients in the treatment arm do not receive targeted treatment because it is not available or patients do not consent. Instead they also receive SOC. Applying the intention-to-treat principle hence involves comparing the outcomes of patients receiving a mixture of treatment and SOC to patients receiving solely SOC. Suppose that each patients' progression can be described by a Cox model via its conditional intensity function. Additionally to estimating the conditional intensity for event occurrence in each group, testing procedures on the magnitude of difference between arms are explored.

Methods: Initially, we suppose that the conditional intensity of events follows a Cox model with proportional hazards. The proportional hazards assumptions for the intention-to-treat comparison is discussed. We use Maximum Likelihood methods on the partial likelihoods of the components representing individual patients and both arms simultaneously. Counting process theory as well as martingale theory is used to develop suitable test statistics.

Results: A regression model is presented and respective results are discussed. Testing procedures for survival differences in the treatment arms are stated and the influence of different types of treatment effects is investigated. Several simulation studies show the versatility of the approach as well as possible restrictions in usage.

Conclusion: We propose guidelines on how to account for possibly various treatment arms and presence of mixtures as well as giving insight on when it is neccessary or appropriate to apply a more rigorous model. Testing results enable an overview of model and test statistic choices, especially in regard to disscussions about proportional hazards and the Cox model.

POSTER SESSION B-03:
Survival Analysis

P-B03-09

Translating time to event data using the mean residual function

Parastoo Niloofar[1], Shirin Moghaddam[2], Amirhossein Jalali[3], Alberto Alvarez-Iglesias[4], John Newell[1]
1 School of Mathematical and Statistical Sciences, University of Galway, Galway, Ireland
2 Department of Mathematics and Statistics (MACSI), University of Limerick, Limerick, Ireland
3 School of Medicine, University of Limerick, Limerick, Ireland
4 HRB Clinical Research Facility Galway, University of Galway, Galway, Ireland

Introduction: The Mean Residual Life (MRL) function gives an attractive summary of time to event data as it can provide a clear and simple translation of the expected time remaining.

Methods: Estimating the MRL is difficult due to the presence of administrative censoring. A hybrid estimator for the MRL has been suggested[1] combining a non-parametric Kaplan-Meier estimator with a parametric component. Methods to calculate the variance estimate for such a hybrid estimator at the start of follow-up time t=0 have been proposed[2]. As no closed form for the variance of this hybrid estimator is available, the bootstrap approach can be used to assess the variability in the MRL function at any arbitrary time point. Moghaddam et al[3] proposed a Bayesian approach to impute the censored observations which allows a simple parametric approach to be used to estimate the MRL as censoring is no longer an issue.

Results: The hybrid and Bayesian approaches are discussed and compared using data from clinical trials and observational studies. Variations of the MRL plot including plots of the between group differences and their ratios will be presented. In addition, the authors new R package to create dynamic nomograms for the MRL function conditioning on covariates will be demonstrated.

Conclusion: The concept of Translational Statistics[4] was proposed to facilitate the integration of biostatistics within clinical research to enhance communication of statistical research findings in an accurate and accessible manner to diverse audiences. The use of appropriate visualisation is central to all areas of statistical research. An advantage of the MRL function over plots of the estimated survival curves and hazard rates is that the MRL is presented in units of time rather than probabilities or rates making it a potentially useful translational tool.

References: *[1] Alvarez-Iglesias, A., Newell, J., Scarrott, C., & Hinde, J. (2015). Summarising censored survival data using the mean residual life function. Statistics in Medicine,34(11), 1965-1976. [2] Gong, Q., & Fang, L. (2012). Asymptotic properties of mean survival estimate based on the Kaplan Meier curve with an extrapolated tail. Pharmaceutical Statistics, 11(2), 135-140. [3] Moghaddam, S., Newell, J., & Hinde, J. (2022). A Bayesian approach for imputation of censored survival data.Stat,5,89–107. [4] McCabe, G. P., & Newell, J. (2022). The art of translational statistics.Stat,11(1).*

POSTER SESSION B-03:
Survival Analysis

P-B03-10

Examination on analyses for progression free survival as interval-censored data

Tomomi Nishikawa[1], Masako Nishikawa[2], Yusuke Saigusa[3], Toshihiko Morikawa[4], Kouji Yamamoto[3]
1 National Cancer Center Hospital East, Kashiwa, Japan
2 The Jikei University School of Medicine Clinical Research Support Center, Japan
3 Yokohana City University School of Medicine Department of Biostatistics, Yokohama, Japan
4 Biostatistics Center, Kurume University (formerly), Japan

Background/Introduction: A variety of statistical methods for interval censored data have been actively investigated and developed in various disease areas since 1970s, when research on the new infectious disease HIV/AIDS attracted attention. In oncology, progression free survival times are obtained as partially interval censored data, and the usual survival time analysis method is often applied after imputing right point of the censoring interval. However, other single point imputation methods such as left point or midpoint imputation also could be candidates, and a lot of research has been done on the statistical properties of single/multiple imputations, but it has been limited only to simulation studies.

Methods: We examined the characteristics of deterministic (left-, mid, and right point) imputations from a theoretical aspect when true distribution is the exponential distribution. Then, we compared the performance of these three deterministic single point imputation methods and Finkelstein's proportional hazards model without imputation by simulation, in the setting of the selection design, so called a pick the winner design, in which the most promising treatment is identified based on the point estimate. We considered realistic situations in practice on simulation, setting allowance of inspection date, several combinations of censoring proportion to each arm, and probability with which patients miss scheduled inspection, under the constant/non constant inspection intervals over the entire study period. As evaluation criteria, the correct selection probability and bias(MSE) of a regression coefficient over 10000 simulations were used.

Results: Theoretically, in the case of single point imputation, it was shown that not right-point, but mid-point imputation is preferable. In the simulation results, in the case of different censoring proportion between two groups, if the inspection should be performed at the date of study discontinuation, mid-point imputation showed the best performance of selection design among four comparative methods including Finkelstein's model. On the other hand, if the inspection is not performed at the date of study discontinuation, left-point imputation showed the best performance of selection design. On Finkelstein's method, the same results as each imputation of mid or left-point were obtained. The distribution of the regression coefficient estimates was skewed and showed a discrepancy from a normal distribution, depending on the simulation scenario.

Conclusion: Midpoint imputation was theoretically the best. In our simulation, midpoint or left-point imputation showed the best performance, depending on the simulation condition.

Reference: *[1] Nishikawa et al. (2024). Theoretical examination and simulation study on analyses for progression free survival as interval-censored data. Japanese Journal of Biometrics, 45 in press(in Japanese).*

POSTER SESSION B-03:
Survival Analysis

P-B03-11

Information sharing in fractional polynomial network meta-analysis where studies report different summaries and subgroup categories: 2nd line treatments for Non-Small Cell Lung Cancer (NSCLC)

Hugo Pedder, Ayman Sadek, Nicky Welton
Population Health Sciences, University of Bristol, Bristol, United Kingdom

Background/Introduction: Network meta-analysis (NMA) using fractional polynomials allows comparison of multiple treatments from trials reporting time-to-event outcomes, without assuming proportional hazards. Ideally we would have individual participant data (IPD) for each study, or reconstructed IPD from published Kaplan-Meier plots. However, in a review of studies of 2nd line treatments for advanced/metastatic non-small cell lung cancer (NSCLC), different studies report results for different PD-L1 status categories, and some studies only report a hazard ratio (HR). This study aimed to combine the reported evidence in a unified fractional polynomial NMA model to compare 2nd line treatments for advanced NSCLC in 12 distinct subgroups depending on histology, PD-L1 status, and genetic variant.

Methods: We fitted fractional polynomial NMA models to progression-free survival (PFS) where the powers of the fractional polynomials were shared across the different comparisons and subgroups. We used a shared-parameter approach to account for different reporting of PD-L1 subgroups, where studies reporting combined categories were modelled as a weighted average of the more detailed categories. Studies which only a reported a HR, provided information on the relative effects for the first fractional polynomial parameter only, which implicitly assumes proportional hazards for those treatments. We assessed the validity of the proportional hazards assumption for treatments where reconstructed IPD were available.

Results: We estimated plots of implied HRs over time which are helpful to highlight how combined proportional and non-proportional hazard models are synthesised. As expected, immunotherapies display clear evidence of non-proportional hazards compared to docetaxel, justifying the use of the fractional polynomial NMA approach. However, for other treatments where Kaplan-Meier data are available, the proportional hazards assumptions was more reasonable. This supported the inclusion of studies reporting HRs alone. PDL1 status is a clear effect modifier for immunotherapies, with patients with lower PDL1 clearly eliciting less benefit. The sensitivity of our approach for synthesising studies with different PDL1 thresholds is shown on PFS predictions.

Conclusion: The information-sharing approach enabled us to combine data reported on subgroups categorised in different ways to make PFS predictions in multiple subgroups defined by histology, PD-L1, and genetic variant, which were required to inform an economic model. However, the sensitivity of results to assumptions made need to be tested. For the NSCLC example it was possible to assume proportional hazards for some treatments but not others, which was plausible given the different mechanisms of action of the treatments, and was supported by our analyses.

POSTER SESSION B-03:
Survival Analysis

P-B03-12

Lead time bias correction in breast cancer screening studies

Marius Robert[1], Brice Amadeo[1], Marie Poiseuil[1], Maja Pohar-Perme[2], Virginie Rondeau[1]
1 University of Bordeaux, Bordeaux, France
2 University of Ljubljana, Ljubljana, Slovenia

Background: Mammography screening plays a crucial role in detecting and diagnosing breast cancer, enabling early treatments and thereby improving survival rates. Evaluating the effectiveness of screening involves comparing the survival of patients diagnosed through screening with those diagnosed after the onset of symptoms. However, this analysis is susceptible to biases that may overestimate the benefits of screening. The lead-time bias arises because screening allows for the early diagnosis of breast cancer, resulting in a longer observed survival without a true improvement in the date of death.

Methods: Two commonly used approaches to solve this bias were applied to real data in order to compare them. The first approach is based on a multi-state model in order to estimate the sojourn time in the preclinical phase, i.e., the period during which the tumor is detectable through screening without causing symptoms. The correction proposed by Duffy et al. (2008) subtracts from the observed survival the expected value of the sojourn time distribution. We propose enhancing the existing correction by considering breast density. The second approach, developed by Abrahamsson et al. (2020), is based on a continuous tumor growth model. Its spirit is to prolong tumor growth in patients diagnosed by screening to estimate the time until a symptomatic diagnosis depending on the tumor volume.

Results: These two methods, applied to breast cancer registry data, do not result in corrections of the same order of magnitude. For screen-detected patients, uncorrected 5-year survivals, corrected using tumor growth and multi-state approaches, are estimated at 0.95, 0.93, and 0.87, respectively. The correction based on tumor growth is the weakest. The second gives a 5-year survival comparable to that of patients diagnosed through symptoms (0.85), which suggests a minimal benefit of screening on survival.

Conclusion: The Duffy correction is generally deemed excessive, and after comparison, we consider it as the least accurate. The Abrahamsson correction is promising but remains underutilized due to its recent introduction and complexity. We provide an R package for lead time correction to simplify the use of both methods.

References: *[1] Abrahamsson et al. (2020). Continuous tumour growth models, lead time estimation and length bias in breast cancer screening studies. Statistical Methods in Medical Research. [2] Duffy et al. (2008). Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. American journal of epidemiology.*

POSTER SESSION B-03:
Survival Analysis

P-B03-13 Estimation of the Kendall's Tau in cluster randomised trials with right censored time to event outcomes

Mathieu Roche[1], Elsa Tavernier[1,2], Hermine Bierme[3], Agnes Caille[1,2]
1 INSERM, SPHERE, U1246, Tours University, Nantes University, Tours, France
2 Clinical Investigation Centre, INSERM 1415, CHRU Tours, Tours, France
3 CNRS, IDP UMR 7013, Tours University, Tours, France

Introduction: and Objective(s): In cluster randomized trials (CRT), clusters of individuals are randomized rather than individuals themselves. In these trials, the correlation of individual's outcomes in a cluster must be taken into account in the sample size calculation and in the analysis and a measure of correlation must be reported. The intra cluster coefficient (ICC) is the most used measure of correlation in CRT. A study made by Kalia et al [1] showed that the estimation of the ICC in CRT with right-censored time-to-event outcomes using either censoring indicators (binary outcomes) or observed event times (continuous outcomes) leads to a negative bias, which increase proportionally with the increasing frequency of administrative censoring when using observed event times. The Kendall's tau can be used as an alternative to the ICC, however we presently lack knowledge regarding its estimation with censored observations in a multivariate context. The objective is to assess the association of Kendall's tau with the censoring rate and to compare several estimation methods of Kendall's tau in CRT with right-censored time to event outcomes.

Method(s): We will conduct a simulation study as in the Wang and al [2]. Correlated time to event outcomes will be generated using a Clayton's copula in a simulated CRT. We will vary the number of clusters, the cluster sizes, the censoring rates (administrative and random). We will estimate the Kendall's tau with both a copula based estimator and a gamma frailty based estimator.

Results: We will report the Kendall's tau according to the simulation's parameters. We will report the bias and the standard error of the two estimators with respect to the censoring rate.

Conclusion: The final objective of the study is to help the planification and the reporting of CRT.

References: *[1] Kalia S et al. "On the estimation of the intracluster correlation for time-to-event outcomes in cluster randomized trials", Statistics in medicine, 35(30): 5551-5560. [2] Wang X et al. "Improving sandwich variance estimation for marginal Cox analysis of cluster randomized trials", Biometrical Journal 2022; 65: 2200113.*

P-B03-14 Identification of predictive factors using tree-based models for survival outcomes

Maral Saadati[1], Nicholas Schreck[1], Julia Krzykalla[3], Wiebke Hielscher[2], Axel Benner[1]
1 German Cancer Research Center (DKFZ), Heidelberg, Germany
2 Hochschule Mannheim, University of Applied Sciences, Mannheim, Germany
3 Boehringer Ingelheim, Biberach an der Riss, Germany

In many clinical applications the identification of variables that modify response to treatment (predictive factors) is an important task with the goal to discern subgroups of patients that may benefit from a specific drug or not. In statistical terms this problem is often addressed by including an interaction term in the model. However, it is known that this approach is very much lacking in power, cf. Brookes et al.[1]. The identification of predictive factors has previously been tackled by Krzykalla et al. [2] using a model-based tree approach by Seibold et al. [3] for binary and normally distributed endpoints in a randomized clinical trial setting. In the current project, we extend the work by Krzykalla et al. to accommodate for time-to-event endpoints, such as overall survival and event-free survival, which are particularly relevant for clinical trials. Using the model-based trees approach it is not possible to apply the ubiquitous semi-parametric Cox proportional hazards model. Two solution strategies are considered: (1) fully parametric Weibull survival models, (2) removal of censoring via pseudo values and subsequent model fit via generalised estimating equations. We present a comparison of the proposed strategies in a simulation study as well as real-life data from a randomized trial in acute myeloid leukemia with mutation information.

References: *[1] Brookes et al., Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess. 2001;5(33):1-56. DOI: 10.3310/hta5330 [2] Krzykalla et al., Exploratory identification of predictive biomarkers in randomized trials with normal endpoints. Statistics in Medicine. 2020;39:923–939. DOI: 10.1002/sim.8452 [3] Seibold et al. Model-based recursive partitioning for subgroup analyses. Int J Biostat. 2016;12(1):45-63. doi: 10.1515/ijb-2015-0032*

POSTER SESSION B-03:
Survival Analysis

P-B03-15

Peeling off the hazard layers: A comparison of relative survival techniques to disentangle different components of excess mortality in vulnerable groups during a crisis

Marije Sluiskes, Eva Koster, Jelle Goeman, Hein Putter, Mar Rodríguez-Girondo, Liesbeth de Wreede
Leiden University Medical Centre, Leiden, Netherlands

Background/Introduction: Vulnerable or frail groups face a heightened mortality risk compared to the age- and sex-matched general population. In extreme situations, such as war or a pandemic, the general population itself may also encounter increased mortality rates. It is relevant to investigate whether vulnerable groups experience an excess hazard during a crisis, that is, excess hazard on top of the sum of their existing excess hazard attributed to their vulnerability and the excess hazard experienced by the entire population during the crisis. We consider care-home residents in the Netherlands: did they experience an additional layer of excess mortality during the Covid-19 pandemic?

Methods: To answer this question, we use and expand upon relative survival techniques. Relative survival models split the observed overall mortality hazard in a background and an excess part. Existing population lifetables, stratified by age and sex, are used for the background hazard. We move beyond the classical setting: we do not only compare survival of a vulnerable group (i.e. care-home residents) with that of the matched general population in the same time period, but also compare survival of the same population in different time periods (i.e. before and during the pandemic). We compare two approaches for estimating the excess hazard. The first involves fitting three conventional Cox-based semiparametric relative survival models sequentially, where in each step one of the excess hazard components is estimated. The second approach entails fitting a single additive hazards model. This approach can deal with negative excess hazards, and when one has access to individual-level population data, it offers the additional advantage of an easy incorporation of multiple adjustment variables in the background hazard.

Results: We illustrate the two approaches using data from Statistics Netherlands describing all Dutch inhabitants aged 65+ and discuss methodological properties and model fit. We show whether care-home residents in the Netherlands experienced excess mortality during the Covid-19 pandemic.

Conclusions: The models we present allow to disentangle different components of excess mortality in vulnerable groups during crises. Both Cox-based relative survival models as well as additive hazards models can be used to estimate the excess component. With additive hazards-based methods the excess hazard can be estimated with a single model, which makes it possible to incorporate covariates in the background hazard in a straightforward way, and they can deal with negative excess hazards.

POSTER SESSION B-03:
Survival Analysis

P-B03-16

How to make fair comparisons of population-based survival before, during, and after the COVID-19 pandemic: The role of stage-standardisation

Rachael Stannard[1], Paul Lambert[1,2], Sam Khan[1], Mark Rutherford[1]
1 University of Leicester, Leicester, United Kingdom
2 Karolinska Institutet, Stockholm, Sweden

Background: Patient and Public Involvement and Engagement (PPIE) ensures that research is designed and conducted in a way that is most beneficial to those it will impact. PPIE is commonplace in applied research but is seen much less often in statistical methodology research. Barriers to PPIE input in methodological work include less direct patient benefit or pre-conceived notions that public contributors are unable to contribute meaningfully to complex statistical methods. Even where researchers recognize the importance of PPIE in their methodology research, they may not know where to start or fear 'tokenism'.

Methods: The University of Leicester PPI-SMART group has developed a series of resources to facilitate the implementation of PPIE in statistical methodology research. An initial survey was conducted to identify researchers' perceptions of PPIE and barriers faced. An animation was developed to explain to public contributors how PPIE can influence methodological research in an accessible manner. A secondary animation also describes examples of using PPIE in methodological research in practice. To help public contributors' understanding of technical language, a glossary of statistical terminology was developed using simple ('lay') terminology. To disseminate these findings and developed resources to methodological researchers, a workshop was hosted combining statistical and public contributor perspectives.

Results: The survey found polarizing attitudes towards PPIE and highlighted the necessity for guidance and resources to aid implementation. Public contributors provided feedback throughout the development of the animations and glossary, ensuring they effectively communicated to the target audience. The workshop provided an opportunity for researchers and public contributors to discuss the most effective ways to incorporate PPIE into methodological research and identify areas for further exploration.

Conclusion: Developing these resources has identified significant scope for improving the uptake and delivery of PPIE in methodological research, and increasing awareness of the benefits to researchers, patients, and the public. We have developed resources to facilitate better understanding of statistical methodology research by public contributors, to enable them to engage with, and therefore influence, methodological research. Involving patients and the public in developing these resources has led to adapting communication styles to suit different audiences, including the use of animations, analogies, and infographics. Although methodological research has less apparent direct patient benefit, methodological research is the foundation for appropriate clinical decision making. As such, the research community is responsible for ensuring that patients and the public are informed and engaged with methodological research.

POSTER SESSION B-03:
Survival Analysis

P-B03-17

Season of neonatal discharge and risk of unplanned Paediatric Intensive Care Unit admission: A use of flexible parametric modelling

Yuhe Wang[1], Tim van Hasselt[1], Shalini Ojha[2], Cheryl Battersby[3], Chris Gale[3], Elizabeth Draper[1], Sarah Seaton[1]
1 University of Leicester, Leicester, United Kingdom
2 University of Nottingham, Nottingham, United Kingdom
3 Imperial College London, London, United Kingdoms

Introduction: Children born very preterm (before 32 weeks of pregnancy) are at a higher risk of admission to Paediatric Intensive Care Units (PICUs) after discharge home from neonatal (newborn) care. We aimed to examine how the season of the year of neonatal discharge may alter this risk using a survival analysis approach.

Methods: We included nationwide data for very preterm babies, admitted to neonatal units in England and Wales in 2013-2018, with a follow-up period until their second birthday. We proposed a flexible parametric model (FPM) approach using splines on the log cumulative hazard scale to examine unplanned PICU admission after discharge from the neonatal unit using linked national datasets (National Neonatal Research Database and Paediatric Intensive Care Audit Network). The model included eight clinically recommended risk factors (e.g., gestational age, gender) and we used a time-dependent effect for season after considering likelihood ratio tests and plotting of Schoenfeld residuals. The number and locations of knots for the FPM were selected using comparisons of Akaike Information Criterion and Bayesian Information Criterion.

Results: We included 38,706 very preterm babies in our analysis, of whom 1,823 had an unplanned PICU admission. We fitted a FPM with a baseline effect of four knots and a time-dependent effect for season with three knots. Knots were positioned for optimal AIC and BIC, including two boundary knots at minimum and maximum event times, two internal knots at the 33rd and 67th percentiles for the baseline hazard function, and one internal knot at the 50th percentile for the time-dependent function. The hazard ratio (HR) of unplanned PICU admission for a baby born at 24 weeks is 2.14 (95%CI: 1.70 to 2.69) compared with a baby born at 31 weeks. Neonatal discharge in both autumn and winter exhibit decreasing trends in HR over time. Due to the time-dependent effect, the HR for autumn started at 2.30 (95% CI: 1.80 to 2.90) and decreased to 1.71 (95%CI: 1.48 to 1.96) by 100 days.

Conclusion: We have used a FPM approach to explore the risk of PICU admission for babies born very preterm, which is a novel statistical approach in this clinical area. The season of neonatal discharge on the unplanned PICU admission was significant and had a time-dependent effect. Results from these analyses benefit both clinicians and parents when considering the hazard of PICU admission for these children over time.

POSTER SESSION B-03:
Survival Analysis

P-B03-18 Bayesian inference of Case II interval-censored data with non-proportionality

Yuh-Jenn Wu[1], Li-Hsueh Cheng[2], Jr-Chang Chen[3], Wei-Quan Fang[4], Chin-Fu Hsiao[2]

1 Department of Applied Mathematics, Chung Yuan Christian University, Chung Li, Taiwan
2 Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan Town, Miaoli County, Taiwan
3 Department of Computer Science and Information Engineering, National Taipei University, New Taipei City, Taiwan
4 Division of New Drug, Center for Drug Evaluation, Taipei, Taiwan

Background: The case II interval-censored data, a mixture of left-, right-, and interval-censored observations, are everywhere in clinical practice but more complicated than right-censored data that are often used in clinical researches. Although Cox model has been widely applied in right-censored data under proportionality assumption, little is known about statistical inference on case II interval-censored data under non-proportionality.

Methods: In this work, we proposed a Bayesian approach to infer the Cox proportional hazard model for case II interval-censored data, while considering time-varying coefficients which allow us to address non-proportionality. Specifically, the time-varying coefficient prior is constructed with a piecewise constant function, and the cumulative baseline hazard prior with a Bernstein polynomial function.

Results: We provide a theoretical proof for the Kullback-Leibler divergence type consistency of the posterior model. We evaluated the model performance with a simulation study. We further illustrate our model with the analysis of a case II censored hemophilia data. All these analytical results coincided with each other, that is, achieving statistical consistency, thus rendering a profile for robustness of the proposed approach.

Conclusion: In this work, we prove the consistency when two covariates are independent and discrete. However, implementing the similar approach can show the consistency even when two covariates are not independent as long as their joint mass function is bounded. Also, this can be generalized to continuous covariates.

POSTER SESSION B-03:
Survival Analysis

P-B03-19

Effective sample size for the Kaplan-Meier estimator: An intuitive measure of uncertainty?

Toby Hackmann[1], Doranne Thomassen[1], Saskia le Cessie[1,2], Hein Putter[1], Liesbeth de Wreede[1], Ewout Steyerberg[1]
1 Department of Biomedical Data Sciences, Leiden, Netherlands
2 Department of Clinical Epidemiology, Leiden, Netherlands

Background/Introduction: Clinical prediction models for shared decision-making and risk prediction are becoming increasingly popular. Most prediction models prioritize the accurate estimation and clear communication of point predictions. Uncertainty around the point prediction is often expressed by confidence intervals or left out altogether. To present prediction uncertainty in a possibly more intuitive way, the concept of effective sample size was recently introduced for linear and generalized linear models, but not yet for survival outcomes.[1] Our goal is to provide estimates of the effective sample size for the Kaplan-Meier estimator.

Methods: Effective sample size is defined as the hypothetical sample size of patients with the same characteristics (with respect to the model) as a new patient whom the prediction is for, such that the variance in that sample would be the same as the prediction variance. Using this definition, we estimate the effective sample size for the Kaplan-Meier estimator, based on Greenwood's formula for the variance. We investigate how the effective sample size estimation is affected by events, censoring and follow-up time, analytically and in simulations.

Results: Without censoring, the effective sample size for the prediction of survival at a specific time is equal to the full sample size at time 0. Effective sample size for the Kaplan-Meier estimator decreases when variance increases due to censoring. We demonstrate the use of effective sample size on the 'colon' data from the 'survival' package in R, where we find it to be higher than the number of patients at risk, as often reported below Kaplan-Meier plots. While the number at risk informs us about the uncertainty of future changes in the survival function, the effective sample size quantifies the uncertainty from time 0.

Conclusions: Effective sample size can express the uncertainty for the Kaplan-Meier(-based) estimator. Future studies should clarify its role to communicate uncertainty around the point prediction of survival probabilities. Extensions to effective sample size for individualized predictions based on Cox' proportional hazards model or Aalen's additive hazard model can be further developed, starting from the basis of effective sample size for the Kaplan-Meier estimator.

Reference: *[1] Thomassen, D., le Cessie, S., van Houwelingen, H. C., & Steyerberg, E. W. (2024). Effective sample size: A measure of individual uncertainty in predictions. Statistics in Medicine. https://doi.org/10.1002/sim.10018*

POSTER SESSION B-03:
Survival Analysis

P-B03-20

The estimation of COVID-19 vaccine effectiveness by time since vaccination in a population-level observational cohort study using linkage of electronic medical records and survival models with time varying exposures

Reshma Kassanjee[1], Mary-Ann Davies[1,2,3,4], Alexa Heekes[1,4], Hassan Mahomed[5,6], Anthony J Hawkridge[7], Milani Wolmarans[8], Erna Morden[4,9], Theuns Jacobs[4], Cheryl Cohen[10,11], Harry Moultrie[12], Richard J Lessells[13], Nicolette Van Der Walt[14], Juanita O Arendse[5,14], Hilary Goeiman[15], Vanessa Mudaly[2,15], Nicole Wolter[10,16], Sibongile Walaza[10,11], Waasila Jassat[17,18], Anne von Gottberg[10,16], Patrick L Hannan[19], Petro Rousseau[8], Daniel Feikin[20], Keith Cloete[15], Andrew Boulle[1,2,3,4]

1 Centre for Infectious Disease Epidemiology and Research, School of Public Health, University of Cape Town, South Africa
2 Division of Public Health Medicine, School of Public Health, University of Cape Town, South Africa
3 Centre for Infectious Diseases Research in Africa, University of Cape Town, South Africa
4 Health Intelligence, Western Cape Government Department of Health and Wellness, South Africa
5 Division of Health Systems and Public Health, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa
6 Metro Health Services, Western Cape Government Department of Health and Wellness, South Africa
7 Rural Health Services, Western Cape Government Department of Health and Wellness, South Africa
8 South African National Department of Health, South Africa
9 School of Public Health, University of Cape Town, South Africa
10 Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, South Africa
11 School of Public Health, University of the Witwatersrand, South Africa
12 Centre for Tuberculosis, National Institute for Communicable Diseases of the National Health Laboratory Service, South Africa
13 KwaZulu-Natal Research Innovation & Sequencing Platform, University of KwaZulu-Natal, South Africa
14 Emergency & Clinical Services Support, Western Cape Government Department of Health and Wellness, South Africa
15 Western Cape Government Department of Health and Wellness, South Africa
16 School of Pathology, University of the Witwatersrand, South Africa
17 Health Practice, Genesis Analytics, South Africa
18 Division of Public Health Surveillance and Response, National Institute for Communicable Diseases of the National Health Laboratory Service, South Africa
19 Division of Epidemiology and Biostatistics, School of Public Health, University of Cape Town, South Africa
20 Department of Immunizations, Vaccines, and Biologicals, World Health Organization, Switzerland

Background: Limited data exist on the real-world effectiveness of COVID-19 vaccines and boosters in resource-limited settings, characterised by high levels of prior infection, limited vaccine coverage, and competing health service priorities. By leveraging available observational data, we assessed the association between vaccination and severe COVID-19 among adults in South Africa's Western Cape province. Ongoing updates of this study's results were utilised in real-time to inform the vaccination programme.

Methods: The Provincial Health Data Centre consolidates person-level longitudinal data from multiple public sector healthcare information systems and utilises data points to infer health conditions. By linking records with vital registration, SARS-CoV-2 testing, and COVID-19 vaccine and outcome data, we conducted an observational cohort study of adults (aged ≥18 years) in the Western Cape during 2020-2022. Multivariable Cox models were employed to estimate the association of Pfizer-BioNTech's two-dose BNT162b and Janssen's one-dose Ad26.COV2.S vaccination with severe COVID-19. The primary exposures were time-varying vaccination status and time since the most recent vaccine/booster. Severe COVID-19 was identified by SARS-CoV-2 diagnoses sufficiently close to hospitalisations or death.

continued

POSTER SESSION B-03:
Survival Analysis

P-B03-20 The estimation of COVID-19 vaccine effectiveness by time since vaccination in a population-level observational cohort study using linkage of electronic medical records and survival models with time varying exposures

continued from previous page

We adjusted for demographic characteristics, time-varying health conditions, socio-economic status, and healthcare utilisation proxies. We separately analysed periods dominated by different SARS-CoV-2 (sub)lineages and studied 6-week rolling windows of time.

Results: The cohort (n=2,429,927) comprised 60% females, had a median age of 38 years, and carried a high burden of health conditions. During the five waves of COVID-19 outcomes, 0.5% of this cohort experienced death (outcome 1), 0.7% a hospitalisation with severe disease or death (outcome 2), and 1.6% any hospitalisation or death (outcome 3); yet by end 2022, only 41% had completed vaccination and 8% received a booster. Recent vaccination/boosting (within 6 months) was associated with reductions in COVID-19 outcomes during all periods: vaccine effectiveness was 46-92%, 45-92% and 25-90% (estimates across periods) for outcomes 1 to 3 in order. However, the protective effect was lower or absent at longer times since vaccination/boosting. We found no evidence of differences in vaccine effectiveness for the 13% of people living with HIV.

Conclusions: Continued regular COVID-19 vaccination and boosting is important even in resource-limited settings with widespread COVID-19 infection before vaccine availability, and there should be affordable access to newer vaccines. While using observational data poses challenges such as unmeasured confounding and selection or information bias, we were able to provide ongoing monitoring of vaccines at a population-level in one of the largest such studies in Africa.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-01 Multicentre flexible calibration curves with binary outcomes using random effects meta-analysis

Lasai Barreñada[1], Laure Wynants[1,2,3], Ben Van Calster[1,3,4]
1 Department of Development and Regeneration, KU Leuven, Belgium, Leuven, Belgium
2 Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands
3 Leuven Unit for Health Technology Assessment Research (LUHTAR), KU Leuven, Leuven, Belgium
4 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

Background: When externally validating clinical prediction models, flexible calibration curves assess the correspondence between estimated risks and outcome proportions. External validation studies using multicenter or IPD (individual patient data) are increasingly common, and important to study performance heterogeneity. The aim of this study is to present approaches to construct flexible calibration curves while considering clustering due to multicenter or IPD data.

Methods: In the first approach, we pool center-specific quantiles using bivariate random effects meta-analysis of event proportion and mean estimated risk. In the second approach, we develop a two-step model where we first fit a flexible curve per center and in a second step we use a random effects meta-analysis to obtain a multicenter curve. In the third approach, we use a mixed model with restricted cubic splines to estimate observed proportions with random intercept for each center. We generated synthetic data (N = 10.000.000, 10 centers) based on real data to evaluate the ADNEX model that estimates the risk of malignancy of an ovarian tumor. We generated two sets of true risks based on linear (logistic regression) and non-linear (random forest) models. We sampled 200 patients from each center and estimated the observed proportion according to logistic calibration (ignoring clustering), logistic calibration per center (flexible and non-flexible) and the three proposed methodologies. We evaluated the mean squared difference between the predicted risks and the true risks per patient and repeated the process 1000 times. We also conducted a simulation study with a data generating mechanism based on a logistic regression model with random intercepts and evaluate the performance and fitting of the different methodologies in an hypothetical average center under different settings.

Results: Our results on synthetic data showed that the best working methodology was the mixed calibration model with median error of 0.021 (IQR 0.020-0.022) when the truth was linear and 0.026 (IQR 0.025-0.027) when it was non-linear. Logistic calibration ignoring clustering and quantile methods got the worst results in both situations. Simulations showed that the best performing method for estimating the observed proportion in an average center was always the mixed model, with better performance when more centers and more intra-cluster-correlation was present.

Conclusion: We found that methods accounting for clustering estimate better the true risk and hence allow a more accurate calibration assessment when clustering is present. According to our simulation and case study we recommend using mixed calibration models for clustered validation data.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-02

Maternal Early Warning Scores for detecting deterioration in women during pregnancy: A systematic review

Mae Chester-Jones[1], Gary Collins[1], Stephen Gerry[1], Peter Watkinson[2], Marian Knight[3], Shona Kirtley[1], Fema Er[1], Vidoushee Jogarah[1], Yu Qiao[1], Ruth Tunn[1], Naomi Vides[1]

1 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre,University of Oxford, Oxford, United Kingdom
2 Kadoorie Centre for Critical Care Research and Education, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
3 NIHR Policy Research Unit in Maternal and Neonatal Health and Care, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

Background: In 2024, UK maternal death rates rose to their highest level in 20 years (13.4 per 10,000 pregnancies). Although maternal deaths remain rare, for every woman who dies, 100 women suffer severe maternal morbidity that will often lead to long-term health problems. Modified Obstetric Early Warning Scores (MOEWSs) are a type of prediction model developed specifically for use in pregnant women to detect their risk of deteriorating. They are simple scores calculated from a patients' physiological markers (e.g., blood pressure, heart rate). However, there is no national UK MOEWS and existing scores are inconsistent and highly variable.

Methods: The primary aim of this study (PROSPERO CRD42023396218) was to critically appraise the methodological conduct (design and statistical methods), examine the completeness of reporting, and assess risk of bias of studies describing the development or external validation of one or more MOEWSs. We defined MOEWS as a score used to identify women at risk of deteriorating and experiencing maternal death, ICU admission or a composite of two or more maternal morbidity outcomes. Four databases were searched to identify relevant studies published between the years 2000-2023. Five reviewers extracted data on risk of bias, open science and items based upon the CHARMS checklist (type of study, population, sample size and model building).

Results: After de-duplication and abstract review of 18,874 studies, 80 studies were included for full text review and 22 studies met our eligibility criteria: 10 development studies (including 5 development and external validation studies), and 12 external validation-only studies. Less than half of the MOEWS were developed using statistical modelling (4/10), the remaining scores were based on clinical consensus (6/10). Three of these studies carried out internal validation but none reported the discrimination or calibration, and one study described their assumption of the missing data mechanism. 63% (10/16) of external validation studies validated a score presented as an example in the appendix of an audit of maternal deaths in 2007. The sample size varied between 184 and 54,429 pregnant patients. The description of the dataset characteristics was poorly reported: 13/16 (81%) studies reported the maternal age, 6/16 (38%) reported gestational age and 6/16 (38%) reported ethnicity. Only four studies (25%) assessed discrimination.

Discussion: The review found that there was a lack of evidence based, robustly developed statistical models in this area. Existing models are largely of poor quality and poorly reported. Due to the lack of evidenced based scores, there has been repeated efforts to adapt or validate a score that was not developed using robust methods. Our findings show that known characteristics and indicators of maternal deterioration (ethnicity, age) were rarely considered or included in MOEWS. The assumptions and methods to handle missing data were poorly reported with no studies discussing methods to handle missing data at model deployment. Key performance metrics for internal and external validation (discrimination and calibration) were poorly reported.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-03

Accounting for between-study heterogeneity when developing prediction models using IPD meta-analysis: A comparison of methods and recommendations for practice

Gordon Forbes, Daniel Stahl, Andrew Pickles
Department of Biostatistics and Health Informatics, Kings College London, United Kingdom

Background: Developing prediction models using data from multiple studies in an individual participant data meta-analysis o6ers advantages over using a single dataset. In addition to increased sample sizes, the generalisability of the model may be assessed using the development data by comparing model performance in the different studies. As well as good overall model performance, low between-study heterogeneity in model performance is required to provide convincing evidence that a model will generalise to a new setting. Existing methodological work has focused on overall model performance with less attention paid to between-study heterogeneity. Whilst the difference between marginal predictions (averaging over study effects) and predictions made conditional on study have been highlighted, issues related to making conditional predictions in new, unseen, studies have not been addressed.

Methods: A combination of analytical reasoning, simulation, and analysis of real data is used to compare fixed-effect and random-effect models. By focusing on continuous outcomes, and heterogeneity in intercepts only, we can make use of analytical reasoning to explore differences. We compare both marginal predictions, averaging over intercepts, and conditional predictions, using new data to make predictions conditional on study. Model performance is calculated separately within each study and meta-analysed to give estimates of overall performance and between-study heterogeneity in performance.

Results: Between-study heterogeneity in development data leads to heterogeneity in model performance for marginal predictions. This can be mitigated by using conditional predictions and incorporating new data to estimate intercepts. Using a random intercept model, with empirical Bayes predictions of intercepts, can substantially reduce the amount of new data required. When predictors are highly correlated within clusters, model performance may be improved by using a modelling approach that separates within-study and between-study associations. When there is a small number of studies random intercept models can perform poorly.

Conclusions: If there is a large amount of data available when making new predictions, there is little di6erence between random and fixed intercept models. When there is a small amount of data for intercept estimation random intercept models should be used with empirical Bayes prediction. When cluster-level covariates are available, random intercept models should be favoured over fixed intercept models. For predictors with a high correlation within study, between and within study associations should be separated..

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-04

Prediction performance effect of added-value biomarkers in multivariable heart failure prognostic models: A systematic review

Amy Magona[1,2], Charlotte Smith[1], Nia Roberts, Clare Taylor[1], Kathryn Taylor[1], Maria Vazquez Montes[1]
1 University of Oxford, United Kingdom
2 National Institute for Health and Care Research (NIHR), United Kingdom

Heart failure (HF) is a long-term health condition, estimated to affect more than 64 million people worldwide [1]. The prognostic ability of certain biomarkers has been demonstrated in recent years and new HF prognostic models have been developed with the inclusion of these biomarkers, however this area has not yet been systematically reviewed.

This review aims to evaluate the added-value of the B-type natriuretic peptide (BNP) and N-terminal pro-B-type natriuretic peptide (NT-proBNP) biomarkers by comparing the performance of a base model with and without these biomarkers. We will include studies of the predictive performance of models which aim to predict 1-year mortality in individuals with chronic heart failure. We will search Medline, Embase, SCOPUS and Web of Science. Eligible studies will have either developed, validated or updated HF prognostic models.

Data will be extracted using an extraction form which has been developed by combining the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling studies (CHARMS) checklist [2] and items from the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [3] reporting guideline. Included studies will be assessed for bias and applicability using the prediction model risk of bias assessment tool (PROBAST) [4] which is used to assess the risk of bias and concerns in models which either develop, validate or update multivariable diagnostic or prognostic models.

The predictive performance of the included models will be reported narratively, due to the anticipated heterogeneity between models. We will extract information on the discrimination and calibration of the models, including concordance statistics, area under the curve, observed vs expected outcomes ratio, and calibration plots. The review will also critically appraise the methodology and conduct of the studies.

References: *[1] Savarese, G. et al. (2022) 'Global burden of heart failure: A comprehensive and updated review of Epidemiology', Cardiovascular Research, 118(17), pp. 3272–3287. doi:10.1093/cvr/cvac013. [2] Moons, K.G. et al. (2014) 'Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The charms checklist', PLoS Medicine, 11(10). doi:10.1371/journal.pmed.1001744.] [3] Collins, G.S. et al. (2015) 'Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement', BMJ, 350(jan07 4). doi:10.1136/bmj.g7594. [4] Moons, K.G.M. et al. (2019) 'PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration', Annals of Internal Medicine, 170(1). doi:10.7326/m18-1377.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-05 Prognostic models for heart failure progression: Systematic review and meta-analysis

Giuseppe Occhino[1,2], Alessandro Musa[1], Anita Andreano[3], Paola Rebora[2], Maria Grazia Valsecchi[2], Lucia Bisceglia[1], Antonio Giampiero Russo[3]

1 Strategic Regional Agency for Health and Social Care of Apulia (AReSS Puglia), Bari, Italy
2 Bicocca Center of Bioinformatics, Biostatistics and Bioimaging (B4 Center), School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy
3 Epidemiology Unit, Agency for Health Protection of Milan, Milan, Italy

Introduction: Heart failure (HF) is one of the most prevalent cardiac disorders and, as the population ages, its prevalence is rising. Many prognostic factors of HF progression have been identified. Both statistical and machine learning approaches have been used to provide tools for the prediction of heart failure progression starting from Health Administrative Databases and/or specific medical databases. An estimate of risk of progression for each HF patient would be useful to tailor tertiary prevention, adapt the setting of care and follow-up. In order to provide evidence regarding what model may be better suit different data we performed a systematic review on the available prognostic models for predicting heart failure outcomes.

Methods: Protocol was registered on PROSPERO. The systematic review was conducted using CADIMA software. The bibliographic search was carried out in PubMed and EMBASE biomedical sources. All studies which developed or internally and/or externally validated models on hospitalization or all-cause mortality in adults with HF were included, with the exception of cross-sectional studies. The outcome prediction was within 1 to 5 years. Two authors blindly screened, extracted the data and evaluated the quality of the studies. After screening, data about study characteristics, included population and model predictive capacity were extracted using a piloted form, and the methodological quality of each study was assessed by the PROBAST tool. C–statistic was used to evaluate and meta-analyse model discrimination.

Results: The bibliographic search produced 2,145 results, with 22 additional records identified through references. After title and abstract screening, 80 (3.7%) full-text articles were assessed for eligibility and 24 (1.1%) articles, with 34 different models, were included in the final review. Half (n=17) of the models were internally and/or externally validated. Almost all models were statistical-based approaches (n=32, 94.1%), in particular the Cox proportional hazards model in 90.6% of those. Only 2 (5.9%) models used the decision tree as machine learning-based approach. The most commonly considered outcome was all-cause mortality (61.8%), followed by hospitalization (14.7%). The remaining 23.5% models considered a composite outcome of death or hospitalization. As regards to all-cause mortality, for which studies were more numerous, the 1- and 5-year pooled C–statistics were 0.84 (95% CI: 0.60–0.95) and 0.79 (95% CI: 0.68–0.86), respectively.

Conclusion: Statistical and machine learning predictive models presented good discrimination for all-cause mortality risk in adults with HF, while for the risk of hospitalization few studies were available.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-06

Prediction models for prognostic outcomes in myelodysplastic syndromes and acute myeloid leukaemia: A systematic review

Jyoti Sehjal, Paula Dhiman, Jie Ma, Biruk Tsegaye, Shona Kirtley, Simon Stanworth
University of Oxford, Oxford, United Kingdom

Background: Myeloid malignancies, such as the myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML) are a group of stem cell disorders characterised by abnormal growth of immature blood cells that build up in the bone marrow, leading to low counts of all blood cell types. Many prognostic scoring systems, such as the revised International Prognostic Scoring System (IPSS-R), have been developed to predict overall survival and leukaemia-free survival in patients with MDS/AML. However, less attention is paid to predicting non-survival events, such as bacterial infection and alloimmunisation. There is also little understanding of the types of statistical models used to develop prediction models in MDS/AML.

Methods: We conducted a systematic review to evaluate the reporting and methodological quality of studies developing or validating a prognostic model for patients with MDS/AML. Eligible studies published in MEDLINE (via Ovid), Embase (via Ovid) and Web of Science, from 01/01/2010 to the date the search was run, were included. We assessed reporting quality using the TRIPOD reporting checklist [1] and risk of bias using the PROBAST tool [2]. We also evaluated the reported rationale for model development and their intended use in clinical practice, and any description of algorithmic fairness.

Results: Our findings confirmed that many existing prognostic models for MDS/AML use survival methods, namely Cox regression, to predict survival after disease diagnosis or the start of treatment, such as allogeneic haematopoietic stem cell transplantation. Outcomes other than survival included early relapse in AML, bleeding in AML, and infection after azacitidine treatment. However, many models and their validations were considered as high risk of bias and were poorly reported. The systematic review also highlighted the emergence of machine learning methods and incorporation of genetic predictor (e.g., TP53, FLT3, RUNX1 and ASXL1 mutations) for developing prognostic models for MDS/AML. Finally, very few studies discussed or considered measures to assess algorithmic fairness, such as reporting predictive performance measures across subsets of demographic factors where possible imbalances in prognostic outcomes may occur.

Conclusion: Prediction in MDS/AML predominantly focusses on survival related outcomes; however, the reporting and methodological quality of models predicting survival and non-survival related outcomes needs improvement and need to adhere to current guidelines in prediction model research. Results from this systematic review also emphasise the need for guidelines that evaluate algorithmic bias and predictive fairness for clinical prediction models where differences in model performance are likely to exist across subgroups of the population of interest.

References: *[1] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015; 13(1): 1. [2] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019; 170(1): 51-58.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-07 A sample size calculation for developing a risk prediction model for a binary outcome

Gareth Ambler, Menelaos Pavlou, Rumana Omar
University College London, London, United Kingdom

Background: It is important that risk prediction models are developed using datasets of appropriate size. Over the last few years, several equations have been suggested for calculating the sample size based on performance metrics such as model calibration and predictive accuracy [1].

Methods: In this study, we derive a new sample size equation based on the statistical significance of the individual predictors. We investigate the performance of this equation using simulation, varying factors such as the number of predictors, model strength (c-statistic) and outcome prevalence. We then make comparisons with other approaches including Riley's shrinkage equation [1].

Results: The simulation results suggest that our new sample size equation calculates sample sizes that are generally very similar to that produced by Riley's shrinkage equation. In most scenarios, the resulting risk prediction models had good calibration, discrimination and predictive accuracy.

Conclusion: This new sample size equation can lead to risk prediction models with good calibration, discrimination and predictive accuracy in many scenarios.

Reference: *[1] Riley, Richard D. "Calculating the sample size required for developing a clinical prediction model." BMJ 2020; 368:m441.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-08 How to improve model stability when calculating the sample size and accuracy when estimating a prediction model

Menelaos Pavlou, Rumana Z. Omar, Gareth Ambler
University College London, London, United Kingdom

Background: It is important to use an appropriate sample size when developing risk prediction models. Recent sample size recommendations to limit model overfitting (among other metrics) can contribute to the development of more reliable risk models. They aim to calculate the sample size based on anticipated model characteristics (number of predictors, c-statistic, outcome prevalence) to ensure that the calibration slope (CS), as a measure of model overfitting, will on average meet a target value (typically 0.9), when the model is fitted using Maximum Likelihood Estimation (MLE). An important aspect of model performance that is not controlled by the existing sample size calculations is model stability (i.e. variability in CS).

Methods: In this study we:
a) use simulations to investigate model stability as a function of model features (number of predictor variables, outcome prevalence and model strength) at the recommended sample sizes when the outcome is binary. Model stability can be quantified, for example, by the variability in the CS over simulations, the root mean square distance of the CS from the ideal value of 1 (RMSD), or the probability of obtaining a sufficiently good CS.

b) investigate an adaptation of existing sample size calculations to control model stability, in addition to ensuring good performance on average.

c) explore whether applying simple uniform shrinkage (using bootstrapping) following an MLE fit (with the recommended sample size), may lead to increased model accuracy/stability.

Results: When adhering to the recommended sample sizes, the variability in the CS increased substantially as the number of predictors decreased. As a result, even if the target CS is met on average, the danger of obtaining a poorly calibrated model can be high, especially when the number of predictors is small. A sample size calculation which ensures that the probability of obtaining a well-calibrated model is sufficiently high can improve model stability. This calculation can be performed either by simulation or analytically. Post-estimation uniform shrinkage (using bootstrapping) led to a much lower RMSD, and a higher probability of obtaining a well-calibrated model compared to MLE, unless the number of predictors was very small (≤5).

Conclusion: Sample size calculations for the development of prediction models should aim to control model stability, in addition to performance on average. Application of shrinkage, at the recommended sizes, may be beneficial in obtaining a more accurate model.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-09 Selection and verification bias adjustments for clinical risk model validation

Donna Ankerst
Technical University of Munich, Munich, Germany

Background: Clinical risk prediction tools are frequently developed from large studies in order to improve public health monitoring, doctor-patient decision-making and clinical trial management. Trends in online posting of tools facilitate rapid external validation across heterogeneous patient populations. Variation of external validation performance, however, can lead to confusion over utility and generalizability of tools.

Methods: This talk formalizes the concepts of reproducibility and transportability of risk models by defining selection and verification bias in study cohorts. When individual level information from both the training and validation cohorts is available, weighted versions of the standard calibration and discrimination metrics to adjust for selection and verification bias can be calculated.

Results: Methods are illustrated through the development and validation of models for five-year prostate cancer risk using two large North American prostate cancer prevention trial cohorts.

Conclusion: Adjustment for selection and verification bias yielded reverse validation conclusions from the unadjusted validation analyses, showing the importance of providing individual-level training data for online risk tools and bias adjustments in external validation studies.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-10

Integrating genomic markers for enhanced prediction: Uncertainty assessment of clinical usefulness in early breast cancer

Frank Doornkamp, Ewout Steyerberg
LUMC, Netherlands

Background: Genomic markers may improve targeting of chemotherapy to patients at increased risk beyond traditional clinical factors in early-stage breast cancer. We aimed to evaluate the uncertainty in clinical usefulness of two multi-genomic marker tests, MammaPrint and OncotypeDX, in guiding treatment decisions.

Methods: The MINDACT and TAILORx trials enrolled 6693 and 10253 patients with earlystage breast cancer respectively. These trials prospective validated the MammaPrint (a binary test classifying patients as high vs low risk) and OncotypeDX (a continuous score) respectively. From these trials, we constructed two datasets with 1 Million virtual patients to explore the added clinical value of each test. A risk-based approach is advised in this disease, which implies that chemotherapy should be initiated if the predicted benefit of chemotherapy is a reduction of the 10-year risk of distant metastasis by at least 5%. For every patient, we assessed if treatment was indicated based on clinical characteristics alone and after including either one of the genomic tests. The clinical usefulness of genomic markers was determined by assessing the improvement in treatment allocation, quantified as a weighted sum between correctly and incorrectly treated patients (i.e. Net Benefit) [1]. Uncertainty in the improved Net Benefit was assessed by using subsampling equal to the corresponding trial sample size and evaluation of the Net Benefit for every subsample (200 repetitions). This subsampling allowed the construction of 95% confidence intervals around point estimates.

Results: The reference model, including only clinical information, would result in a positive balance between preventing distant metastasis and overtreatment in both trial settings (5.6 [-1.2 – 11.6] net metastasis prevented per 1000 patients in MINDACT, 3.2 [- 3.2 – 9.2] per 1000 in TAILORx). Multi-genomic marker tests improved patient selection in both trial settings to a different extent: +3.1 [-2.4 – 7.1] vs +1.3 [-2.8 – 5.5] in MINDACT, +2.6 [-2.6 – 4] vs +1.2 [-2.4 – 5.3] in TAILORx per 1000 OncotypeDX vs MammaPrint tests.

Conclusion: - Multi-genomic marker tests improve identifying patients who benefit from chemotherapy in early-stage breast cancer. The large uncertainty motivates further evaluation of their clinical usefulness and cost-effectiveness.

Reference: *[1] Vickers, A. J., Kattan, M. W., & Daniel, S. (2007). Method for evaluating prediction models that apply the results of randomized trials to individual patients. Trials, 8, 14. https://doi.org/10.1186/1745-6215-8-14*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-11

Statistical pitfalls and errors in external validation studies of clinical prediction models: A case study with the SAPS II and SAPS 3 prognostic models for critically ill patients

Evangelos Kritsotakis[1], Stamatios Kokkinakis[2], Ioannis Xanthakis[3], Augusto Morgia[4], Maria Samaritaki[5], Konstantinos Lasithiotakis[2]

1 Laboratory of Biostatistics, School of Medicine, University of Crete, Heraklion, Greece
2 Department of General Surgery, School of Medicine, University of Crete, Heraklion, Greece
3 Master of Biostatistics Program, Medical School and Department of Mathematics, National Kapodistrian University of Athens, Athens, Greece
4 Department of Public Health and Infectious Diseases, Specialization School of Health Statistics and Biometrics, Sapienza Università Di Roma, Roma, Italy
5 Department of Anaesthesiology, University Hospital of Heraklion, Heraklion, Greece

Introduction: Successful application of a clinical prediction model requires external validation by applying the model as originally specified to different patient populations or settings and then quantifying the accuracy of the predictions made. Empirical accounts examining how researchers apply statistical methods for this purpose are lacking. Herein we report statistical pitfalls and errors identified from a systematic review of the prognostic performance of the Simplified Acute Physiology Score (SAPS) II and SAPS 3 in intensive care units (ICUs). The SAPS models rely on ICU admission data to estimate the probability of subsequent in-hospital death with a logistic regression equation.

Methods: PubMed and Scopus were searched for studies externally validating the discrimination and calibration performances of SAPS II or 3 in ICU patients, since 2003. Forward citation searches were performed and previous systematic reviews of SAPS II/3 were surveyed for eligible studies. Data were extracted using the CHARMS checklist. Risk of bias (RoB) was assessed with the PROBAST tool. PROSPERO CRD42023446721.

Results: In all, 89 external validation reports involving 495,443 patients in 1,157 ICUs in 33 countries were assessed. Most studies were performed in Europe (n=38; 43%), Asia (n=26; 29%), or South America (n=13; 15%). Most involved prospective data collection (n=57; 64.0%), sampled consecutive ICU admissions (n=84; 94%), and were designed either as cohort (n=69; 78%) or registry-based (n=20; 22%) studies. In the PROBAST analysis domain, 73 reports (82%) were classified as high RoB and 8 (9%) were unclear RoB, whereas only 8 reports (9%) were low RoB. Most concerns were raised due to inappropriate handling of missing data by either unjustified complete case analysis (n=43; 48%) or single imputation by normal/reference values (n=12; 13%). Other concerns were inadequate evaluation of model performance measures (n=30; 34%), low numbers of outcome events (n=29; 33%), and not including all enrolled patients in the analysis (n=20; 23%). Most studies did not report a formal sample size calculation (n=85; 96%). Calibration was assessed solely with a statistical test in 13 cases (15%). Discrimination was examined with the AUC in all studies, but 20 (22%) calculated a classification measure after dichotomizing SAPS with some cut-off threshold. Decision curve analysis was performed in only 3 studies (3.4%).

Conclusion: External validation studies of prognostic models are viewed as necessary and are pursued by the medical research community. However, this investigation reveals an urgent need for improved

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-12 Examining model stability across modelling approaches to predicting poor functioning for individuals at-risk of psychosis

Yanakan Logeswaran[1], Dominic Oliver[1], Matthew Kempton[1], Mark van der Gaag[2], Lieuwe De Haan[3], Barnaby Nelson[4], Anita Riecher-Rössler[5], Rodrigo Bressan[6], Neus Barrantes Vidal[7], Marie-Odile Krebs[8], Merete Nordentoft[9], Stephan Ruhrmann[10], Gabriele Sachs[11], Bart Rutten[12], Jim van Os[12], Lucia Valmaggia[4], Philip McGuire[13], Paolo Fusar-Poli[1], Daniel Stahl[1], EU-GEI High Risk Study[14]

1 King's College London, London, United Kingdom
2 VU University Amsterdam, Amsterdam, Netherlands
3 AMC Amsterdam, Amsterdam, Netherlands
4 Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, Australia
5 University of Basel, Basel, Switzerland
6 UNIFESP, São Paulo, Brazil
7 Universitat Autònoma de Barcelona, Barcelona, Spain
8 University Paris Descartes, Paris, France
9 University of Copenhagen, Copenhagen, Denmark
10 University of Cologne, Cologne, Germany
11 Medical University of Vienna, Vienna, Austria
12 University of Maastricht, Maastricht, Netherlands
13 University of Oxford, Oxford, United Kingdom
14 EU-GEI High-Risk Study Group

Background: Primary indicated prevention in individuals at Clinical High Risk for psychosis (CHR-P) has largely focussed on transition to psychosis. However, many individuals do not develop psychosis but still experience substantial functional decline. Given the associated personal and societal burden, clinical prediction models have been developed to inform preventive strategies and have observed good prognostic accuracy. However, recent reviews have highlighted that many model development studies are at high risk of bias, and that this may hamper their external generalisability and utility. Given the use of small development datasets and inadequate internal validation, there are concerns that the individual model predictions of these models may be unstable. Thus, this study aims to compare model performance and instability across modelling approaches for predicting poor functioning in CHR-P individuals.

Methods: CHR-P individuals were recruited as part of the EU-GEI study, a prospective observational study. The outcome was determined by the GAF disability/impairment scale at the last follow-up visit. We will develop each model utilising three cumulative blocks of five a priori chosen predictors. Logistic regression, regularized logistic regression, and random forest models will be developed using a 10-fold cross-validation framework to predict poor functioning (GAF<61). We will assess the models' performance using internal validation, indexed by discrimination (area under the curve, balanced accuracy, and Brier score) and calibration (calibration slope and Brier score). Using bootstrapping, we will examine model instability of individual predictions through (i) prediction instability plots, (ii) calibration instability plots, and (iii) mean absolute prediction error (MAPE) instability plots.

Results: 220 CHR-P individuals (mean[SD] age at study entry = 22.7[4.8] years; 119[54%] male; 161[73%] white ethnicity) were included. 108(49%) had poor functioning at last follow-up visit (median[IQR] years from baseline to last visit = 2.0[1.1-2.2]). Regarding CHR-P criteria, 12 (6%) met criteria for Brief Limited Intermittent Psychotic Symptoms, 179 (86%) met Attenuated Psychotic Symptoms criteria, and 17 (8%) met criteria for genetic risk vulnerability. Regarding comorbid psychopathology at baseline, 66/214 (31%) had major depression, 109/212 (51%) an anxiety disorder, and 26/208 (13%) a personality disorder. Further, 63/194 (32%) had been prescribed antidepressants, 23/194 (12%) anxiolytics, and 17/190 (9%) antipsychotics.

Discussion: This study may emphasise the importance of reporting model stability in this field and perhaps emphasise the importance of harmonisation of measures across consortia to provide larger datasets for more stable, better performing clinical prediction models that can improve care for CHR-P individuals.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-13 Development and validation of a novel longitudinal prognostic model for depressive relapse using passively measured biomarkers

Ewan Carr[1], Faith Matcham[2], Nicholas Meyer[3,4], Daniel Stahl[1], Matthew Hotopf[5,6]

1 Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom
2 School of Psychology, University of Sussex, Falmer, United Kingdom
3 Insomnia and Behavioural Sleep Medicine Service, University College London NHS Foundation Trust, London, United Kingdom
4 Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
5 Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
6 South London and Maudsley NHS Foundation Trust, London, United Kingdom

Background: Major depressive disorder (MDD) is a heterogeneous condition characterised by highly variable course trajectories with periods of remission followed by relapse[1]. Prognostic models for depression enable timely and targeted interventions, but typical markers rely on patient self-reports, which are time-consuming and often incomplete. Mobile health technologies, including wearable and smartphone devices, enable convenient passive collection of objective physical markers. Sleep or circadian function are leading candidate markers for early identification of depressive relapse[2,3]. While existing models demonstrate the potential of sleep parameters for depression[4], they rely on small, cross-sectional samples[5,6] and do not incorporate recent innovations in longitudinal prediction modelling. This study aimed to develop and validate a novel longitudinal, multivariable, prognostic model to detect depressive relapse and symptom severity in people with recurrent MDD.

Methods: Data RADAR-MDD7 was an international cohort study of people with recurrent MDD. Participants (n=623) at three sites (UK, the Netherlands, and Spain) wore a Fitbit device and provided regular outcome assessments for two years. Measures Outcomes were measured repeatedly every three months: depressive relapse and depressive symptom severity (IDS-SR). Sleep features were measured continuously and summarised over the three months preceding each outcome assessment (e.g., median, variance, slope). Candidate features included total sleep time, sleep efficiency, sleep fragmentation, onset latency, midpoint, and social jet lag. We also considered common demographic and clinical factors, including age, gender, education, partnership, and medication and alcohol use. Statistical analyses We developed longitudinal prediction models using three approaches: Regularised Generalised Linear Mixed Models (glmmLasso8), Binary Mixed Model forest (BiMM9), and Generalized Mixed Models with Likelihood-Based Boosting (GMMBoost10). We also considered a random forest model that ignored clustering. Models included non-linear transformations and relevant interactions. Each algorithm was evaluated by combining internal-external cross-validation (leave-one-site-out) with cluster bootstrapping11 (clustered on participant). Models were assessed in terms of performance (R2), discrimination (c-statistic), calibration (calibration-in-the-large, slope), and model stability (MAPE12). Results Participants had a median age of 49 years, and 24% were male. For depression relapse, there were 1260 assessments and 62 relapse events (5%). For depression severity, there were 3743 assessments. Our results (forthcoming) will compare optimism-corrected model performance, calibration, and stability across the four algorithms.

Conclusion: Accurate prediction of symptom fluctuation in recurrent MDD could enable proactive, personalised interventions and reduced relapse rates. Our study is the first to develop prognostic models for recurrent depression using longitudinal mobile health assessments.

References: *[1] Verduijn J, Verhoeven JE, Milaneschi Y, et al. Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule. BMC Medicine 2017; 15: 215. [2] Pandi-Perumal SR, Monti JM, Burman D, et al. Clarifying the role of sleep in depression: A narrative review. Psychiatry Research 2020; 291: 113239. [3] Daghlas I, Lane JM, Saxena R, et al. Genetically Proxied Diurnal Preference, Sleep Timing, and Risk of Major Depressive Disorder. JAMA Psychiatry 2021; 78: 903–910. [4] Otte Andersen T, Skovlund Dissing A, Rosenbek Severinsen E, et al. Predicting stress and depressive symptoms using high-resolution smartphone data and sleep behavior in Danish adults. Sleep 2022; 45: zsac067. [5] Chikersal P, Doryab A, Tumminia M, et al. Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection. ACM Trans Comput-Hum Interact 2021; 28: 3:1-3:41. [6] Lyall LM, Sangha N, Zhu X, et al. Subjective and objective sleep and circadian parameters as predictors of depression-related outcomes: A machine learning approach in UK Biobank. Journal of Affective Disorders 2023; 335: 83–94. [7] Matcham F, Barattieri di San Pietro C, Bulgari V, et al. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019; 19: 72. [8] Groll A, Tutz G. Variable selection for generalized linear mixed models by L1-penalized estimation. Stat Comput 2014; 24: 137–154. [9] Speiser JL. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. J Biomed Inform 2021; 117: 103763. [10] Tutz G, Groll A. Generalized Linear Mixed Models Based on Boosting. In: Kneib T, Tutz G (eds) Statistical Modelling and Regression Structures. Heidelberg: Physica-Verlag HD, pp. 197–215. [11] Bouwmeester W, Moons KGM, Kappen TH, et al. Internal Validation of Risk Models in Clustered Data: A Comparison of Bootstrap Schemes. American Journal of Epidemiology 2013; 177: 1209–1217. [12] Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. Biometrical Journal 2023; n/a: 2200302.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-12

Development and validation of treatment effect functions in observational studies: Application to predictive analysis of amiodarone in out-of-hospital cardiac arrest patients

Ryo Emoto[1], Mitsuaki Nishikimi[2], Shigeyuki Matsui[1]
1 Nagoya Graduate School of Medicine, Nagoya, Japan
2 Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

Background: Although amiodarone is widely used for resuscitating out-of-hospital cardiac arrest (OHCA) patients with shockable rhythm, several recent randomized clinical trials have reported different results regarding the effects of amiodarone. One possible explanation for this discrepancy may be the heterogeneity of the therapeutic effect of amiodarone in these patients.

Methods: We analysed data from the Japanese nationwide OHCA registry of patients with shockable rhythm at hospital arrival. The primary endpoint was a favorable neurological outcome at 30 days. We applied a predictive analysis with direct modelling of treatment-covariate interactions. To address confounding in observational data, we utilized an inversed-propensity weighting based on a set of possible confounding factors to obtain a treatment effect function to inform treatment selection for individual patients, mimicking a treatment effect function that would be obtained from a randomised clinical trial. As an internal validation, we devised a cross-validation based on the propensity score and adopted a permutation procedure to test for the variation in treatment efficacy as well as treatment effects in subgroups indicated by the derived treatment effect function.

Results: Out of 68,111 OHCA cases, 2,333 were eligible for our analysis. Our model identified several factors increasing the treatment effect of amiodarone, including higher age, longer interval between the call to the emergency medical services and hospital arrival, absence of a witness, no defibrillation prior to hospital arrival, hypothermia at hospital arrival, and pre-hospital epinephrine administration. The cross-validation analysis revealed some variability in the treatment effect of amiodarone across different subgroups.

Conclusions: Our method to develop and validate the treatment effect function based on the propensity score can be useful in the predictive marker analysis using observational data. Issues in more comprehensive prediction analysis for OHCA patients include incorporation of the effects of other drugs and the timing of treatment initiation.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-15

A comparison of regression models for static and dynamic prediction in electronic health records in the absence of censoring

Shan Gao[1], Elena Albu[1], Hein Putter[2], Pieter Stijnen[3], Frank Rademakers[1], Veerle Cossey[1,3], Christel Janssens[3], Yves Debaveye[3], Ben Van Calster[1,2], Laure Wynants[1,4]
1 KU Leuven, Belgium
2 Leiden University Medical Center, Netherlands
3 University Hospitals Leuven, Belgium
4 Maastricht University, Netherlands

Introduction: Electronic health records (EHR) data used for predicting events during hospitalization often have no censoring in the sense that hospitals register information until discharge or death. We aimed to compare different logistic and survival regression modeling approaches in both baseline and dynamic context to estimate the risk of central line-associated bloodstream infection (CLABSI) in EHR data without censoring but with competing events.

Methods: We used data from the EHR of the University Hospitals Leuven (Belgium), comprising 30,862 catheter episodes from 2012 and 2013 to predict 7-day risk of CLABSI for hospitalized patients with catheters in both static and dynamic way. Model performance was assessed using 100 random 2:1 train-test splits, with missing data imputed via an adapted missForest algorithm 1. Static models included logistic regression, multinomial logistic regression, Cox regression, cause-specific hazard regression, and Fine-Gray regression. Survival models were fitted with or without administrative censoring at 7 days to reduce the impact of eventual non-proportional hazards. Dynamic models included the same type of models using a landmark supermodel approach, and regularized multi-task learning2,3. Models were fitted with administrative censoring at prediction interval of 7 days to make the models robust against violations of proportional hazards assumption. We assessed discrimination, calibration, and overall performance, treating outcomes as binary due to the absence of censoring.

Results: Among static models, the cause-specific model without administrative censoring at 7 days had the highest AUC (mean 0.721), while the multinomial logistic and cause-specific model with administrative censoring had the best calibration. Several dynamic landmark supermodels reached peak dynamic AUCs between 0.741 and 0.747. Calibration deteriorated over landmark time, however Fine-Gray and cause-specific landmark supermodels maintained stable observed over expected (O/E) ratios. Both for static and dynamic models, Cox models had the poorest discrimination and calibration.

Conclusions: In this comparative study, categorical and time-to-event outcome definitions exhibited similar model performance in the static and dynamic setting, except for Cox models. Ignoring competing risks caused problems for risk prediction in the time-to-event framework (Cox), but not in the categorical framework (logistic regression).

References: *[1] Albu, Elena. 2023. missForestPredict: Missing Value Imputation Using Random Forest for Prediction Settings. https://github.com/sibipx/missForestPredict. [2] Houwelingen, Hans, and Hein Putter. Dynamic Prediction In Clinical Survival Analysis. Boca Raton: CRC Press, 2012. [3] Cao, H., J. Zhou, and E. Schwarz. 2018. "RMTL: An r Library for Multi-Task Learning." Journal Article. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty831.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-16 Diagnostic tool to assess projected occupation probabilities in illness-death model

Tommi Härkänen, Mrs. Heli Tapanainen, Laura Sares-Jäske, Laura Paalanen
Finnish Institute for Health and Welfare, Helsinki, Finland

Background/Introduction: Accurate predictions are vital in decision making. Predicting the risk of, for example, ischemic heart disease (IHD) allows for early lifestyle changes and medication, potentially preventing IHD. Predicting health outcomes in a population, future healthcare needs can be identified helping to allocate healthcare resources more effectively. External validation of model assumptions with respect to data collected later than the data used for parameter estimation can provide important information on temporal changes in the associations of risk factors and health outcomes, which can influence predictions at least in population subgroups.

Methods: We apply microsimulation methods to predict occupation probabilities of an illness-death Poisson regression model[1] with multiple timescales, accounting for parameter and prediction uncertainties. We perform external validation using older population surveys for parameter estimation and latest population surveys to compare the predictions and observed outcomes. We predict the occupation probabilities using parametric bootstrap and simulation of event times.[1] We assess the predictions using the multinomial regression model diagnostics based on randomized quantile residuals (RQR).[2] We compare these residuals with the standard normal distribution using normality tests, heteroscedastic linear models and regression trees to assess and identify problematic covariates. As the RQR have not been presented for nominal outcomes, we perform the diagnostics for all 24 permutations of the four possible states and use Bonferroni correction to control the familywise type I error rate. We utilize the Health 2000 (n=5975) and FINRISK 2007 (n=4874) surveys for parameter estimation, and the FINRISK 2012 (n=4812) and FinHealth 2017 (n=5123) surveys for the validation. We used basic demographic variables in addition to the most important risk factors of IHD. The time-to-event data on deaths and IHD incidences are obtained from registers linked with the survey datasets. The follow-up ended in 2011, 2014, 2021 and 2021 for these surveys, respectively.

Results: Most discrepancies were found with respect to age. Among the oldest individuals, the predicted IHD incidence was higher than the observed incidence indicating a missing interaction of calendar time and age in the model.

Conclusions: Our approach provides a diagnostic tool to validate a multistate model against another dataset to detect missing interactions or other inadequate model assumptions.

References: *[1] Carstensen, Plummer (2011). Using Lexis Objects for Multi-State Models in R. J. Stat. Softw., 38(6), 1-18. URL https://www.jstatsoft.org/v38/i06/. [2] Dunn and Smyth, Randomized quantile residuals, J. Comput. Graph. Stat. 5 (1996), 1–10.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-17

Quantifying the impact of underdiagnosis in outcome data when developing and validating a clinical prediction model: A simulation study

Joyce (Yun-Ting) Huang, Matthew Sperrin, David A. Jenkins
University of Manchester, Manchester, United Kingdom

Background: Clinical prediction models (CPMs) are valuable in aiding healthcare decision-making. However, the data used to develop and validate CPMs are susceptible to measurement errors and missing information. When such errors affect the outcome of a CPM, such as underdiagnosis in conditions like diabetes, this may have a serious effect on the accuracy of the CPM. This study aims to investigate the impact of varying degrees of underdiagnosis on the predictive performance of CPMs through a simulation study.

Methods: A simulation study was conducted to investigate a prognostic model for a binary outcome. Development and validation datasets included 100,000 individuals and 3 predictors. The prevalence of the binary outcome was varied at 1%, 5%, 10%, 30%, and 50%. Different scenarios of underdiagnosis, 1%, 5%, 10%, 30%, and 50%, were simulated in the binary outcome of the development data, while the validation data remained free of underdiagnosis. One of the predictors was binary and assumed to influence underdiagnosis, to represent a sensitive attribute impacting the outcome and diagnosis. In the presence of this predictor, individuals were assumed to be less or equally likely to experience underdiagnosis. Predictive performance was evaluated in the validation dataset by calculating calibration intercept, calibration slope, and the area under the ROC Curve (AUC).

Results: When outcome prevalence exceeded 10%, calibration intercepts exhibited significant miscalibration for all degrees of underdiagnosis. Conversely, when the prevalence was below 10%, underdiagnosis of 5% did not significantly impact model performance. In the scenario where the outcome prevalence was 10%, a 30% underdiagnosis proportion yielded a calibration intercept of 0.40 (95% CI:0.35,0.45) and a calibration slope of 1.04 (95% CI: 0.99,1.10). Reducing underdiagnosis to 5% resulted in a calibration intercept of 0.06 (95% CI: 0.01,0.10) and a calibration slope of 1.01 (95% CI: 0.96,1.06). The calibration slope only exhibited miscalibration when the outcome prevalence was 30% or 50% and the underdiagnosis was ≥10%. Despite these variations, no differences were observed in the AUC across different degrees of underdiagnosis or outcome prevalence. When the predictor impacted underdiagnosis, we observed better predictive performance for calibration and a small decrease in AUC, despite unchanged significance.

Conclusion: Our findings suggest all levels of underdiagnosis are concerning when CPMs come to clinical utility, except for a low prevalence disease with a low proportion of underdiagnosis. Care needs to be taken to ensure this measurement error (underdiagnosis) is avoided in the data used to develop CPMs. Funding This work was funded by an Innovate UK GM levelling up grant. The views expressed in this publication are those of the authors and not necessarily those of the UK Research and Innovation.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-18 Prognosis value of immunocyte ratios in neoadjuvant chemotherapy-treated breast cancer

Seyed Amirhossein Jalali[1,2,3], David Miresse[4], Matthew Fahey[4], Niamh Mhaonaigh[4], Andrew McGuire[4], Emer Bourke[4,5], James Brown[2,3,5,6]
1 School of Medicine, University of Limerick, Limerick, Ireland
2 Health Research Institute, University of Limerick, Limerick, Ireland
3 Limerick Digital Cancer Research Centre, University of Limerick, Limerick, Ireland
4 Lambe Institute for Translational Research, University of Galway, Galway, Ireland
5 Centre for Chromosome Biology, University of Galway, Galway, Ireland
6 Bernal Institute, University of Limerick, Limerick, Ireland

Background: Breast cancer represents a heterogeneous condition in which the interaction between the host immune response and primary oncogenic events can impact disease progression. NLR (neutrophil-to-lymphocyte ratio) has been shown to be prognostic in a variety of cancers, including breast cancer, however, the role of other immunocyte ratios remains to be fully explored. We studied the prognostic value of five key blood marker ratios (i.e. NLR, LMR, NWR, LWR, MWR and PLR) for overall survival (OS) and disease-free survival (DFS).

Methods: A neoadjuvant chemotherapy-treated cohort (2006–2017) of histologically confirmed breast cancer patients (from Galway University Hospital, Ireland) is analysed to investigate the relationship between each blood marker (immunocyte) ratio and 3-year OS and 3-year DFS. Descriptive statistics were performed to summarise the variation in each marker ratio by cancer subtype. Optimal (subtype-specific) thresholds were chosen for each marker ratio to maximize prognosis differences between low- and high-blood marker ratios. Receiver Operating Characteristic (ROC) cut-off points for each blood marker ratio were chosen to ensure at least 80% sensitivity for 3-year OS and 3-year DFS, and the corresponding specificity was calculated.

Results: Data consists of 353 neoadjuvant chemotherapy-treated patients including patients with Luminal A (n = 108), Luminal B (n = 122), HER2-positive (n = 41), Triple negative/basal (n = 43) followed for a period of at least 3 years. LMR (5.29 cutoff point) and MWR (0.06 cutoff point) were associated with Luminal A OS (p = 0.004 and p = 0.022) and DFS (p = 0.004 and p = 0.022), and Luminal B OS (p = 0.027 and p = 0.008) and DFS (p = 0.005 and p = 0.007). NLR (1.79 cutoff point) and LWR (0.30 cutoff point) were associated with HER2-positive OS (p = 0.013 and p = 0.043). NLR (1.79 cutoff point) and NWR (0.62 cutoff point) were associated with DFS (p = 0.035 and p = 0.021). No significant association was observed between any immunocyte ratio in the triplenegative cohort. Conclusions: The prognostic value of immunocyte ratios in neoadjuvant chemotherapy-treated breast cancer is investigated, and the subtype-specific prognostic value of immunocyte ratios is discussed. The results show significant variation in the application of specific immunocyte ratios by breast cancer subtype, indicating the need for a large multi-national study to validate the use of subtype-specific immunocyte ratios as clinically relevant prognostic markers.

References: *[1] Jalali, A., Miresse, D., Fahey, M. R., Ni Mhaonaigh, N., McGuire, A., Bourke, E.,... & Brown, J. A. (2022). Peripheral Blood Cell Ratios as Prognostic Indicators in a Neoadjuvant Chemotherapy-Treated Breast Cancer Cohort. Current Oncology, 29(10), 7512-7523.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-19 Use of statistical process control to monitor calibration-in-the-large of a clinical prediction model

David Jenkins[1], Glen Martin[1], Mamas Mamas[2], Niels Peek[1], Matthew Sperrin[1]
1 University of Manchester, United Kingdom
2 Keele University, United Kingdom

Background: Clinical prediction models (CPMs) are useful tools used across healthcare but a CPMs coefficients are often time invariant post development. It is often unclear when in time a model should be updated, or no longer used. As such, updating is currently done at arbitrary time points. Therefore, this study aimed to explore the use statistical process control (SPC) methods as a way of identifying when a CPM should be updated.

Methods: We describe the use of statistical process control (SPC) to continuously monitor a CPMs calibration-in-the-large, and generate an alert when the calibration-in-the-large reaches some predefined threshold (control limit). We also extend the method by proposing a burn in period and a resenting procedure for the limits when the model is well calibrated to control error. A simulation study was performed based on an existing CPM for 30-day mortality after percutaneous coronary intervention. We generated simulated outcomes under varying degrees of miscalibration-in-the-large. SPC was then used to monitor the model performance, and repeated 1000 times, for each miscalibration scenario. Time of an alert, generated from using 3 and 4 standard deviations (SDs) as the control limits, was recorded for each iteration. The number of individuals that would have incorrectly been classified using a risk threshold was recorded.

Results: When the intercept of the data generating model was miscalibrated by 0.2, resulting in an average systematic over prediction 1.2-fold the true risk, an alert was triggered on average after 7510 time units (95% CI: 294, 27136) and 16022 units (95% CI: 1863, 42160) for the 3SD and 4SD control limits, respectively. When the data generating model was miscalibrated-in-the-large by 2, an average overestimation of risk 4.7-fold an individual's true risk, the median and 95% quantile for the 3 and 4SD control limits were 246 (95% CI: 175, 320) and 246 (95% CI: 175, 321), respectively. When the there was no miscalibration, 10% of the iterations resulted in an incorrect alert for the 3SD control limit, compared to 1.2% using the 4SD control limit. For all risk thresholds, the 4SD and 3SD control limit incorrectly classified up to 17 persons and 3 persons, respectively. Conversely, no control limit resulted in up to 6000 misclassifications, depending on degree of miscalibration.

Conclusion: This study shows how SPC can be used as a way of continually monitoring the calibration-in-the-large of a CPM, to suggest when a model should be updated.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-20

Predicting individualised treatment effects: A comparison of different modelling approaches and performance metrics

Carolien Maas[1], David M. Kent[2], David van Klaveren[1]
1 Erasmus MC, Rotterdam, Netherlands
2 Tufts Medical Center, Boston, United States

Background: Measuring the performance of models that predict individualized treatment effects is challenging because the outcomes of two alternative treatments are inherently unobservable in one patient. We previously proposed metrics that defined observed treatment effect as the difference between the outcomes in pairs of matched patients with different treatment assignments. We aimed to assess the usefulness of these metrics when comparing the performance of different models designed to predict individualized treatment effects in randomized clinical trials (RCTs).

Method: We used publicly available RCT databases in diverse medical fields (cardiovascular disease, diabetes, and prostatic hyperplasia) to analyse seventeen unique treatment comparisons that had significant overall treatment effects. Individualized treatment effects for each comparison were predicted with a risk model (including a constant relative treatment effect) and a penalized effect model (including all interactions between risk factors and treatment) fitted with LASSO regression. Discrimination, calibration, and overall performance were evaluated at apparent and 5-fold cross-validation, using the previously proposed metrics: C-for-benefit, E-for-benefit, and Brier-for-benefit, with confidence intervals computed using 100 bootstrap samples. We performed sensitivity analyses to assess the influence of the matching strategy by changing the distance measure from Mahalanobis to Robust Mahalanobis, and the matching order from randomly to matching the closest pairs first.

Results: The sample sizes of RCTs ranged from 715 to 10,251, with the outcome rate ranging from 5.3% to 35.9%, and the number of risk factors ranging from 4 to 16. At apparent validation, effect models often outperformed risk models numerically based on the C-for-benefit (15 out of 17 treatment comparisons) and Brier-for-benefit (9 out of 17 treatment comparisons). However, at 5-fold cross-validation, the risk modelling approach often performed numerically best based on the C-for-benefit (13 out of 17 treatment comparisons) and Brier-for-benefit (16 out of 17 treatment comparisons). These findings were consistent across all matching strategies, but model performance did not differ significantly. Calibration measures were highly sensitive to the matching strategy. For example, the median difference in E50-for-benefit across the RCTs was 0.01 when we compared random order matching using Mahalanobis versus robust Mahalanobis distance.

Conclusion: Despite improvements in apparent validation, effect modelling including interactions do not add substantially to risk modelling in our sample of RCTs at cross-validation. The C-for-benefit and Brier-for-benefit give consistent numerical results across matching strategies. However, the E-for-benefit is sensitive to the matching strategy and further research is needed to propose alternative calibration measures.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-21

Predicting study duration based on dynamic predictions using joint models in clinical trials with a time-to-event endpoint

Ryunosuke Machida[1], Kentaro Sakamaki[2], Takashi Sozu[3]
1 Biostatistics Division, Center for Research Administration and Support, National Cancer Center, Japan
2 Faculty of Health Data Science, Juntendo University, Japan
3 Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Japan

Background: In event-driven clinical trials with time-to-event endpoints such as overall and progression-free survivals, the primary analysis is conducted when the required number of events is observed. Thus, the study duration is uncertain, and its evaluation is essential to manage the costs, personnel, and labour of planned trials in practice. Several methods for predicting study duration have been developed, considering the enrolment scheme and distributions of time-to-event outcome and censoring. In such cases, the precision of predicting the study duration would be enhanced if a longitudinally measured covariate associated with the time-to-event outcome was available after starting the trial and incorporated into the prediction method for the study duration. Methods: We propose a method for predicting the study duration based on dynamic predictions using joint models for the time-to-event outcome and longitudinally measured covariates. Simulations were conducted to evaluate the operating characteristics.

Results: We predicted the study duration, assuming prostate cancer clinical trials in which the prostate-specific antigen level (continuous variable) of each participant was longitudinally measured. Participants at a certain time point were classified into three conditions: (a) participant with an event, (b) participant without an event, and (c) potential participant not enrolled, if lost to follow-up was not assumed. The proposed method comprises of four steps. Step 1: Estimate the parameters of the enrolment scheme, distribution of baseline covariates, and joint model using the interim data of participants (a) and (b). Step 2: Generate and impute enrolment dates and baseline covariates of potential participants (c). Step 3: Generate time-to-event data based on the framework of dynamic predictions using joint models for participants (b) and (c), conditioning the accumulated data until the current time point [1]. Step 4: Calculate the duration until the required number of events is observed (study duration). By repeating Steps 2—4, the summary statistics of the realised values for the study duration can be calculated, enabling the assessment of its uncertainty.

Conclusion: We demonstrate that the proposed method increases the precision of predicting the study duration compared with methods without covariates. The proposed method, based on dynamic predictions using joint models, would be useful for accurately predicting the study duration if a longitudinally measured covariate is available.

Reference: *[1] D. Rizopoulos, Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data, Biometrics, 67, 2011, 819-829.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-22

Tremor classification by digital spiral analysis in a large population sample

Roberto Melotti[1], Giulia Barbieri[1,2], Franziska Hopfner[3], Yuri D'Elia[1,4], Martin Goegele[1], Michele Filosi[1], Fabiola Del Greco-M[1], Peter Pramstaller[1], Cristian Pattaro[1], Gregor Kuhlenbaeumer[5]

1 Institute for Biomedicine, Eurac Research, Bolzano/Bozen, Italy
2 Unit of Epidemiology and Medical Statistics, Department of Diagnostics and Public Health, University of Verona, Verona, Italy
3 Department of Neurology, Ludwig-Maximilians University, Munich, Germany
4 Prusa Research a.s., Prague, Czech Republic
5 Department of Neurology, Kiel University, Kiel, Germany

Introduction: Spiral drawing is a clinical diagnostic tool to aid classification of action tremor types, which are typical of essential tremor and other neurodegenerative disorders of complex etiology. However, objective tremor classification in the population is scant. In the Cooperative Health Research in South Tyrol (CHRIS) study, we derived novel metrics through digital spiral analysis (DSA) and assessed their performance for tremor classification.

Methods: Archimedes' digital spiral drawing was performed with both hands by 10,983 adult CHRIS participants, resulting into multiple DSA measures of tremor amplitude, frequency, acceleration, speed, pressure, direction, and power. Both random and opportunistic sampling of 5,810 paired spirals from 2,905 participants (median age 50 years; age range:18-93 years; 53% females) were visually rated on screen in random order by an expert neurologist on a scale between 0 (no tremor) and 9 (maximum tremor) with 1 to 3 assignments per spiral. Aggregating small relative frequencies in extreme scores, we performed ordinal random forest (ORF) analysis on four ordinal outcomes: scoreH,5={<2,2,3,4,5+}, and scoreH,6={<2,2,3,4,5,6+}, taking the highest value; scoreR,5={<2,2,3,4,5+}, and scoreR,6={<2,2,3,4,5,6+}, taking one value at random, over repeated assignments per spiral. We investigated the classification performance of a selection of 37 DSA heterogeneous metrics, and both task-related and individual tremor determinants, totalling to 66 features. Five-fold cross-validation scheme was used to assess ORF model performances based on the rank probability score function. Linearly weighted kappa statistics between the observed and the predicted classes were obtained in the testing folds for average classification performance and based on individual class probability distributions over 1000 simulations for individual class forecasting performance. For each outcome, features were ranked by importance (I) for classification accuracy. We then run a final optimized ORF for out-of-sample classification on the whole dataset.

Results: Spiral scores ranged between 0 and 9 (median=2; IQR=2-3). Kappa statistics were fair and internally consistent for both scoreH (prediction range=0.397-0.483, forecasting range=0.281-0.342) and scoreR (prediction range=0.373-0.443, forecasting range=0.256-0.326). Out of 20 top-ranked variables for classification accuracy, 13 to 14 were in the amplitude domain for either scoreH or scoreR (I=0.0026-0.0168). Importance values of age, sex, and drawing-hand dominance had generally lower rankings (I<0.0045). As expected by opportunistic sampling design, final out-of-sample best class predictions were lower than the top-tremor classes for all score classifications.

Conclusion: A mixture of DSA semi-automated derived metrics can build a stable classification model for action tremor in large scale studies of tremor related disorders.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-23 Construction of decision tree using prior information

Asanao Shimokawa
Tokyo University of Science, Tokyo, Japan

Background/Introduction: The prediction of interesting outcome based on patient information is an important problem in medical research. Decision tree is widely used as one choice to achieve this problem, and the classification and regression tree (CART) algorithm is a popular method to construct these models. The CART algorithm allows the automatic construction of optimal combination of splitting rules to construct subsets of covariates spaces. A key setting in the algorithm is how to measure the reduction of impurity in a node due to splitting of covariate space and creating two child nodes. The reduction of impurity in a node due to splitting is estimated by using the estimates of the probability that any case falls into a child node and predetermined criteria of impurity within a node like mean squared error or Gini index. Compared with the effect of differences between criteria of impurity within a node the effect between estimation methods of the probability that any case falls into a child has not much research. In many cases, nonparametric estimators with simple empirical distributions are used. In recent years, however, advances in data collection and storage technology have meant that there is many information of covariate distribution available for model building other than learning data. For example, it could be the distribution of age and medical history from the census or real-world data collected by various sensors. If such information can be used for model construction, it is expected that more accurate models can be constructed.

Methods: In order to construct decision trees using prior information on covariates, we will examine the methods that use estimators of the covariate distributions obtained from data other than learning data in the splitting and pruning steps. Three estimation methods are compared: maximum likelihood estimation when assuming a parametric distribution, non-parametric estimation by kernel density estimation, and Bayesian estimation. A theoretical comparison of model building methods using these estimators and a simulation comparison under various situations were conducted.

Results: Comparative studies have shown that in many situations, model building methods using prior information provided more accurate models than conventional methods.

Conclusion: In this research, we proposed the construction method using prior information of covariate distributions. From comparative studies, the proposed method has higher accuracy in many situations than conventional method.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-24 Development and validation of a model predicting post-operative shoulder stiffness by patients undergoing an arthroscopic rotator cuff repair in Switzerland

Thomas Stojanov[1], Soheila Aghlmandi[2], Laurent Audigé[1], Philipp Moroder[3], Alexandre Lädermann[4], Study Group ARCR_Pred[1], Andreas Marc Müller[1]
1 University Hospital Basel, Switzerland
2 University Children's Hospital of Basel, Basel, Switzerland
3 Schulthess Clinic, Zürich, Switzerland
4 La Tour Hospital, Meyrin, Switzerland

Background: Post-operative shoulder stiffness (POSS) is one of the most frequent adverse events (AE) after an arthroscopic rotator cuff repair (ARCR). Predicting POSS occurrence could support healthcare in closely monitoring patients with a high risk of occurrence of AE after surgery. We aimed to update and validate a clinical prediction model for POSS, using data from a multicenter cohort setting (ARCR_Pred study).

Methods: Between June 2020 and November 2021, 973 primary ARCR patients were prospectively enrolled in a multicenter cohort across 18 Swiss and one German orthopedic center. POSS was defined and approved by a panel of 44 Swiss surgeons as a composite outcome (1/ A persisting post-operative restriction at 6 months in passive motion in at least two range of motion planes or 2/ symptomatic stiff shoulder leading to deviation from the routine postoperative management up to 6 months). A set of 35 factors including baseline characteristics and operative details was identified via database screening. Prognostic relevance of each variable was assessed by a panel of 44 Swiss surgeons. To identify the set of factors maximizing the apparent and bootstrap validated area under the receiver-operating characteristics curve (AUC), a backward elimination procedure was performed. Associations were reported in terms of Odds Ratio (OR) along with their 95% confidence intervals.

Results: Overall, 105 patients (10.8%) out of 973 had a POSS occurring within the six months post-surgery. The backward elimination procedure led to retaining 10 factors in the final model. Being a female (OR = 1.52 [0.96;2.40]), being a current (OR = 1.56 [0.91;2.61]) or former smoker (OR = 1.70 [0.98;2.90]), drinking alcohol daily (OR = 1.78 [0.92;3.27]), carrying moderate (OR = 1.56 [0.97;2.53]) or heavy (OR = 1.71 [0.91;3.17]) loads in working activities under the head, pre-operative medication (OR = 1.58 [1.03;2.43]), traumatic tears (OR = 1.56 [0.94;2.60]) and longer symptoms duration (OR = 1.79 [1.09;2.98]) were associated with a higher risk of POSS. Greater acromiohumeral distance (OR = 0.85 [0.77;0.93]), better pre-operative functional scores (OR =0.95 [0.93;0.98]) and greater passive motion in abduction (OR = 0.99 [0.97;1.00]) were associated with lower risk of POSS. The apparent and bootstrap-validated model performances were AUC = 0.73 and 0.68, respectively.

Conclusions: This model combining ten various baseline factors will be useful for healthcare eager to provide their patients with individualized predictions of risks of post-operative AE. Model performance improvement is expected by including the 6-week data.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-25

Can a clinical model be used to update a radiomics deep learning model for prediction of lung nodule malignancy?

Cristian Tebe[1,2], João Carmezim[1], Sonia Baeza[2], Guillermo Torres[3], Carles Sanchez[3], Débora Gil[3], Antoni Rosell[2]
1 Germans Trias i Pujol Research Institute and Hospital (IGTP), Badalona, Spain
2 Hospital Universitar Germans Trias i Pujol, Badalona, Spain
3 Universitat Autonoma de Barcelona, Bellaterra, Spain

Background: Radiomics is a promising area of research for the diagnosis of lung cancer. However, current radiomic models are limited in their accuracy, mainly because they are based only on image features. Our aim is to evaluate the integration of a clinical prediction model and a radiomic model based on deep learning to predict the malignancy of lung nodules. Methods A cross-sectional study of 87 resected nodules from 90 patients was performed. Clinical data included demographic variables, epidemiological risk factors and pulmonary function tests. The region of interest of each chest CT containing the nodule was extracted and analysed. The radiomic model was estimated by a convolutional neural network using MobileNetV2 [1] based on 333 visual features. The clinical model was estimated by a logistic regression model using the epidemiological and lung function risk factors. The malignancy probability from the clinical model was used as the best estimate of the pre-test probability of disease to update the malignancy probability of the radiomics model using a nomogram for Bayes' theorem [2].

Results: The mean age of the patients was 65 years, ranging from 35 to 84 years, 65% were men and 43% had a family history of cancer. The nodules had a mean diameter of 17.9 mm, histology: 59% adeno, 15% squamous, 2% other malignant, 1% carcinoid and 23% benign. The incidence of nodal malignancy in the study was 77% (95% CI: 67%-85%). The radiomic model had a positive predictive value of 0.89 (95% CI: 0.77-0.96), an AUC of 0.68 (95% CI: 0.52-0.83), and an accuracy of 0.69 (95% CI: 0.59-0.79). The clinical model identified DLCO, obstruction index and smoking status as the most consistent clinical predictors associated with outcome. Using the clinical model estimate as the pre-test probability significantly improved the estimate of lung nodule malignancy. The final positive predictive value was 0.95 (95% CI: 0.84-0.99), the AUC was 0.78 (95% CI: 0.66-0.90) and the accuracy was 0.66 (95% CI: 0.55-0.75).

Conclusion: Integrating clinical information into a deep learning radiomic model for the assessment of malignant lung nodules improved the model's predictive performance. Limitations include the need for a large patient sample and external validation of the model to gain confidence in these results. This study supports the potential of combining image-based and clinical features to improve lung cancer diagnosis.

References: *[1] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Jun;4510–20. [2] Fagan TJ. Letter: Nomogram for Bayes's theorem. N Engl J Med. 1975 Jul 31;293(5):257.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-26

Stronger penalties on treatment-covariate interactions improve the ability to predict treatment effect

David Van Klaveren[1,2], Ewout Willem Steyerberg[3], David Michael Kent[2]
1 Department of Public Health, Erasmus MC University Medical Center, Rotterdam, Netherlands
2 Predictive Analytics and Comparative Effectiveness Center, Tufts Medical Center, Boston, United States
3 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

Background: Prediction models with treatment covariate interactions (effect-modifiers) can identify patient groups with differential treatment effects, but often result in exaggerated variation in treatment effect predictions, potentially leading to treatment mistargeting. We have shown previously that LASSO regression does not fully prevent this overestimation of treatment effect heterogeneity We aimed to propose shrinkage techniques that further reduce overestimation of treatment effect heterogeneity.

Methods: To obtain a uniform shrinkage factor for treatment interactions, we first decomposed model predictions on the logit scale into: a prognostic index; an average treatment effect; and the deviation from the average treatment effect due to treatment covariate interactions We then used cross validation to estimate shrinkage factors for these components Alternatively, we used adaptive LASSO regression to obtain a multiplicative penalty factor for interaction effects (versus 1 for main effects) based on the size of interaction compared to main effects in a preliminary model fit. We exploited a previously published simulation framework (samples of 3,600 patients; 25% binary outcomes; 12 binary covariates) - both without and with 6 true treatment-covariate interactions - to compare these approaches with a risk model including a constant relative treatment effect and with a standard LASSO model with treatment-covariate interactions. We measured discrimination in a superpopulation of 1,000,000 patients by the arithmetic difference between the true absolute treatment effect in the fourth and first quarter of predicted absolute treatment effect (extreme quarter difference; "EQD"). We measured calibration by the arithmetic difference between predicted and true absolute treatment effect in the fourth quarter of predicted absolute treatment effect ("bias").

Results: In the absence of true treatment-covariate interactions, the median uniform shrinkage factor (0.29) and the median adaptive penalty factor (2.39) for treatment co variate interactions were strong. Discrimination and calibration of the uniform shrinkage technique (median-EQD 1.5%; median bias 2.2%) and especially adaptive LASSO (median-EQD 2.5%; median-bias 1.2%) were closer to the correctly specified risk model (median-EQD 2.6%; median-bias 0.3%) than the standard LASSO (median-EQD 0.7%; median-bias 2.5%). In the presence of true treatment covariate interactions, the median uniform shrinkage factor (0.62) and the median adaptive penalty factor (1.60) for the treatment interactions were still substantial. The predictive performance of the proposed shrinkage techniques (median-EQD 8.6% and 7.4%; median-bias 0.9% and 0.4%) was similar compared to the standard LASSO (median-EQD 8.0%; median-bias 0.6%) and much better compared to the risk model median (EQD-1.4%; median-bias 0.7%).

Conclusion: Data driven techniques targeted at shrinkage of treatment interaction effects improve a model's ability to predict treatment effects.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-27

Development and validation of postpartum cardiovascular disease (CVD) risk prediction model in women with a history of pregnancy incorporating reproductive and pregnancy-related candidate predictors.

Steven Wambua[1], Kym I E Snell[1,2], Francesca Crowe[1], Shakila Thangaratinam[3,4], Colin McCowan[5], Sinead Brophy[6], Christopher Yau[7,8,9], Krishnarajah Nirantharakumar[1], Richard D Riley[1,2]

1 Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom
2 National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, United Kingdom
3 WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, United Kingdom
4 Department of Obstetrics and Gynaecology, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, United Kingdom
5 School of Medicine, University of St Andrews, St Andrews, United Kingdom
6 Data Science, Medical School, Swansea University, Swansea, United Kingdom
7 Big Data Institute, University of Oxford Li Ka Shing Centre for Health Information and Discovery, United Kingdom
8 Nuffield Department of Women's & Reproductive Health, University of Oxford, John Radcliffe Hospital, United Kingdom
9 Health Data Research, London, United Kingdom

Background: Current risk prediction models for cardiovascular disease (CVD) do not include pregnancy-related factors in the models for women despite existing evidence showing that pregnancy complications (e.g., hypertensive disorders of pregnancy, placental abruption, preterm birth, gestational diabetes mellitus, stillbirth) and reproductive factors (e.g., early age at menarche and polycystic ovary syndrome) are associated with risk of CVD in women. We aimed to determine whether adding pregnancy factors to a prediction model with established risk factors for CVD improves 10- year risk prediction of CVD in postpartum women using the QRISK3 risk equation as a benchmark.

Methods: We used a population-based retrospective cohort of women aged 15 to 49 with a history of pregnancy from the Clinical Practice Research Datalink (CPRD) primary care database. We identified established risk factors for CVD in the general population such as age, ethnicity, deprivation, family history of CVD, diabetes mellitus and some medications use from the QRISK3 risk prediction model. We then identified additional risk factors specific to women with a history of pregnancy from an umbrella review and from discussions with clinicians and patient research partners. We then externally validated QRISK3 in our cohort of women and developed updated risk prediction models for CVD using Cox-proportional hazards models including established risk factors from QRISK3 and comparing with a model that included additional pregnancy-related factors. Models were evaluated using measures of discrimination, calibration and clinical utility.

Results: Among 567,667 women with a history of pregnancy between 15-49 years of age who were CVD free at 6 months after delivery, 2,175 (0.38%) experienced a CVD event during follow-up. Although adding pregnancy factors to those from QRISK3 led to small improvements in model performance (QRISK3 c-statistic: 0.703 (95% CI 0.688 to 0.718), Updated model with traditional risk factors c-statistic: 0.721 (95% CI 0.707 to 0.736), Updated model with additional pregnancy factors c-statistic: 0.728 (95% CI 0.713 to 0.743), calibration and the clinical utility of updated models was better.

Conclusions: Adding pregnancy and reproductive history to established risk factors of CVD from QRISK3 did not significantly improve discrimination of the risk prediction models in the low-risk population of young postpartum women. This could be due to the known association of pregnancy-related complications with established risk factors of CVD. The additional factors however led to better calibrated models. Future research could evaluate and demonstrate feasibility of introducing CVD risk assessment in postpartum women into clinical practice.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-28

Comparison of different strategies in using Lasso in clinical prediction models for rare outcomes: A simulation study

Junfeng Wang[1], Haoyue Wang[2], Jos P. Kanning[3], Shengfeng Wang[4]
1 Utrecht University, Utrecht, Netherlands
2 Amsterdam University Medical Centers, Amsterdam, Netherlands
3 University Medical Center Utrecht, Utrecht, Netherlands
4 Peking University, Beijing, China

Background: The Lasso regression algorithm is a common and efficient method for fitting models and selecting variables simultaneously and is widely used in prediction model development. In order to shrink more coefficients to zero, the penalty parameter needs to be increased but the over-shrunken coefficients may lead to underestimation of the event probability. There are currently no clear recommendations for model refitting when variable selection is performed using an increased penalty factor, and whether penalization is necessary during refitting to avoid overfitting.

Methods: We conducted a simulation study to compare the performance of variable selection and prediction of eight strategies in applying Lasso. In total eighty-one scenarios were generated, and model performance was assessed as the average in 100 simulation runs.

Results: The results show that classification-based selection is worse than AUC based selection; 1se models had better variable selection than optimal models; model refitting can improve the model performance; penalization in model refitting may bias the predicted probability.

Conclusions: Lasso is a useful tool for variable selection in prediction model development. When the number of candidate predictor is large and event rate is low, the penalty parameter yields the best cross-validated performance may fail in excluding noise predictors. In this case, further increasing the penalty parameter is recommended, and model refitting is needed. Model refitting without further penalty shows good discrimination and calibration, while model refitting with penalty overestimates the predicted probability and needs correction.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-29

Predictive accuracy metrics in the context of interval censoring and competing risks

Zhenwei Yang[1], Dimitris Rizopoulos[1], Eveline Heijnsdijk[1], Lisa Newcomb[2], Nicole Erler[1]
1 Erasmus Medical Center, Rotterdam, Netherlands
2 Fred Hutchinson Cancer Center, Seattle, United States

Background: Evaluating prediction models, in terms of calibration and discrimination, has been widely discussed in the standard setting and with competing events. However, in medical research, interval-censored events are common. With interval censoring coming along, the ambiguity of the underlying unobserved event times brings challenges in defining cases and controls, which is often overlooked. This study takes the time-dependent AUC as an example to enhance the general applicability of the predictive accuracy measure. In our case, a joint model is used to predict the primary event, cancer progression, which is interval censored by periodic biopsies, and the (exactly observed) competing event, early treatment initiation. For many patients it is unknown whether they experienced the primary event within the interval of interest for calculating the time-dependent AUC, (t, t+dt]. For instance, even when a patient experiences the competing event, he may still have had undetected cancer progression. We develop a model-based approach to determine the time-dependent AUC, adapt the inverse probability of censoring weighting (IPCW) method that deals with right-censoring to our interval-censored competing risk setting, and explore a non-parametric hybrid approach.

Methods: While in the IPCW approach, the AUC calculation is restricted to patients for which the event of interest is known to be in (t, t+dt], our model-based approach takes advantage of the whole sample. Patients are considered to experience cancer progression before, in or after (t, t + dt] with probabilities according to their estimated cumulative risk in the corresponding time intervals. This approach has the advantage over IPCW that it makes use of the longitudinal biomarkers which typically are essential in predicting the underlying event. A limitation is, however, that the model is used in its own evaluation which may result in overly optimistic performance estimates. Therefore, we additionally explore the use of a non-parametric approach as an alternative to obtain the probabilities with which a patient has the event in, before or after (t, t+dt]. The methods are demonstrated using data from the Canary PASS and compared in a simulation study.

Results: For the Canary PASS data, the model-based approach resulted in a time-dependent AUC of 0.61 for the interval between years 1 and 4, while the corresponding estimate from the IPCW approach was 0.64.

Conclusion: While the methods estimated similar predictive performance in the real data, the simulation study is needed to compare the estimates to the "true" AUC (in absence of censoring).

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-30 Stop before you start: A checklist for those thinking about developing a clinical prediction model

Lucinda Archer[1,2], Rebecca Whittle[1,2], Kym Snell[1,2], Paula Dhiman[3], Gary Collins[3], Richard Riley[1,2], Joie Ensor[1,2]
1 Institute of Applied Health Research, University of Birmingham, United Kingdom
2 National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, United Kingdom
3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom

Background: Clinical prediction models are currently being published at an alarming rate, but most are of poor quality and have no impact on clinical practice. A major reason for this is that models are often developed with little prior consideration of clinical need, study design, and potential requirements for implementation. This ultimately hampers their predictive performance, uptake, and acceptability.

Methods: To improve research standards and building on previous work [1], we describe a new checklist of key issues to consider prior to developing a clinical prediction model. This checklist was formed through consultation and consensus among experts in the prediction modelling field, both methodological and clinical. We outline the potential impact of neglecting each item on the

Results: The checklist covers key issues researchers should consider when planning to develop a predictive performance, clinical utility, and acceptability of the developed model. clinical prediction model and is applicable irrespective of clinical area or modelling approach. Items surrounding study design include: recruitment of and consultation with a relevant Patient and Public Involvement and Engagement (PPIE) group; full consideration of the research question under investigation (e.g., using PICOTS: Population, Implementation, Competing models, Outcomes, Timing, Setting) with input from key stakeholders; and identification of the necessary sample size to develop a stable and generalisable model [2,3]. Regarding consideration of model implementation, items cover: ensuring model outcomes and potential predictors are both measurable and acceptable; understanding why any existing models in the field are not currently being used, and what improvements are needed; and assessing potential barriers to implementation in clinical practice, both on a population-level and in key subgroups.

Conclusion: Prediction model development should be preceded by crucial preparatory work to help ensure that the resulting models are statistically robust and clinically relevant. The checklist should be considered at the pre-protocol stage of a research project and can help guide researchers on whether they should be developing a prediction model or not. Overall, the checklist aims to ensure investigators achieve maximum relevancy and utility of their clinical prediction models, while reducing the risk of their work becoming research waste. Key words: prediction, model development, research planning

References: *[1] Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. BJOG. 2017. doi: 10.1111/1471-0528.14170. [2] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins GS, van Smeden M. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020. doi: 10.1136/bmj.m441. [3] Riley, RD, & Collins, GS. Stability of clinical prediction models developed using statistical or machine learning methods. Biometrical Journal. 2023. doi: 10.1002/bimj.202200302.*

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-31 LB

Uncertainty in clinical risk prediction: perspectives and approaches

Richard Riley[1,2], Gary Collins[4], Laura Kirton[3], Kym Snell[1,2], Joie Ensor[1,2], Rebecca Whittle[1,2], Paula Dhiman[4], Lucinda Archer[1,2]

1 Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom
2 NIHR Birmingham Biomedical Research Centre, Birmingham, United Kingdom
3 Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, United Kingdom
4 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, University of Oxford, Oxford, United Kingdom

Background: Each year, thousands of prediction models are published in the medical literature aiming to inform diagnosis or prognosis in a particular target population. These models enable an individual's risk of a health-related outcome to be estimated, though most provide only a point estimate of this risk and do not present any information on the corresponding uncertainty in their estimate (e.g., through uncertainty intervals or distributions). Before a prediction model can be considered for use in clinical practice, it should be critically appraised and rigorously evaluated. Unfortunately, many published models are not fit for purpose due to poor methodological standards that lead to high levels of uncertainty in predictions (e.g., use of small sample sizes, inappropriate model development techniques, and a lack of external evaluation). There is inconsistency in the presentation of uncertainty around risk estimates in practice, and much debate surrounding its usefulness. In general, uncertainty in predicted risk is ignored.

Methods: We outline perspectives on the presentation of uncertainty in risk estimates from a clinical prediction model and propose methods to quantify this uncertainty in practice.

Results: Deriving uncertainty intervals fully conditional on key patient attributes is the most appropriate option for tailoring uncertainty estimates to the individual, though is computationally complex. We demonstrate how such intervals can be derived using the variance-covariance matrix of the parameter estimates (following a frequentist approach) or by sampling from the posterior distribution for an individual's risk (when fitting the model in a Bayesian framework). More generally, bootstrapping can be used to derive uncertainty estimates conditional on the individual's predicted risk, regardless of model development approach, though such estimates may not suitably account for patient-level characteristics.

Conclusion: Presentation of uncertainty in predicted risk allows end-users and stakeholders to evaluate and critically appraise a prediction model, and can direct further research for developing and updating these models. Although point estimates of risk are open sufficient for individual decision making, acknowledging uncertainty may also enhance the patient-doctor consultation, though more research is also needed on how best to communicate uncertainty information to patients.

POSTER SESSION B-04:
Prediction and Prognostic Models

P-B04-32 LB

Uncertainty-based sequential sample size calculations for developing, validating and updating clinical prediction models

Amardeep Legha[1,2], Joie Ensor[1,2], Lucinda Archer[1,2], Rebecca Whittle[1,2], Kym Snell[1,2], Paula Dhiman[3], Gary Collins[3], Richard Riley[1,2]

1 Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom
2 National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, Birmingham, United Kingdom
3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

Background: Clinical prediction models estimate an individual's risk of particular outcomes to inform clinical decision-making. An adequate sample size is required to ensure model predictions are robust and predictive performance is precisely estimated. Current sample size calculations are mainly done before data collection, leading to a fixed sample size target based on sensible assumptions. However, adaptive sample size calculations can be used during data collection, to sequentially examine when additional data are (not) required based on information obtained. This may lead to more reliable models being developed and evaluated than using a fixed sample size approach.

Methods: We propose uncertainty-based sequential sample size calculations, based on a value of information approach for individual-level decision making. Our method focuses on the uncertainty of individual-level predictions and classifications based on risk thresholds for decision making. Beginning at the minimum sample size recommended before data collection, we derive and sequentially update prediction and classification instability plots as new participants are recruited. A stopping rule is based on the perceived value of additional information; crucially this is context specific, that can be guided by the level of uncertainty and classification errors that stakeholders (e.g. patients, clinicians) are willing to accept. Our approach is illustrated with real examples of studies developing, validating and updating a model. The number of predictors and model development approach are varied, including regression and machine learning approaches.

Results: Our findings show that the sample size required is strongly dependent on the risk thresholds of interest for decision making, and the level of prediction and classification instability deemed acceptable by stakeholders. The sequential approach often leads to much larger sample sizes than the fixed sample size calculated prior to data collection, especially when models are based on machine learning approaches that allow more complexity. As individual-level uncertainty can often be wide, we also demonstrate the impact of stopping rules based on targeting suitable precision in groups of individuals defined by the same or similar predicted risk.

Conclusions: An uncertainty-based sequential sample size approach allows users to dynamically monitor and identify when enough participants have been recruited to reliably develop, evaluate or update their prediction model. This new approach for studies carrying out prospective data collection helps ensure more reliable models that optimise clinical decisions for individuals.

POSTER SESSION B-05:
Diagnostic and Screening Tests

P-B05-01 Inference on the symmetry point-based optimal cut-off point and associated sensitivity

Alba M. Franco Pereira, M. del Carmen Pardo, Christos T. Nakas, Benjamin Reiser
Universidad Complutense De Madrid, Spain

Background/Introduction: When considering a continuous biomarker, clinicians need an optimal threshold to classify subjects into one of the two diagnostic groups. One such method is based on the symmetry point. The symmetry point has been used in applications but typically without the addition of confidence intervals. Methods We have considered both marginal confidence intervals (CIs) for the symmetry point and its associated sensitivity and confidence regions (CRs) for these jointly. For the marginal CIs we examined both parametric and nonparametric approaches. The parametric approaches are based on the binormal model and implemented using the Box – Cox transformation for cases when normality assumptions are not met. Other are based on bootstrap techniques. Two nonparametric methods are examined. The first due to López- Raton et al. (2016) is based on empirical likelihood while the second is a new proposal based on the chi-square test statistic. For the CRs only nonparametric approaches are considered. The first due to Adimari and Sinigaglia (2020) is based on empirical likelihood. We propose an alternative based on empirical chi-square.

Results: One bootstrap method worked well for the associated sensitivity for a wide range of distributions except for the mixtures. However, nonparametric methods, which give very similar results, are recommended for constructing CIs for the symmetry point. Simulations indicate that although the two methods for CRs perform similarly, our proposal generally provides results with observed coverage closer to the nominal level. We illustrate these procedures using part of the dataset from a published study on SARSCoV- 2 antibody levels post vaccination.

Conclusion: In addition to the symmetry point approach there are many other methods proposed in the literature for obtaining the optimal cut-off point value and there is an inherent problem in choosing the appropriate method for the selection of an optimal cutoff point that can be used in everyday practice. The availability of CIs and CRs for the symmetry point approach should help practitioners using this method in their data analyses.

References: *[1] Adimari, G. and Sinigaglia, A. (2020). Nonparametric confidence regions for the symmetry point-based optimal cutpoint and associated sensitivity of a continuous-scale diagnostic test. Biometrical Journal, 62, 1463– 1475. [2] López-Ratón, M., Cadarso-Suárez, C., Molanes-López, E. M. and Letón, E. (2016). Confidence intervals for the symmetry point: An optimal cutpoint in continuous diagnostic tests. Pharmaceutical Statistics, 15, 178–192.*

POSTER SESSION B-05:
Diagnostic and Screening Tests

P-B05-02 A joint model of tumour volume and mode of detection for a cohort of incident breast cancer cases

Letizia Orsini, Kamila Czene, Keith Humphreys
Karolinska Institutet, Stockholm, Sweden

Introduction: In recent years, as an alternative to multi-state models, several continuous tumour growth models have been developed for the analysis of breast cancer screening data. These represent a relatively novel and useful approach for analysing screening data. Continuous tumour growth models use random effects and growth functions to model the (latent) growth in the size of tumours.

Methods and Results: Based on a continuous tumour growth model and using previous findings regarding stationary distributions of tumour size and growth rate distributions in non-screened populations [1], we develop an analytical expression for the proportion of interval breast cancer cases (cases diagnosed due to symptoms between screening rounds) among regularly screened women. We show how the probability of symptomatic detection within a screened population can be jointly modelled with the tumour volume at detection conditioning on the screening history. We fit a model on 3493 cases diagnosed in Sweden between 2001-2008. Our approach avoids relying on estimated background cancer rates. We make specific parametric assumptions concerning tumour growth and detection processes (screening or symptoms), but our framework easily accommodates alternative assumptions. Our methodology allows us to estimate the distribution of tumour sizes at the most recent screening for interval cancers. Importantly, we find that our model-based expected incidence of interval breast cancers aligns closely with observed patterns in our study and a large Nordic screening cohort [2]. Finally, we evaluate the association between screening interval length and the interval cancer proportion.

Conclusions: An analytical expression for the proportion of interval breast can be used to jointly model the tumour volume and the mode of detection, and to give insights into the effectiveness of population-based breast screening.

References: *[1] Isheden, G., & Humphreys, K. (2022). A unifying framework for continuous tumour growth modelling of breast cancer screening data. Mathematical Biosciences, 353, 108897. [2] Hofvind S, Tsuruda K, Mangerud G, Ertzaas AK, Holen AS, and Pedersen K. The Norwegian Breast Cancer Screening Program, 1996-2016: celebrating 20 years of organised mammographic screening. Cancer in Norway 2016 :S1–S72*

POSTER SESSION B-05:
Diagnostic and Screening Tests

P-B05-03

Agreement of a latent class model with an expert panel for defining the reference standard in a diagnostic imaging study

Tom Parry[1], Sue Mallett[1], Nandini Dendukuri[2]
1 UCL, London, United Kingdom
2 McGill University, Montreal, Canada

Introduction: Evaluating the accuracy of diagnostic imaging tests is complex without a single "perfect" test to define the reference standard for all participants. Frequently, studies use an expert panel of clinicians to combine multiple "imperfect" tests into a reference standard. We aim to assess using a latent class model (LCM) as an objective and efficient alternative method. We assessed the agreement of an LCM with a panel in a multi-centre diagnostic imaging study of terminal-ileal Crohn's disease (TICD) activity.

Methods: We analysed data from participants prospectively recruited with high suspicion of active TICD (METRIC [ISRCTN03982913]). All participants received an MRI and ultrasound as index tests. A panel defined the reference standard based on at least one of colonoscopy, histology, C-reactive protein, or faecal calprotectin. We prespecified tests for the LCM with clinical expertise, including MRI, ultrasound, the tests used by the panel, and the participant-reported Harvey-Bradshaw index. We also prespecified a posterior probability threshold of ≥10% to classify participants with active TICD. After multiple imputation, we fit a 2-class Bayesian LCM with flat priors and a random effect to account for conditional dependence among tests. We estimated the prevalence, positive-specific, negative-specific, and overall agreement of the LCM and the panel.

Results: In our analysis of 284 participants, the panel identified active TICD in 69% of participants and the LCM estimated it in 79%. Relative to the panel, the LCM correctly classified 82% (95% CI 77, 86) of participants, had an 88% (83, 91) probability of agreement on a participant with active TICD, and had a 65% (59, 70) probability of agreement on a participant without active TICD. Disagreement arose because the LCM weighted MRI and ultrasound as the most diagnostically important tests.

Conclusion: Our analysis highlights the potential of LCMs to define the reference standard in diagnostic imaging studies. Future analyses will assess LCMs in studies with different diseases, panel methods, and sample sizes.

POSTER SESSION B-05:
Diagnostic and Screening Tests

P-B05-04

Reporting completeness of systematic reviews and meta-analysis of diagnostic test accuracy studies published in 2022 based on the PRISMA-DTA Reporting Guideline: An empirical study

Ioanna Tarelli[1], Natalia Papadimitriou[1], Eirini Pagkalidou[1], Sofia Tsokani[1,2], Anna-Bettina Haidich[1]
1 Department of Hygiene, Social-Preventive Medicine & Medical Statistics, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece
2 Methods Support Unit, Cochrane CET, United Kingdom

Background/Introduction: In the world of evidence-based medicine, the reporting completeness of systematic reviews (SRs) and meta-analyses (MAs) plays a critical role in ensuring the transparency and reliability of research findings. Despite the existence of reporting guidelines like PRISMA-DTA, adherence to reporting guidance remains problematic.

Methods: We are currently conducting a systematic review of MAs of diagnostic test accuracy (DTA) studies published in 2022. We aim to evaluate the reporting completeness of the eligible publications and explore various factors (e.g., protocol existence, funding, journal impact factor, etc.) based on the PRISMA-DTA checklist. We will assess adherence to PRISMA-DTA checklist using a total score, based on whether each of the 27 PRISMA-DTA items are reported or not.

Results: Our search on MEDLINE via PubMed, Epistemonikos, Scopus and Cochrane Library yielded 1603 publications in 2022, eligible for inclusion. We will present our findings using descriptive statistics, regression analyses, and graphical representations. All statistical analyses and graphical depictions will be performed using R software.

Conclusions: This study emphasizes the importance of thorough and transparent reporting in DTA research. Adhering to reporting guidelines improves the quality of reports, aiding informed decision making in healthcare.

References: *[1] Isheden, G., & Humphreys, K. (2022). A unifying framework for continuous tumour growth modelling of breast cancer screening data. Mathematical Biosciences, 353, 108897. [2] Hofvind S, Tsuruda K, Mangerud G, Ertzaas AK, Holen AS, and Pedersen K. The Norwegian Breast Cancer Screening Program, 1996-2016: celebrating 20 years of organised mammographic screening. Cancer in Norway 2016 :S1–S72*

POSTER SESSION B-05:
Diagnostic and Screening Tests

P-B05-05

Automated data screening and data quality checks using the dataquieR R package

Carsten Oliver Schmidt, Elisa Kasbohm, Joany Mariño, Elena Salogni, Stephan Struckmann
University Medicine Greifswald, Germany

Background/Introduction: A statistician's analysis workflow usually begins with the checking and cleaning of a data body before proceeding to any substantive scientific analyses. This starting point frequently consumes a large part of the analysis resources. This work shows how the efficiency and transparency of this essential step in the analysis workflow can be improved based on the recently updated R dataquieR package.

Methods: The R package dataquieR consists of almost 20 core functions and more than 250 support functions, which enable an automated creation of complex reports. The support functions also ensure robustness to a priori unknown data errors. For the output, dataquieR renders HTML files using htmltools and uses plotly to enable interactive graphs. The package is available on the comprehensive R archive network (CRAN) [1].

Results: dataquieR integrates a wide range of data checks in a one-stop-shop, including, amongst others descriptive statistics, range and contradiction checks, unit and item missingness overviews, univariate and multivariate outlier assessments, checks of expected associations and distributions, mixed models and nonparametric regression techniques to inspect observer, device or centre effects, as well as time trends. Assessments are controlled based on a metadata model containing knowledge and assumptions about (1) individual variables (e.g. range violations, expected range of point estimates), (2) groups of variables (e.g. contradictions), (3) data segments (e.g. expected case count in an examination), (4) data tables (e.g. expected variable selection), (5) classification of missing codes by linking study specific missing value codes with a generic ontology. Each metadata table can be treated as a spreadsheet in an Excel workbook for easy editing. The contained information triggers the entire analysis pipeline or selected parts of it. dataquieR has been applied to cohort study as well as health registry data.

Conclusion: The workflow implemented in dataquieR illustrates an often neglected approach: the systematic separation of assumptions and knowledge underlying analyses from the actual analysis code. Adhering to this separation not only increases efficacy, but also contributes to a more systematic implementation and significantly greater transparency of data quality checks and their reporting.

References: *[1] Struckmann S, Mariño J, Kasbohm E, Salogni E, Schmidt CO: dataquieR [2]: An updated R package for FAIR data quality assessments in observational studies and electronic health record data. 2024 Zenodo (preprint under review) https://zenodoorg/records/10722214.*

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-01

Bayesian Spatial modelling of mental health incidence data: A comparison of approaches

Hollie Hughes, David Hughes
University of Liverpool, Liverpool, United Kingdom

Background/Introduction: Depression is one of the most prevalent mental health disorders worldwide and many factors can affect the risk of occurrence, including both social and geographical factors. The Besag York Mollie (BYM) model has provided a popular solution to the spatial auto-correlation problem in modelling and mapping data such as depression incidence. Markov Chain Monte Carlo (MCMC) and Integrated Nested Laplace Algorithm (INLA) approaches are the most frequently chosen approaches for fitting a BYM model in Bayesian spatial statistics. However, both MCMC and INLA offer limitations to Bayesian inference in spatial statistics. MCMC provides a substantially large issue with computational cost, and although INLA was proposed as an alternative to MCMC to reduce complexity, the accuracy of INLA is somewhat unclear. Concerns around INLA are raised by Khan et al. (2021), who discuss the lack of research into how reproducible the output from INLA is, as well as the lack of reporting on how INLA's accuracy is measured (1). Mean Field Variational Bayes (MFVB) techniques provide an alternative to MCMC and INLA, alleviating the problem of computational cost whilst still promising a high accuracy.

Method: The MFVB approach differs from the previous approaches through enforcing the mean field restriction to estimate the posterior distribution, through optimisation of a product of less complex density functions. In this presentation we will provide a comparison of MFVB, INLA and MCMC approaches to analysing the BYM model. Accuracy of results will be assessed both in terms of distributional similarity and the accuracy of point estimates. To do this, mental health data from the Place-based Longitudinal Data Resource (PLDR) will be used with a combination of freely available United Kingdom (UK) national demographics. We will compare the three aforementioned modelling approaches through assessing the computation time for each method and accuracy of the results obtained.

Results: Results will include time taken to fit models along with comparisons of accuracy, and a discussion of the accuracy tests involved. Evaluation is ongoing but full results will be presented at the conference.

Conclusion: We believe the MFVB approach may offer a solution to much faster Bayesian inference on spatial data with the potential of being substantially as accurate as the successful MCMC approach, in a much quicker time through the alleviation of sampling.

References: *[1] Khan K, Luo H, Xi W. Computing with R-INLA: Accuracy and reproducibility with implications for the analysis of COVID-19 data. arXiv; 2021.*

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-02 Data augmentation for improving parameter estimation in mixed models

Tina Košuta[1], Rok Blagus[1,2]
1 Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
2 University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia

Introduction: Generalized linear mixed models (GLMM) offer a flexible framework for modelling complex data with some correlation structure within the observations belonging to the same cluster, however, the estimation of the model parameters may often present a challenge. The parameter estimates, which are usually estimated by the maximum likelihood (ML) method, may lie on the boundary of the parameter space, the random effects covariance matrix estimate being more prone to this issue than fixed effects random coefficients. One of the possible strategies to obtain the non-boundary parameter estimates is to penalize the likelihood and to find the estimates by maximizing the penalized likelihood. A penalty term, chosen in a way that guarantees some desired frequentist properties, such as lower mean squared error or non-boundary estimates, is added to the model likelihood to obtain the penalized likelihood.

Methods: We propose an approximate Bayesian analysis using data augmentation prior on the random effects covariance matrix, to prevent it from taking the boundary estimates, where the prior acts as the penalization term added to the likelihood. To avoid implementing the maximization of the penalized likelihood, a non-trivial task due to the presence of intractable integrals in the model likelihood, the aim is to represent the penalty term as pseudo-observations whose contribution to the likelihood represents the penalty term. Contrary to simpler models, the task is non-trivial due to the hierarchical structure of the model and the intractable integral within the likelihood. When pseudo-observations are generated, they are added to the original data set and are jointly analysed by commonly used statistical software using ML for parameter estimation. The estimates obtained by ML on the augmented dataset, coincide with the penalized ML estimates (pML) and the maximum a posteriori (MAP) estimates.

Results: We show the relationship between the penalty term and pseudo-observations and the relationship between ML, pML, and MAP. We illustrate the performance of our proposed method using simulation studies and real-world datasets for different outcome distributions and various levels of complexity of random effects.

Conclusions: Our approach allows for simple approximate Bayesian GLMM, without requiring familiarity with Bayesian statistical software. Furthermore, it guarantees non-boundary estimates of random effects covariance matrix within a frequentist framework, potentially improving the inference of fixed effects regression coefficients.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-03

Injurytools: A toolkit for sports injury data analysis

Dae-Jin Lee[1], Lore Zumeta-Olaskoaga[2]
1 IE University, Madrid, Spain
2 Biogipuzkoa Health Research Institute, Donostia - San Sebastian, Spain

Background: Sports injuries pose significant challenges to athletes, affecting their performance, longevity in sports, and overall well-being. Traditional methods of injury data analysis often involve complex and time-consuming processes that may not effectively leverage the vast amount of data collected in sports contexts. The development of comprehensive tools for sports injury data analysis is crucial for identifying injury patterns, understanding risk factors, and implementing effective prevention strategies. The 'injurytools' package addresses these needs by offering a robust toolkit designed to streamline the analysis of sports injury data.

Methods: 'injurytools' R package is developed with the aim of providing an accessible, yet powerful, set of routines and utilities specifically tailored for sports injury data analysis. The toolkit encompasses functions for the initial preparation of injury data, enabling users to clean and structure their datasets for analysis efficiently. It includes sophisticated visualization tools to help analysts and researchers identify trends and patterns within their data intuitively. Furthermore, it facilitates both descriptive and model-based analyses, allowing for the examination of injury occurrences and the exploration of potential risk factors through statistical modeling approaches. The package is built to accommodate various types of sports injury data, ensuring broad applicability and flexibility in research and practical settings.

Results: Utilization of 'injurytools' significantly simplifies the process of sports injury data analysis, from data preparation to the execution of complex statistical models. Early adopters have reported reduced time spent on data processing and enhanced clarity in the interpretation of results. Through its informative visualizations, users have been able to communicate findings effectively, fostering a deeper understanding of injury mechanisms and prevention strategies among sports professionals and stakeholders. The package's comprehensive approach to data analysis has also facilitated the identification of previously unrecognized injury patterns and risk factors, contributing valuable insights to the field of sports medicine and injury prevention.

Conclusion: The 'injurytools' package represents a significant advancement in the field of sports injury data analysis. By providing a suite of standardized routines and utilities, it enables researchers, clinicians, and sports organizations to analyze injury data more efficiently and effectively. This, in turn, supports the development of targeted injury prevention strategies and promotes the health and safety of athletes. Future developments of 'injurytools' will continue to enhance its capabilities, further enriching the sports community's resources for combating the challenge of sports injuries. Injurytools is available on R CRAN at https://cran.r-project.org/web/packages/injurytools/

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-04

A case study of combining cluster analyses in clinical trial data

Gabriele Fresia[2], Anikó Lovik[1,2], Sofia Pazmino[3], Michael Doumen[3], Elias De Meyst[3], Patrick Verschueren[3]

1 Karolinska Institutet, Sweden
2 Leiden University, Netherlands
3 KU Leuven, Belgium

Background: Rheumatoid arthritis (RA) is a chronic, autoimmune disease that in most cases can be controlled by medication. After starting treatment, most patients have low disease activity or are in remission, although even among these patients some report residual symptoms. For this reason, it was assumed that a pooled cluster analysis over the entire study period would be the most useful in separating patients with adverse disease trajectories from those who were responding well to treatment.

Methodological issue: Combining clusters is far from straightforward. Potential problems include: 1) potentially differing number of clusters in different analyses, 2) the difficulty of matching clusters, for example, based on cluster centroid), 3) deciding on how to calculate cluster centroids of the pooled results, 4) observations potentially changing cluster membership and 5) subsequent difficulty in calculating cluster size.

Methods: A proof-of-concept analysis on a clinical trial dataset of 379 patients of the with rheumatoid arthritis participating in the CareRA trial with data collected at 8 visits. Hierarchical k-means clustering was combined from 2000 datasets obtained by taking 100 independent samples of each of 20 multiply imputed datasets. Clusters were aligned using the centroids produced by the cluster analysis applied to the centroids of all datasets as the "gold standard".

Results: Based on majority vote, four clusters were selected. As expected, the largest subset of patients comprised those who responded positively to the treatment and reported a relatively low disease burden. The method also identified cluster of patients with unmet psychological needs and patients with chronic inflammation.

Conclusions: Through cluster analysis, it is possible to identify patients with a shared disease activity pattern and similar patient-reported outcomes which might allow practitioners to employ individual treatment plans. This analysis study provides a proof-of-concept for the workflow of similar analyses.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-05

Comparison of the statistical performance of generalized pairwise comparisons, LWYY and Cox models for composite endpoints including recurrent and terminal events in clinical trials

Nicolas Sauvageot, Shu-Fang Hsu Schmitz
Johnson & Johnson, Switzerland

Introduction: The most common approach for analyzing composite endpoints in clinical trials is for time to first event using a Cox model. Subsequent events occurring after the first event are ignored with such an approach, which does not adequately capture patients' burden of disease and may lead to substantial loss of important information. As an alternative, recurrent events data can be analyzed using a Lin-Wei-Yang-Ying (LWYY) model, which also considers available information beyond the first event. The interpretation of treatment effect based on recurrent events should be done with caution in the presence of a terminal component event like death. To overcome this concern, the approach of generalized pairwise comparisons (GPC) might be an alternative way of analyzing composite endpoints. Potential advantages of the GPC include a hierarchical structure that account for relative priorities of the components, the flexibility of incorporating different types of component outcomes, and no need for strong assumptions as are needed for parametric methods. The objective of this work is to assess the statistical performances of the three above-mentioned approaches using real data and simulations.

Methods: The three analysis approaches were applied to data from a randomized placebo-controlled clinical trial. The composite endpoint included all-cause death, disease-related hospitalizations, and initiation of a specific rescue therapy. Change in risk score from baseline to month 12 in a continuous scale was added as the lowest priority component in the GPC analyses. The power performance was investigated using simulations under various scenarios (different component endpoints, hierarchical orders, magnitudes of treatment effect).

Results: For the given trial data, while the Cox model and the GPC analysis did not show statistically significant treatment effects, the LWYY model did. The addition of the post-hoc continuous component increased the treatment effect, drastically reduced the percentages of ties, and resulted in a statistically significant treatment effect using GPC. In the scenarios explored by simulations, the highest power was obtained from the LWYY model, followed by the GPC method and the Cox model. When adding a continuous component to the GPC analyses, treatment effects increased, percentages of ties decreased, and power increased.

Conclusion: In the considered scenarios, the GPC method showed higher power than the Cox model. Incorporating an additional continuous component further increased the power of the GPC method. However, the LWYY model still showed the highest power.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-06 CASc winner

Interval estimation of median odds ratio for measuring contextual effects in multilevel data using binary logistic model

Shafayet Khan Shafee[1,2], M. Shafiqur Rahman Rahman[1]
1 Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka, Bangladesh
2 Pathao Limited, Dhaka, Bangladesh

Introduction: In multilevel data, where individuals are nested within clusters, one of the important concerns is to assess clustering effects (often termed contextual effects) for making cluster-level inferences. Multilevel modelling (MLM) is a widely used approach that incorporates contextual effects as a part of the model component. Variance partition coefficient (VPC) is commonly used for measuring contextual effects in MLM with continuous outcomes, however, it is not straightforward to estimate VPC for the models with binary outcomes. As an alternative, the median-oddsratio (MOR) - the median relative change in odds of having the outcomes when comparing identical subjects from two randomly selected different clusters that are ordered by risk, offers a more intuitive interpretation of contextual effects that facilitates direct comparison with the fixed effects. The existing literature discussed the point estimate of MOR only for the two-level binary logistic model, however, an interval estimate is necessary for inferential statistics.

Methods: This study discusses the interval estimation of MOR for a two-level binary logistic model and extends the method to the three-level logistic model, for evaluating two different contextual effects. Simulations were conducted to evaluate the performance of MOR estimates under different data structures. Further, the methods are applied to estimate the contextual effects of both administrative district (3?? level) and enumeration area (2?? level) on having access to internet by a woman (1?? level) by using data from the database of the 2019 Multiple Indicator Cluster Survey (MICS).

Results: Simulation results show that the percentage of relative bias of MOR estimate decreased with increasing cluster size and number of clusters. For those scenarios, the confidence interval of the MOR provided approximately equal 1 coverage to the 95% true nominal coverage. The results also suggest that the small-sample bias induced by the MOR is due to the bias in the estimate of the variance of the random effect of the model. The application findings show that both district and EA have significant contextual effects on women's having internet access.

Conclusion: This study discussed the interval estimation of MOR with an extension of the methods to the three-level models. The application shows that the estimate of the MOR has an intuitive interpretation, and the methods can be applied to assess the contextual effect of hospitals and doctors on patients' management in multi-center studies.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-07

Graphical inference tests revisited: Using two-dimensional grids in nonparametric testing for hypothesis evaluation and educational insights

Lubomir Stepanek[1,2], Filip Habarta[2], Ivana Malá[2], Luboš Marek[2]
1 Department of Biomedical Statistics, Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, Prague, Czech Republic
2 Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

Background/Introduction: Nonparametric tests often rely on straightforward concepts like ranking observations or dichotomous refinement of pre-post changes. These tests, such as the Mann-Whitney test or signed-rank tests, offer numerical precision but can also be interpreted graphically. While graphical techniques can't fully substitute numerical performance, they aid in comprehending the test logic and may lead to practical rules of thumb formulation.

Methods: In this work, we revisit the graphical inference testing for selected nonparametric tests, both two-sample and paired ones. The graphical testing logic involves transforming a test statistic construction into (orthogonal) directional changes on a two-dimensional finite-step grid. Also, the graphical pathway creation depends on what the test statistics focus on the observations. While the mutual distribution of both the samples' rankings and sample affiliation's changes as counted from the lowest to the highest rank matter in the case of two-sample tests, the sequence of either positive or negative pre-post changes plays a role in paired tests. The changes are plotted as unit steps in two orthogonal directions on the grid. Usually, a graphical representation of a null hypothesis claiming no difference between the sample's observed data and a constant value or other sample's data is determined by almost regular alternating of both grid directions. Also, directions of the graphical pathways of a given inference test instance on a given two-dimensional grid follow the binomial distribution and, consequently, could be depicted within a probability framework. As an aspect of novelty and our contribution, we apply Popoviciu's inequality to derive an upper bound of a probability of getting observations contradicting the null hypothesis in the same or even more substantial way, i.e., the p-value, which may provide an insight into statistical power of such an inference test.

Results: R language functionality was developed for computations and two-dimensional plotting of grids used in graphical inference testing. Regions within the grids were highlighted for typical situations such as null hypothesis rejection. Test grids can reflect asymmetric null hypotheses for the signed-rank test by upper-bounding the maximum "traffic" for a given direction. Various simulations were conducted, considering different sample pairs or pre vs. post situations.

Conclusion: Graphical inference testing enables a better understanding of the nonparametric hypothesis test logic or could serve education purposes. The R functionality for graphical testing will soon be melted into an R package.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-08 Prediction of language growth scores from a two-stage population survey with longitudinal data

George Vamvakas, Courtenay Norbury, Andrew Pickles
King's College London, United Kingdom

Background: An important objective in studies of children's language growth is to predict individual outcomes of assessments from expressive and receptive language modalities. Describing the uncertainty of the predicted outcomes is crucial in demonstrating whether a child with developmental language disorder is making greater-than-expected progress with implications for intervention. Age-referenced predictions and prediction intervals for language development from a large population of children with various language abilities have not been previously described.

Methods: The work was motivated by SCALES, a two-stage UK population survey of children (N=7,267) who were followed through their primary school years. The survey consisted of an initial screening phase and a subsequent longitudinal component of in-depth assessments. Predictions were sought for the in-depth assessments of a child for whom at least one initial measurement existed. Point predictions and 95% prediction intervals were generated using population estimates from a structural equation model that combined partially observed Normally distributed repeated measurements with fully observed count data from the screening phase. The count data were used as auxiliary variables to correct for biases when data were missing by design and to improve the estimates' efficiency. We explored the statistical behaviour of the point predictions and the prediction intervals with simulation and application to real data.

Results: With predictions further into the future, the prediction uncertainty increased. Utilising more assessments for the child informed the predicted trajectory more clearly and increased the prediction's precision. When only screening information was available, using a few degrees of freedom for the centile in the prediction interval was needed. When Normally distributed information was also available, Normal centiles should be used. We generated predicted trajectories of language outcomes for individual children and characterised their uncertainty within a 95% prediction envelope. We plotted the trajectories against growth charts built from our model's estimates to express the ranking of the child's ability against age referenced population norms.

Conclusion: Our approach applies to situations when data are missing at random and analysts seek to use auxiliary variables with maximum-likelihood-based models to make a prediction for a future outcome and describe its uncertainty. Our methods are unique in their ability to assess whether a child's progress is following the expected trajectory, whether the child is catching up with peers, or whether progress has plateaued. This research has potential applicability to many developmental outcomes typically collected in psychopathology.

POSTER SESSION B-06:
Statistical Analysis of Complex Data Structures

P-B06-09

Estimation and application of derivative multivariate functional principal component analysis

Yueyun Zhu[1], Andrew Simpkin[1], Norma Bargary[2], Steven Golovkine[2]
1 University of Galway, Galway, Ireland
2 University of Limerick, Limerick, Ireland

Dai and Müller (2018) proposed a Derivative Functional Principal Component Analysis (DFPCA) approach to recover the derivatives of univariate functional data, by estimating the derivative functional principal components (DFPCs) and the corresponding scores. Happ and Greven (2018) proposed a Multivariate Functional Principal Component Analysis (MFPCA) approach for multivariate functional data. Following their studies, we aim to develop a Derivative Multivariate Functional Principal Component Analysis (DMFPCA) approach which estimates the derivatives of multivariate functional data. The DMFPCA approach starts from applying Mercer's theorem to the covariance of derivatives of multivariate functional data. The eigenvalues of the covariance represent the amount of variability in the derivatives explained by the derivative multivariate functional principal components (DMFPCs), while the derivative multivariate functional principal component scores (DMFPC-scores) serve as weights of DMFPCs in the multivariate Karhunen-Loève expansion of the derivatives. We establish a relationship between DFPCA and DMFPCA, which offers an estimation strategy to calculate DMFPCs and DMFPC-scores based on their univariate counterparts. The asymptotic properties for the proposed estimates are derived. Moreover, an alternative approach, which directly takes the derivative of multivariate Karhunen- Loève expansion of the original functional data, is compared with DMFPCA in simulation. It emerges that DMFPCA provides more accurate recovered derivatives and a more reasonable explanation for understanding data dynamics. Analysing the dynamics of functional data through their derivatives is of interest in biostatistics. For example in cardiology, angiograms provide two functional variables - the quantitative flow ratio (QFR) and diameter - measured across the length of a vessel. The dynamics of these functional data are of key interest to assess coronary artery disease. However, there are currently no methods to deal with the derivatives of multivariate functional data. The DMFPCA we propose extends the existing methods for estimating the derivatives of univariate functional data to multivariate cases. By applying DMFPCA to the angiogram data, it allows us to explore the joint dynamics between QFR and diameter. It turns out that DMFPCA effectively captures important modes of variation in the change of diameters and the velocity of QFR, which helps to assess coronary artery disease.

POSTER SESSION B-07:
Validation of Synthetic Data

P-B07-01 Calibrating representations of expert knowledge with patient data in latent spaces for synthetic trajectories

Hanning Yang[1], Meropi Karakioulaki[2], Cristina Has[2], Moritz Hess[1], Harald Binder[1]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany
2 Department of Dermatology, Medical Faculty and Medical Center, University of Freiburg, Freiburg, Germany

Background: In longitudinal clinical cohorts, e.g., on the natural history of rare diseases like Epidermolysis Bullosa (EB), a range of interrelated measures has to be taken into account, such as blisters, wounds, inflammation, anaemia, and body growth. Modelling patient trajectories is often hampered by few observations and complex missingness patterns in individual patient data (IPD). Therefore, representing expert knowledge from the scientific literature via quantitative models could be useful for augmenting IPD by synthetic data. However, this still requires some calibration with real data. We here utilize ordinary differential equations (ODEs) to represent knowledge on the natural history of EB, specifically simulating biomarker evolution over time.

Methods: Based on expert knowledge from a systematic literature search, we constructed a system of ODEs, aiming for an accurate modelling of the dynamics of critical biomarkers such as C-Reactive Protein (indicative of inflammation levels), haemoglobin (for anaemia), and variables related to growth in EB. Calibration of this ODE system with real data is hindered by the relatively large number of markers, and therefore we use an autoencoder, i.e., a neural network-based approach, for dimension reduction. We consider different variants of how to quantify similarity between the synthetic data from the ODE and the real data in the latent space of the autoencoder to obtain gradients for calibrating the ODE parameters. In addition, we vary the amount of noise that is put on top of the ODE solution to assess the impact on calibration. Similarly, we assess the effect of different missingness patterns in the real data.

Results: We find that the proposed approach allows to calibrate our rather large ODE system. However, we encounter limitations with our approach when calibrating in two-dimensional latent space, whereas three-dimensional latent space provides a good visual match between real and synthetic data. The results even hold up when introducing a large amount of noise and challenging missingness patterns. This is backed up by investigations of the gradients which indicates stability in the results.

Conclusion: We demonstrate how the challenge of calibrating a complex expert-informed synthetic data model can be addressed by investigating similarity of real and generated observations in a low dimensional latent space as learned by a neural network. This enables naturalistic synthetic IPD while having control over the generative process, e.g., for potentially synthesizing realistic control groups.

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-01 Estimating treatment effects when there are competing risks using real-world data: Application to prostate cancer

Caroline Chesang, Linda Sharples, Thomas Cowling, Ruth Keogh
London School of Hygiene and Tropical Medicine, London, United Kingdom

Background: Cause-specific and sub-distribution hazard ratios are commonly used treatment effect measures when there are competing risks, but recent statistical literature has shown that they lack a clear causal interpretation and has recommended alternative estimands. This challenge has been addressed in the context of RCT's, but extensions are required to estimate these estimands using observational data in the presence of confounding and time-varying treatment initiation. Motivated by a question about treatment effects for prostate cancer, our objective was to define appropriate causal estimands and demonstrate their estimation using a clone-censor-weight (CCW) approach and an extension based on landmarking. These methods are applied to emulate a target trial (PR07) comparing the effects of RT+HT (radiotherapy initiated at any time within 8-weeks of randomisation in addition to hormone therapy) and HT (hormone therapy alone), on disease specific survival for men with high risk prostate cancer using UK cancer data.

Methods: We developed a protocol for emulating a target trial, closely related to the PR07 trial, using UK cancer data. The estimand of interest was the risk of death due to prostate cancer up to 7 years without elimination of competing events. In PR07, the timing of treatment initiation relative to diagnosis was not specified. In the emulated trial we used diagnosis as the time origin. As few patients initiated radiotherapy within 8 weeks of diagnosis in the observational data, we considered emulated trials in which the grace period for initiating radiotherapy ranged from 4-6 months and used CCW for estimation. We also used the landmarking approach in combination with CCW to consider different 'time zeros' after diagnosis, and to accommodate a shorter grace period as in the trial.

Results: The PR07 trial reported cumulative incidence functions (CIFs) at 7 years of 9% (RT+HT) and 19% (HT) for prostate cancer deaths versus 17% (RT+HT) and 16% (HT) for non-prostate cancer deaths. Corresponding results from our CCW analysis using a grace period of 6 months gave corresponding CIFs of 5.9% (RT+HT) and 13.2% (HT) for prostate cancer deaths and 14.2% (RT+HT) and 19.8% (HT) for non-prostate cancer deaths. Landmarking CIFs will be presented.

Conclusions: Overall, these methods provided similar all-cause and prostate cancer CIF estimates to PR07 at 7 years. Differences in other deaths CIFs could be attributed to patient selection, treatment being protective, unmeasured confounding, and improvements in prostate cancer care since PR07.

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-02 Penalisation methods for multi-state models: A comparative, simulation study

Chantelle Cornett[1], Niels Peek[2], Alexander Pate[1], Glen Martin[1]
1 University of Manchester, Manchester, United Kingdom
2 University of Cambridge, Cambridge, United Kingdom

Background/Introduction: Overfitting presents a significant challenge in the development of accurate multi-state models, especially with complex datasets with many transitions. This study aims to investigate the efficacy of various penalization methods in mitigating overfitting within Weibull transition-specific models.

Methods: We performed a simulation study, wherein we generated simulated data for n=500,000 individuals, incorporating three baseline characteristics (a categorical variable, a binary variable, and a continuous variable) and simulating time spent in each state for both three-state and five-state multi-state models. Each of the three- and five-state scenarios had one rare transition, incorporating small sample size for one transition to increase the risk of overfitting. These data were split randomly into a development (n=300000) and validation (n=200000) set. To the development set, we fit the following models: unpenalized transition-specific Weibull models estimated under maximum likelihood, a MSM fit under a global shrinkage factor (i.e., applying the heuristic shrinkage factor to the coefficients of each transition model estimated under maximum likelihood), a MSM fit under LASSO penalization, and a MSM fit under a Bayesian framework with a horseshoe prior. We also fit an (unpenalized) MSM to a state-collapsed version of the five-state model (i.e., where two states are 'combined' into one), to investigate this as an alternative to penalisation methods. The predictive performance of all models was then estimated within the validation set, quantified through the C-indexes, transition-specific calibration plots, and time-dependent ROC curves. The mitigation of overfitting for the penalized models was compared to the unpenalized models, serving as a baseline, by comparing the transition-specific calibration slopes as a quantification for the level of overfitting.

Results: At the time of writing this abstract, the simulation is running and should be completed within the next few weeks. Results will be ready in time for the conference. I will present the comparison in overfitting levels between unpenalized and penalised models and discuss the relative differences in model predictive performance. I will present some guidance on choosing an appropriate penalization method, given its potential impact on model accuracy and predictability.

Conclusion: This research underscores the importance of penalization in enhancing model accuracy and provides a foundation for future empirical work to validate these methods across various datasets and model complexities. Future research will expand on these findings, exploring the impact of study characteristics such as sample size and number of parameters, and incorporating cross-validation to further validate the selected penalization methods.

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-03

A comparison between the joint model and landmark approach for dynamic prediction of competing risks in survival analysis with time-dependent covariates

Ömer Faruk Dadaş, Timur Köse
Department of Biostatistics and Medical Informatics, Ege University Faculty of Medicine, Izmir, Turkey

Background: / Introduction: In survival analysis, it is important to carefully consider competing risks when dealing with multiple events. Health studies often use time-dependent measurements and data related to the progression of the disease to monitor the illness. In the management of chronic diseases, it is crucial to dynamically predict competing risks using patient data. Dynamic predictions help to understand how a patient's risk of recurrence or mortality changes over time. Two common approaches for dynamically predicting competing risks are the Joint model and the Landmark approach. Given the frequent occurrence of longitudinal data and competing risks in health studies, it is important to identify the appropriate analysis method.

Methods: To achieve this, an investigation was conducted to determine if there were any differences in patient-specific risk predictions between the two methods and if sample size affected the results. The Joint model and the Landmark approach were used to dynamically predict cumulative incidence probabilities of competing risks. These predictions were then compared across different datasets with varying sample sizes, including a real dataset and a simulation study. All analyses were performed using the R statistical software package.

Results: When comparing all individuals, it was observed that the Landmark approach did not yield similar results to the Joint model at any landmark times. Furthermore, when comparing the computation time for dynamic prediction, it was determined that the Landmark approach had a shorter time.

Conclusion: It has been determined that the Landmark approach cannot be used as an alternative method to the Joint model in general.

References: *[1] Elashoff, R. M., Li, G. & Li, N. (2008). A Joint Model for Longitudinal Measurements and Survival Data in the Presence of Multiple Failure Types. Biometrics, 64(3), 762–771. doi:10.1111/j.1541-0420.2007.00952.x. [2] Rizopoulos, D., Molenberghs, G. & Lesaffre, E. M. E. H. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biometrical Journal, 59(6), 1261–1276. doi:10.1002/bimj.201600238*

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-04 An R function for data preparation for an acyclic multistate model with non-ordered intermediate states

Lorenzo Del Castello[1], Davide Paolo Bernasconi[1,2], Laura Antolini[1], Maria Grazia Valsecchi[1,3]
1 University of Milan-Bicocca, Milan, Italy
2 ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy
3 Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy

Background: The package "mstate" is a popular R library to perform multistate analysis [1]. The function "msprep" provides a simple tool to handle data preparation where a wide-format dataset (one row per patient) must be transformed in long-format (one row for each possible transition). However, "msprep" allows only for triangular transition matrices corresponding to irreversible acyclic Markov processes. Sometimes this is too restrictive. Hazard et Al. [2] used multistate models to analyze care pathways of patients admitted to ICU for severe COVID-19. Patients were allowed to move back and forth between non-absorbing states. Thus, the authors developed a dedicated R function for data preparation which overcomes the limits of "msprep".

Methods: A further complexity, not present in the application by Hazard, could be the case of an acyclic model with no defined order between intermediated states. In our application, patients affected by cardiogenic shock are admitted to ICU and then may be treated at different times with none, one, two or three mechanical circulatory supports, namely IABP, Impella or ECMO. A patient may be treated with any of these devices and there is no specific order between treatments. However, no patient can be treated twice with the same device (e.g., after IABP one could receive another device but then cannot go back to IABP). The final absorbing state is a composite endpoint (first event between LVAD implantation, heart transplant or death). In this context, transition to a certain intermediate state blocks the possibility of future back-transitions to that state.

Results: We developed an R function allowing for data preparation in this situation. The proposed function transforms a dataset from wide to long format that can be passed to the function "msfit". The new dataset will be prepared in order to cause the contribution to the risk sets in time to be consistent with the dynamic update of possible transitions. This is achieved through the definition of a new matrix, called "blockmat", in which the paths that cannot be observed are declared blocked. Without the specification of the blockmat, some subjects would be considered at risk of presenting also transitions which they can never travel. In practice, this would result in the presence of additional unnecessary rows in the long dataset that will cause an underestimation of transition hazards.

Conclusion: We created an R function allowing for data preparation in case of an acyclic multistate model with non-ordered intermediate states.

References: *[1] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007 May 20;26(11):2389-430. doi: 10.1002/sim.2712. PMID: 17031868. [2] Hazard D, Kaier K, von Cube M, Grodd M, Bugiera L, Lambert J, Wolkewitz M. Joint analysis of duration of ventilation, length of intensive care, and mortality of COVID-19 patients: a multistate approach. BMC Med Res Methodol. 2020 Aug 11;20(1):206. doi: 10.1186/*

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-05

Enhancing healthcare understanding from clinical routine data by simplifying the representation of treatment pathways

Maryam Farhadizadeh[1], Zoe Lange[2], August Sigle[3], Holger Dette[2], Harald Binder[1], Nadine Binder[4]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
2 Department of Mathematics, Ruhr University Bochum, Bochum, Germany
3 Department of Urology, Faculty of Medicine, Medical Center, University of Freiburg, Freiburg, Germany
4 Institute of General Practice/Family Medicine, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Background: In clinical routine care data, patients' pathways often exhibit heterogeneity, encompassing the order in which a wide variety of interventions occur in addition to varying sequence length. This presents a challenge in understanding and managing healthcare paths effectively, e.g., when using multistate models, which can only deal with a limited number of states. Therefore, there's a critical need to identify and group similar pathways within this diverse landscape [1, 2].

Methods: To identify typical treatment pathways for inpatient stays, we have developed an algorithmic solution that uses coded clinical data, i.e., diagnoses and procedures. This approach helps to visually represent the data in treatment path diagrams. Specifically, the algorithm detects important procedures by calculating their importance based on patient counts and node significance, comparing them to predefined thresholds. Additionally, the algorithm groups less important procedures to put the focus on essential components of these pathways.

Results: We demonstrate our algorithm using clinical routine data from prostate cancer patients receiving radical prostatectomy in the Department of Urology at the Medical Center, University of Freiburg. To further explore the efficacy of our approach in simplifying the representation of treatment pathways, we evaluate it through an extensive simulation study. This involves varying pathway similarities, i.e., different levels of heterogeneity of interventions, and sequence lengths. We find that a representation with a manageable number of typical pathways can be obtained, and characterize the sensitivity with respect to different tuning parameters of our algorithm.

Conclusion: Our proposed method simplifies identification and visualization of common healthcare trajectories, aiding in understanding patient paths and informing healthcare decisions. Improving our understanding of these pathways can elevate clinical care standards and enhance health outcomes, in particular by enabling subsequent analysis with multistate models and statistical tests.

References: *[1] Binder, Möllenhoff, Sigle, Dette. Similarity of competing risks models with constant intensities in an application to clinical healthcare pathways involving prostate cancer surgery. StatMed 2022 [2] Möllenhoff, Binder, Dette. Testing similarity of parametric competing risks models for identifying potentially similar pathways in healthcare. arXiv:2401.04490v1, 2024*

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-06 Balancing speed and detail: Assessing multi-state prediction models in the face of emerging pathogens

Derek Hazard[1], Oksana Martinuka[1], Hamid Reza Marateb[2], Marjan Mansourian[2,3], Philipp Mathé[1], Siegbert Rieg[1], Manuel Rubio-Rivas[4], Martin Wolkewitz[1]

1 Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany
2 Biomedical Engineering Research Centre (CREB), Automatic Control Department (ESAII), Universitat Politècnica de Catalunya-Barcelona Tech (UPC), Barcelona, Spain
3 Department of Epidemiology and Biostatistics, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran
4 Department of Internal Medicine, Bellvitge University Hospital, Bellvitge Biomedical Research Institute-IDIBELL, University of Barcelona, Barcelona, Spain

A fundamental challenge in providing sufficient preparation for pandemics or emerging pathogens is the rapid generation of valid evidence, preferably in real-time. In a rapidly changing environment, the challenge is developing "forward-looking" strategies that are not based too heavily on "backwards-looking" data. This issue applies prominently in clinical epidemiology in predicting the clinical course of hospitalized patients. The importance is manifold – from flagging changes in the severity of a disease to proper estimation of medical resource needs. Many approaches focus on time-to-event analyses and we have applied multi-state methodology to this unique setting.

We simulate a multi-state model, inspired by a scenario with two transient states (regular ward, intensive care unit (ICU)) and two absorbing states (discharge alive, in-hospital death). Our objective is to compare two prediction approaches. Firstly, modelling transitions among the states non-parametrically utilising an Aalen-Johansen estimator with adjustment for covariates at baseline. Secondly, modelling transitions among the states using time-constant hazards derived from straightforward formulas. Our goal is to investigate the extent of improvement in prediction using individual-level data versus aggregate, routinely collected data. The performance of the models in terms of predicted state occupancy and transition probabilities will be assessed by calculating prediction errors using Brier scores and Kullback–Leibler divergences. Through the simulation, we aim to identify under which circumstances the more detailed model surpasses the simpler model in performance. Additionally, we will simulate the impact of calendar time to reflect changing conditions in a pandemic (new treatments, variants, vaccinations, etc.). We will explore methods to capture this dynamism, including fractional polynomials, splines, time categories, and time weights.

The methodology will be demonstrated with prediction models trained on data collected from hospitalized COVID-19 patients from the Bellvitge University Hospital in Barcelona, Spain during the first year of the pandemic. Covariates collected at admission included age, sex, body mass index, oxygen saturation, and lymphopenia. The models will be externally validated on data from COVID-19 admissions to the University Medical Center in Freiburg, Germany from January 2020 to December 2021. Innovative visualization techniques for external validation will be presented and demonstrated in a Shiny application.

Routinely collected data significantly reduces the burden of gathering evidence in dynamic circumstances. Moreover, aggregate data navigates around data protection and privacy challenges. These elements facilitate rapid research when urgency is paramount. These examples illustrate the situations in which a compromise in performance offsets these benefits.

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-07 Marginal analysis of time spent hospitalised or exacerbating

Jonatan Hedberg
AstraZeneca, Mölndal, Sweden

Introduction: The analysis of recurrent events is common in RCTs when assessing hospitalisations or exacerbations. The number of days patients spend in these states can however be more informative for payers and patients alike, aligning better to cost and quality of life. We investigate two methods to analyse the rate ratio of event days with incomplete follow-up. Based on inference theory, Poisson with sandwich SE should yield asymptotically unbiased results with correct variance estimates [1]. LWYY regression for recurrent events [2], can be proven to also work for any endpoint that is increasing over time, e.g. cumulative event days.

Methods: We evaluate small sample bias and type-I error by simulating continuous time multi-state data with a healthy state and an event state. Constant transition intensities were assumed with subject specific random effects from a gamma distribution. In total 48 scenarios were run with 10,000 iterations each representing RCTs of 365 days. Intensities of 0.001, 0.005, and 0.01 (0.365, 1.825, 3.65 per year), and 0.5 and 0.1 (duration of 2 and 10 days) were used for the transitions to the event state and healthy state respectively. The variance of the random effect varied between 1 and 5, with the random effect only acting on the transitions to the event state in half of the scenarios. For each iteration, Poisson regression with robust SE and LWYY regression were run.

Results: Both regression models yielded unbiased estimates of the rate ratio. LWYY regression had type-I error close to 0.05 for all scenarios ranging from 0.043 to 0.056. Poisson regression generally had higher type-I error ranging from 0.048 to 0.089. Scenarios with the lowest event intensity, and the random effect acting on both transitions resulted in an inflated type-I error. For the larger sample size coupled with higher intensities, the type-I error were similar.

Conclusions: When sample sizes are small, events are scarce, and the between patient variability is high, LWYY outperforms Poisson with robust SE for analysing the cumulative event days even under constant intensities in a Markovian multistate model.

References: *[1] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50(1), 1–25. [2] Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 62(4), 711–730.*

POSTER SESSION B-08:
Competing Risks and Multistate Models

P-B08-08

The challenge of time-to-event analysis for multiple events: A guidance

Sandra Schmeller[1], Alexandra Erdmann[1], Jan Beyersmann[1], Ann-Kathrin Ozga[2]
1 Ulm University, Ulm, Germany
2 University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Background: Clinical trials often compare a treatment to a control group with respect to multiple possible combined time-to-event endpoints like time to hospitalization and time to death. Thereby, the first endpoint may occur more than once (recurrent), whereas the second endpoint is absorbing, that is, after observing the second endpoint an individual can no longer experience the first event. However, usually only the time until the first occurrence of an event for a patient is analyzed. Inclusion of all observed events in the analysis can increase the power and provides a more complete picture of the disease but needs more sophisticated methodology.

Methods: We give a stepwise guidance on how to extend the simple time-to-first event model to complex multistate methodology, where multiple events are incorporated. We thereby consider nonparametric methodology and semiparametric methods and show how they are related. Advantages and disadvantages of the different methods will be introduced and explained on data from the Interdisciplinary Network for Heart Failure, a multi-center randomized controlled trial that investigated the efficacy of a nurse-coordinated disease management program for patients that were first hospitalized for systolic heart failure. Special attention is given on the prerequisites of the models, e.g. the Markov property, and their interpretation. A test for the Markov property is investigated within a simulation study and extensions based on novel non Markov results are suggested. The aim is to give an overview of existing methods, present the assumptions, especially discussing the Markov assumption, and elaborate the difference in interpretation, especially between hazard and rate models.

Results: To quantify a treatment effect in the data from the Interdisciplinary Network for Heart Failure we found that marginal summary measures are easier to interpret than rate models. The treatment group has a shorter time spend in hospital, less number of hospitalizations and a slightly lower empirical probability of death.

Conclusion: We provide a guidance to the researcher for the analysis of multiple time-to-event data.

POSTER SESSION B-09:
Integrative Data Analysis

P-B09-01

Integrated analysis of humoral immune response across several COVID immunisation schemes in COVIDAuRA study

Camilo Broc[1], Jérémie Becker[1], Benoît Levast[1], Carla Saade[2], Inès VuDuc[3], Lucie Saint-Joanis[1], Kendra Reynaud[1], Charlotte Mignon[1], Natacha Mariano[1], Stéphanie Donnat[1], Shin-Yi Yu[1], Joséphine Abi Ghanem[1], Viet-Dung Tran[1], Oxana Svratskikh[1], Céline Couturier[1], Anely Tranchot[1], RICO study group, COVIDSER study group, Karen Louis[1], Laurent Beloeil[1], Cyril Guyard[1], Stéphane Paul[4], Sophie Trouillet-Assant[2], Arnaud Marchant[3]

1 Bioaster, Lyon, France
2 HCL, Hospices civils de Lyon, Lyon, France
3 ULB, Université Libre de Belgique, Brussels, Belgium
4 INSERM, U1111, CNRS, UMR 5308, CIRI– GIMAP, UCBL 1, UJM, Immunology and Immuno-monitoring Laboratory, iBiothera, CIC 1408, Saint-Étienne, France

Background/Introduction: The emergence of multiple SARS-CoV-2 variants as well as the implementation of different vaccination strategies gave rise to multiple immunization schemes in the population and related questions 1) which immunization strategy leads to the most effective long-term immune response against SARS-CoV-2? 2) What is the impact of the emergence of new variants?

Methods: To this end, the COVIDAuRA study aimed to characterize the humoral response following different combination of vaccine and COVID-19 infection using multiple technologies. More specifically, we investigated a set of 180 volunteers enrolled in the CovidSer cohort according to their immunization profile: i) convalescent patients infected during the first wave of the pandemic. ii) convalescent patients who subsequently received one or two dose(s) of ChadOx1 or BNT162b2 mRNA vaccine. iii) COVID-19 naïve individuals fully vaccinated with two or three doses of ChadOx1/BNT162b2. iv) individuals vaccinated with two or three doses of BNT162b2 followed by a breakthrough infection during the Omicron BA.1 wave. This corresponds to a total of nine groups. The humoral immune response was monitored at 6 months post last immunization (i.e. infection or last vaccine dose) based on antibody titers, neutralization antibody assays, IgG glycosylation, antibody Dependant Complement Deposition (ADCD) assays, FcR binding assays and Bio-Layer Interferometry (BLI) avidity. While a single antigen was used for BLI and glycosylation 3 and 8 were tested in neutralization and ADCD, FcR binding assays respectively. To compare the nine groups, multivariate analysis and breadth of coverage across variants were computed.

Results: An integrated multivariate analysis led to the identification of three main immune profiles associated with different breadth of coverages. Although these profiles were correlated with the number of immunizations, the clinical intervention explained a larger fraction of variation. A subset of neutralization, FcR and ADCD parameters were found highly discriminant across these three profiles. Due to the enrolment of the nine groups at different time periods, early and late variants were analysed separately. Using a nested ANOVA, the data further showed that these nine groups could be grouped according to their clinical interventions.

Conclusion: Beyond vaccination impact, the analysis also compares the added value of each technology to answer the biological question. BLI technology and specific antibody related assays are promising for humoral response studies.

POSTER SESSION B-09:
Integrative Data Analysis

P-B09-02

Child malnutrition at Upazila Level in Bangladesh: A small-area estimation approach

Md Mynul Islam, Syed Shahadat Hossain
Institute of Statistical Research and Training, University of Dhaka, Bangladesh, Dhaka, Bangladesh

Background: The importance of promoting good nutrition to eliminate all forms of malnutrition is clearly mentioned in SDG 2. Bangladesh is one of many countries that have made significant progress in reducing malnutrition [1]. The dramatic progress made by Bangladesh means that the country must strengthen its efforts to achieve SDG 2. For effective distribution of food aid, it is useful to assess malnutrition indicators at a geographical level so that unusually high levels of malnutrition can be targeted. The objective of this study is to estimate the prevalence of stunting, wasting and underweight among children under five years at the Upazila level and prepare a map to assist in planning development assistance programs.

Method: This study uses the small area estimation technique (ELL method) developed by Elbers, Lanjouw and Lanjouw in 2003. This study uses the 2019 Multiple Indicator Survey [2] and the 2011 Population and Housing Census of Bangladesh (10%).

Results: The small estimates generated in this report have a low standard error up to the Upazila level. Department and district level estimates are consistent with direct estimates from MICS 2019. However, some exceptions are noted. In case of small area estimate using the developed model, the district estimates of stunting, wasting and underweight for the two or three districts are found to be over-estimated or under-estimated.

Conclusion: It is believed that this study will help decision makers gain insight into Upazilalevel estimates of under-five malnutrition.

References: *[1] Press release, Bangladesh sees sharp decline in child malnutrition, while violent disciplining of children rises, new survey reveals UNICEF, 2020, https://www.unicef.org/bangladesh/en/press-releases/bangladesh-sees-sharpdecline- child-malnutrition-while-violent-disciplining-children [2] Bangladesh Bureau of Statistics (BBS) and UNICEF Bangladesh. 2019. Progotir Pathey, Bangladesh Multiple Indicator Cluster Survey 2019, Survey Findings Report. Dhaka, Bangladesh.*

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-01 Semi-Bayesian methods to compute standard errors in nonlinear mixed effects models for sparse data

Mélanie Guhl[1], Lucie Fayette[1], Julie Bertrand[1], Emmanuelle Comets[1,2]
1 Université Paris Cité, INSERM, IAME, F-75018 Paris, France, Paris, France
2 Univ Rennes, INSERM, EHESP, IRSET - UMR_S 1085, F-35000 Rennes, France, Rennes, France

Introduction: Nonlinear mixed effects models (NLMEM) are used to handle longitudinal data to describe disease progression and/or treatment response. Parameter estimates are often obtained using maximum likelihood estimation (MLE), with their uncertainty estimated via the Fisher Information Matrix (FIM). With limited sample size (N) and sparse design (n), the FIM can underestimate the uncertainty [1], and sampling from the distribution of the MLE with Bayesian inference has shown improved results in this case [2]. In the present work, we propose to embed semi-Bayesian inference in the R package saemix, implementing the stochastic approximation expectation maximization (SAEM algorithm).

Methods: We implemented two semi-Bayesian approaches, one using a Metropolis-Hastings algorithm (SAEM MH) [3] and one using Approximate Bayesian Computation (SAEM ABC), testing different variations. We compared them with frequentist methods (FIM, sampling importance resampling and bootstrap) and a method using a Hamiltonian Monte Carlo algorithm implemented in the stan software (Post). We compared coverage rates and relative standard errors (SE) on a simulation study of 500 pharmacokinetic (PK) datasets (N=12, n=3) in two scenarios, one being more complex (high inter-individual variabilities and correlations between random effects), and on real PK data.

Results: SAEM MH showed good performances until confronted to complex variability structures where acceptation rates dropped near zero rendering the method unreliable to estimate uncertainty. Using block-sampling and random walk improved the acceptation rates, but SE and coverage rates were still downward biased. Post was also challenged by random effect correlations, while SAEM ABC showed improved results and was computationally efficient. On real PK data, all methods gave consistent results, e.g. RSE between 5 and 6% for the clearance suggesting asymptotic conditions were met. On a sparse subset of the data, discrepancies appeared between methods, e.g. RSE between 5 and 12% for clearance illustrating the need to identify a reliable method on sparse design studies.

Conclusion: Semi-Bayesian approaches implemented in the SAEM algorithm were promising to compute uncertainty at finite distance. Further calibration is needed.

References: *[1] Loingeville F et al., New Model–Based Bioequivalence Statistical Approaches for Pharmacokinetic Studies with Sparse Sampling, AAPS J, 2020. [2] Guhl M et al., Impact of model misspecification on model-based tests in PK studies with parallel design: real case and simulation studies, JPKPD, 2022. [3] Guhl M et al., Uncertainty computation at finite distance in nonlinear mixed models: evaluation of a new Bayesian method, Poster WP51, ISCB44, 2023.*

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-02

Bayesian mixed models approach to exploring resilience dynamics: Impact of stress on subjective health and affects over time during the COVID-19 pandemic

Markus Schepers[1], Irene Schmidtmann[1], Jun.-Sarah Schäfer[2,3], Julia Petersen[4], Nora Hettich-Damm[4], Daniela Zahn[5,6], Daniel Wollschläger[1]

1 Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
2 Leibniz Institute for Resilience Research, Mainz, Germany
3 Technische Universität Braunschweig, Braunschweig, Germany
4 Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
5 Preventive Cardiology and Medical Prevention, Center of Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
6 Health Sciences, Hochschule Fulda University of Applied Sciences, Fulda, Germany

Background: In the last decade, societies worldwide have faced a myriad of profound stressors, prominently exemplified by the COVID-19 pandemic (or natural disasters or wars). Due to the increasing stressor exposure, understanding the mechanisms that enable individuals to maintain their health and respond resiliently has gained popularity as a research goal in public health. Various conceptualizations and measures of resilience have emerged, reflecting the complexity of this construct. However, the precise methods and approaches to measure resilience in the context of longitudinal observational studies are still under ongoing refinement and active development.

Methods: Using longitudinal app survey data of the Gutenberg-COVID-19 study, we approach resilience dynamics with mixed models in a Bayesian framework for several ordinal response variables (analysed separately): self-assessed physical health, self-assessed mental health, and the self-reported affects of fear, sadness, anger and energy. The fixed effects are various kinds of individual stressors, and the random effects are the individual study participants. Furthermore, we consider the time series of these response variables, averaged over all app survey participants, with respect to the COVID-19 pandemic and its associated interventions. We describe the distribution and associations between the response variables and the individual-specific stressors.

Results: In total, there were 206 912 app survey responses from 7 386 study participants (age mean 55.09 years, 51.52% women) collected during the course of one year (between October 29, 2020 and October 25, 2021). Overall, subjective health and affects were negatively impacted during the lock-down periods but recovered to the previous level quite quickly thereafter. While the most common sources of stress were work-related, social stressors (i.e., loss of social contacts or conflicts with spouse or partner) had larger negative impacts on subjective health and affects.

Conclusion: Our longitudinal study design and mixed-model analysis support the negative impact of the COVID-19 containment measures on multidimensional health in the short-term, but a regain of mental health over longer periods, suggesting a certain degree of resilience to the pandemic overall. As we found that social stress was the most detrimental to the maintenance of subjective health and had the most negative effect on affect states, future research should examine the influence of protective social resources (e.g., social support) on the random-effects coefficients of the individual study participants.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-03 Longitudinal and joint models for PSA values and time to prostate cancer diagnosis

Birzhan Akynkozhayev, Benjamin Christoffersen, Keith Humphreys, Mark Clements
Karolinska Institutet, Stockholm, Sweden

Background: The prognostic significance of prostate-specific antigen (PSA) levels in the early detection of prostate cancer has been a subject of considerable debate. The study of the rate of change of PSA over time, as opposed to analysing single PSA values, has also yielded conflicting conclusions regarding its diagnostic properties for prostate cancer [1]. In this study, we aim to apply both longitudinal and joint models of longitudinal and time-to-event data to the Stockholm PSA and Biopsy Register, which includes over 2.4 million PSA measurements from more than half a million men. Our objective is to provide a comprehensive descriptive analysis of the dynamics of population-based PSA levels over time and to explore the relationship between trends in PSA levels and the time until prostate cancer diagnosis.

Methods: Ignoring cancer onset and censoring PSA values at cancer diagnosis, we have fitted a linear mixed-effects model with random intercept and random slope terms. The initial parameterisation for the random intercept centred at age 35 years, based on several established prostate cancer screening models. Additionally, we have recently developed a framework that utilises variational approximations for fast and efficient estimation of joint models, VAJointSurv, which is on the Comprehensive R Archive Network. In on-going work, we will use that framework to fit a joint model for longitudinal PSA values and time to prostate cancer diagnosis.

Results: For a random intercept centred at age 35 years, we observed a high correlation between intercept and slope terms when fitting the longitudinal model, which decreased significantly when we centred for the mean age of PSA testing. This finding indicates that the relationship between individual baseline PSA levels (intercept) and the rate of change in PSA levels over time (slope) is influenced by how age is incorporated into the model. Results from the joint modelling are forthcoming.

Conclusions: We provide a population-level description of PSA testing in Sweden. Our findings suggest that several prostate cancer simulation models should revise their PSA submodels. Current software copes well with the increasing quantities of rich longitudinal data, however software continues to be constrained for fitting joint models to larger datasets.

Reference: *[1] Javaeed, A., Ghauri, S. K., Ibrahim, A. and Doheim, M. F. (2020). Prostate-specific antigen velocity in diagnosis and prognosis of prostate cancer - a systematic review, Oncology Reviews 14(449): 64–71. URL: http://dx.doi.org/10.4081/oncol.2020.449*

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-04

A dynamic prediction model for predicting the time to conversion to AD in patients with MCI Based on time-dependent covariates

Yanjie Wang, Zheng Chen
Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China

Background: Alzheimer's Disease (AD) has become a pervasive international public health problem. AD is an irreversible neurodegenerative disease, which has imposed a heavy burden on the family and the society as the aging population continues to grow. Whereas mild cognitive impairment (MCI) is an intermediate state between normal aging and AD, we can alleviate this problem from the perspective of identifying and evaluating prognostic factors and predictive markers affecting the time for patients to convert from MCI to AD. However, most of the available studies are based on Cox models that need to satisfy the assumption of proportional risk, and HR is only a relative value, which does not intuitively predict how much time is left in the future for patients with MCI to convert to AD.

Methods: In this study, we constructed a new two-stage dynamic RMST prediction model based on the time scale indicator Restricted mean survival time (RMST), by combining the two-stage landmark method and the RMST regression model, which takes into account the measurement error of the longitudinal data, utilises more longitudinal information, and achieves the prediction of the mean time for the future transformation into AD for MCI patients at different time points.

Results: The simulation study showed that the regression coefficient estimates of the two-stage dynamic RMST model had a small bias with a coverage of about 95%, and the prediction performance was better than that of the static RMST model. In addition, individual dynamic prediction of time to AD conversion in MCI patients was achieved by analysing the ADNI dataset, and found that four variables, namely education, MMSE score, FAQ score and P-tau protein concentration, had a dynamic effect on the time. The mean C-index of the cross validation reached 0.8, and the mean prediction error of 1.9, all of which were better than the static RMST model.

Conclusion: Therefore, the proposed two-stage dynamic RMST prediction model has better prognostic and predictive performance, and can continuously update the length of mean time for MCI patients to turn into AD in the future according to the change of variables at different time points, which provides a scientific basis for doctors to monitor the patient's condition as well as evaluate the effect of interventions, improves the patient's quality of life, and reduces the burden on the society.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-05

Emulating hypothetical physical activity interventions to assess obesity risk reduction: Applying target trial emulation in the 1958 British Birth Cohort

Jamie Wong[1], Manuel Gomes[2], Michail Katsoulis[3]
1 Divison of Biosciences, UCL, London, United Kingdom
2 Department of Applied Health Research, Institute of Epidemiology & Health Care, UCL, London, United Kingdom
3 Population Science & Experimental Medicine, Institute of Cardiovascular Science, UCL, London, United Kingdom

Background/Introduction: Many studies examining the effect of physical activity on obesity have only used a single baseline measurement of physical activity in their analyses as their exposure. There is a limited number of papers that have evaluated the lifelong effect of physical activity on obesity measured at multiple timepoints, accounting for changes over time. This study aimed to use the target trial emulation framework to estimate the 32-year risk of obesity under hypothetical physical activity interventions from the National Child Development Study (NCDS).

Methods: This population-based cohort study utilised data from NCDS, a British birth cohort involving 13028 individuals born in 1958. Participants were allocated to one of two hypothetical interventions: low (0-3x per month) and high (1-5x per week) physical activity levels. Obesity was defined as a body mass index exceeding 30kg/m2. The study aimed to estimate the per-protocol effect, assessing the impact of physical activity if all participants had fully adhered to their assigned intervention. Multiple imputation was used to address missing data. Non-parametric bootstrap with 500 samples was used to estimate 95% confidence intervals, defined by the 2.5 and 97.5 percentiles of estimated risks. Inverse probability weighting was used to account for adherence, using information from time-fixed and time-dependent confounders. Pooled logistic regression models were utilised to estimate the standardised risk curves in 13,028 men and women with 32 years of follow-up (1981-2013), following individuals from age 23 to age 55.

Results: 7572 participants were assigned to the low physical activity (hypothetical) intervention, and 5456 to the high physical activity (hypothetical) intervention. The estimated standardised risk of obesity over 32 years for those who adhered to low physical activity was 52.58% (95% CI = 37.90% - 85.26%), and 39.37% (95% CI = 37.62% - 40.90%) for those who adhered to high physical activity. The risk difference between the two hypothetical interventions after 32 years was -13.21% (95% CI = -46.23% - 1.77%)

Conclusions: Individuals who adhered to a high level of physical activity exhibited a lower risk of developing obesity over 32 years of follow-up (between 23 to 55 years of age). This study was the first within the British birth cohorts, and among very few cohort studies overall, to utilise the target trial emulation framework to evaluate hypothetical interventions across the life course.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-06 Modelling longitudinal cognitive test data with ceiling effects and left skewness

Denitsa Grigorova, Dean Palejev, Ralitza Gueorguieva
Sofia University, Sofia, Bulgaria

Introduction: Cognitive tests such as the Mini Mental State Examination (MMSE) may result in data with discrete and skewed distributions that necessitate proper statistical models for valid inference. We review different longitudinal approaches to model cognitive decline data in older individuals and provide recommendations for model choice and result interpretation.

Methods: We used data from Alzheimer's Disease Neuroimaging Initiative study and focused on MMSE scores collected on up to four visits over a two-year period in older individuals (mean age 73 years). At baseline individuals were classified as having Alzheimer's disease (AD), early or late mild cognitive impairment, subjective memory concern, or being cognitively normal. We considered generalized additive models for location, scale and shape (GAMLSS) with binomial/beta-binomial response distribution and parametric/non-parametric random effects, selected the best model and used graphs for illustration.

Results: Binomial model with non-parametric random intercept and slope fit the best according to the Bayesian Information Criterion. The three way-interaction between time, age and diagnostic group was statistically significant suggesting that AD individuals had the steepest cognitive decline among all groups, especially in younger individuals. Furthermore, males and APOE4 carriers had worse cognitive performance, while more educated people had better cognitive performance compared to less educated. Various plots were used to illustrate and aid in interpretation of the results.

Conclusion: GAMLSS are an appropriate class of models providing interpretable results for repeatedly measured cognitive test data. We recommend that they are used more widely, accompanied by effect estimation, statistical testing and visualizations for illustration.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-07 A multivariate disease progression model for identifying subtypes in CADASIL

Sofia Kaisaridi[1], Hugues Chabriat[2], Sophie Tezenas du Montcel[1]

1 Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France

2 Département de Neurologie et Centre Neurovasculaire Translationnel, Centre de Référence CERVCO, FHU NeuroVasc, Hôpital Lariboisière, AP-HP, Université de Paris, INSERM, Unité Mixte de Recherche 1161, Paris, France, Paris, France

Introduction and Objective(s): Disease progression models are a promising tool for analysing longitudinal data presenting multiple modalities. Such models can be used to estimate a long-term disease progression and to reconstruct individual trajectories that can account for the variability between patients but also between modalities. This structure allows us to perform a classification to study the underlying disease subtypes. Using such a model (Leaspy) we analysed the evolution of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), a monogenic form of cerebral small artery disease caused by mutations of the NOTCH3 gene, and we identified subgroups with similar evolution.

Method(s) and Results: We implemented the multivariate version of a bayesian mixed effects model (Leaspy) which includes a time reparametrisation to analyse the evolution of CADASIL. We analysed data from 395 patients with 2007 visits, recruited at the French National Referral Centre CERVCO, using Leaspy to assess the temporal variability, resulting from different pace of progression and temporal offset, along with the spatial variability, which takes the form of a reordering in the sequence of events. Our multivariate analysis allowed us to model the disease progression as a 14-dimensional vector, and thus account for the relationships between the dependent variables, that showed variations in the starting deterioration timepoint. Using a GMM clustering algorithm for the individual parameters we identified a subgroup of patients with early onset and rapid progression, who shows motor symptoms and neurological deficits earlier, and a group with late onset and slow progression, who shows cognitive symptoms earlier. Male gender, lower educational level, hypertension and, most importantly, the presence of the mutation in the EGFr domains 1-6, previously identified as an important determinant of disease severity only at a cross-sectional level, are associated with more severe manifestations.

Conclusions: The multivariate model was able to highlight the differences in the evolution of the clinical scores suggesting a gradual but also heterogeneous decline in CADASIL and the profiling revealed two subtypes that imply both the temporal and the spatial variability, and are influenced by gender, education, hypertension, and the location of the mutation. A modification of the method to implement clustering on the fixed effects is ongoing and will be presented.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-08 Use of interrupted time-series analysis to examine the progression of respiratory viruses amidst the SARS-CoV-2 pandemic

Maria De Martino[1], Maddalena Peghin[2], Leonor Suarez[2], Cristina Rovelli[2], Paolo Antonio Grossi[2], Miriam Isola[1]
1 University of Udine, Udine, Italy
2 Infectious and Tropical Diseases Unit, Department of Medicine and Surgery, University of Insubria-ASST-Sette Laghi, Varese, Italy

Introduction: During the COVID-19 pandemic, comprehensive actions were taken to curb the transmission of this virus. At the beginning, predominantly non-pharmacological preventive measures (NPIs – Non Pharmacological Interventions) were enforced, serving as a pivotal tool in combatting the virus. These measures seem to have markedly curtailed the spread of SARS-CoV-2, but simultaneously exerted a noteworthy influence on the seasonal dissemination of other respiratory viruses. The aim of this study is to delineate the epidemiological change of community respiratory infections not related to SARS-CoV-2, diagnosed within the hospitals "ASST dei SeFe Laghi" from January 2022 to May 2023.

Methods: From January 2022 to May 2023, 270 patients with respiratory infections caused by viruses other than SARS-CoV-2 were included. The information regarding the type of virus were collected and the outcome considered was the monthly frequency of infections. An interrupted time-series analysis (ITSA) using ordinary least-squares (OLS) regression-based approach was used to assess the change in incidence of respiratory viruses following the interventions prescribed by various legislative decrees during that period, particularly the legislative decree of October 31, 2022, which removed the vaccination obligation for healthcare professionals. The model was adjusted for seasonality.

Results: In the month following the legislative decree of October 31, 2022, there appears to be a significant increase in the number of infections by 30.5 (95% CI 9.6, 51.4; p=0.010), followed by a rise in the monthly trend of 20.5 (95% CI -3.2, 44.3; p=0.081). As expected, the autumn season is associated with an increase in the number of infections, while the spring season is linked to a reduction.

Conclusions: The implementation of legislative decrees appears to have had an impact on the spread of viruses during the considered period. However, it is crucial to consider the effect of different seasons on virus transmission and the limited time frame of data collection. For this reason, data collection is still ongoing and analyses will be updated in light of new findings.

References: *[1] Ehrenzeller S et al., Impact of SARS-CoV-2 Preven6on Measures on Non-SARS-CoV-2 Hospital-Onset Respiratory Viral Infections: An Incidence Trend Analysis From 2015-2023. Clin Infect Dis. 2023 [2] Linden A., A matching framework to improve causal inference in interrupted time-series analysis. J Eval Clin Pract. 2018*

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-09 LEASPY: LEArning Spatiotemporal patterns in PYthon

Juliette Ortholand, Nemo Fournier, Pierre-Emmanuel Poulet, Sofia Kaisaridi, Nicolas Gensollen, Etienne Maheux, Sophie Tezenas Du Montcel, Stanley Durrleman
Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-7503, Paris, France

Over the past decade, disease progression models have been turning heads as a prominent tool for unravelling the mechanisms of chronic diseases. In longitudinal cohorts, patients are rarely followed for the entire disease duration, thus the data typically consist of snapshots of the disease course. The main challenge is to realign patients on a common disease timeline (temporal aspect of the disease). On the other hand, while describing the typical progression patterns, we also need to assess the remaining heterogeneity between individuals, which is reflected in a variability in the reordering of outcome progression (spatial aspect). To achieve these goals, we introduce Leaspy, an open-source library in Python, which implements spatio-temporal models for disease progression. Leaspy (LEArning Spatiotemporal patterns in PYthon) is built on the Disease Course Mapping model proposed by [Schiratti et al.]. This model is a non-linear mixed-effects model, whose spatial formulation is grounded in Riemannian geometry, and which uses time-warps to perform temporal registration of trajectories. It has been successfully applied in various diseases (Alzheimer's, Parkinson's, Huntington's, Amyotrophic Lateral Sclerosis, Ataxia, …). The library allows fitting multivariate models to different types of input (continuous, ordinal or time-to-event in a joint model), potentially adjusting trajectories for covariates and then making dynamic predictions for new patients. The models' parameters estimation is performed with an MCMC-SAEM algorithm. The library allows users to easily tailor their model for immediate use, while its generic structure eases the integration of new models. This open-source package is developed following modern practices (systematic testing, continuous integration and deployment, peer-review procedures). Leaspy is a ready-to-use Python library for disease progression models that disentangle spatial patterns from temporal patterns. It can therefore be a new, reliable tool for the advancement of clinical research in progressive diseases.

Reference: *[1] Schiratti, et al. A Bayesian mixed-effects model to learn trajectories of changes from repeated* **manifold-valued observations. Journal of Machine Learning Research (2017)**

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-10 Design considerations for continuous time models in progressive diseases

Doug Thompson, Aris Perperoglou, Dave Lunn, Doug Thompson
GlaxoSmithKline, United Kingdom

The Mixed Model for Repeated Measures (MMRM) is used ubiquitously in the analysis of longitudinal clinical trials across many indications in medicine research. It treats time as categorical and therefore makes no parametric assumptions about the time-outcome relationship. Although the advantage of relaxing model assumptions around the functional form of time is appealing, in some settings it may also be seen to be restrictive. Particularly when the goal is to characterise disease progression according to a more principled understanding of the continuous evolution of patient outcomes with progressive conditions such as Alzheimer's disease (AD). Numerous methods may alternatively be used when fitting flexible continuous time models, for instance regression splines, penalised splines, or fractional polynomials. Their suitability depends upon – amongst other things – key characteristics of the adopted study design. In this work we aimed to explore some common design considerations toward a better understanding of the key strengths and weaknesses of continuous time models in longitudinal data analysis. We devised a simulation study to compare a classic MMRM against continuous time models where time is handled using: (i) natural cubic splines; (ii) penalised splines; or (iii) fractional polynomials. We consider scenarios for the true data generating mechanism motivated by pa erns that would be typical of progression amongst treated and untreated patients in clinical trials for AD. Planned visit density is a key design parameter anticipated to influence the utility of continuous time models in contrast to the MMRM. In this presentation we will present findings on statistical power, type 1 error, bias and mean square error under each analysis method and each design scenario. Additionally, we will provide our reflections on alternative estimands made possible under continuous time models (e.g., extrapolation of effect) and discuss some of the possible barriers for their wider adoption.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-11

Impact of different longitudinal data representations on transformer performance in small data applications

Kiana Farhadyar[1,2], Lukas Königs[1], Harald Binder[1,2]
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg im Breisgau, Germany
2 Freiburg Center for Data Analysis and Modeling University of Freiburg, Freiburg im Breisgau, Germany

Introduction: Longitudinal cohort data offer valuable insight into the developmental trajectories of individuals in biomedical research. Yet, analysis of such data may have challenges due to the complexity of the temporal patterns and limited sample sizes. Therefore, applying state-of-the-art approaches in such settings requires special care. We specifically consider the use of transformer architectures, which are based on attention mechanisms, i.e., weighting schemes over the course of time.

Methods: To address the challenge of using transformers in a small data application, we propose converting longitudinal datasets into a discrete event format that is more in line with the application transformers have originally been developed for, i.e., natural language processing. We accordingly define these discrete events as our vocabulary set. In the next step, to achieve a vector representation for each word, i.e., word embedding, we use an autoencoder, which is a dimension reduction approach to the original binary variables. The encoder then learns the correlation structure of single events, leading to a more informed word embedding than the pragmatic one-hot-encoding approach. Moreover, we investigate the use of large language models (LLMs), e.g., pre-trained GPTs, to infuse semantic information about the correlation structure of each two words in real life. We assume that LLMs, being trained on many texts, have a language understanding of the events that tend to happen together. Therefore, using general prompt engineering recommendations, we ask the LLM to imagine themselves as an introduced persona and answer questions on the likelihood of defined discrete events happening together. Afterward, we can infuse the information we have gained on the correlation between these events, i.e., words, into the autoencoder dimensionality reduction method to augment our word embedding.

Results: We illustrate the advantages of such event coding and embedding with data that comprises self-reported stressors collected in a longitudinal resilience assessment study. We see that the word embeddings (at the event level, i.e., reported stressors) and attention scores (at the sequence level) align well with subject matter knowledge on stressors. In particular, when we directly use the transformer on the data, the model only learns to distribute the attention scores equally over sequence, but when we use the proposed embedding, it leads to meaningful attention patterns.

Conclusion: As our results show, the use of the proposed representation scheme and embedding can provide a promising foundation for using the maximum capability of transformers when analyzing longitudinal cohort data with a small number of individuals.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-12 Identifying factors associated with physician assistance in dying for older adults: A cohort study

Komal Aryal, Aaron Jones, Andrew Costa
McMaster University, Hamilton, Canada

Introduction: As more older adults are diagnosed with terminal illnesses, many are actively considering Physician-Assisted Death (PAD), which enables competent adults with incurable, advanced-stage, irreversible medical conditions to administer medical assistance in dying. This study aims to understand the differences between participants who considered PAD but did not undergo the procedure and those who underwent a PAD.

Methods: We conducted a retrospective cohort study comparing all participants who considered a PAD and those who experienced a PAD-related deaths between April 28, 2017, to February 26, 2022, and the Canadian Longitudinal Study on Aging (CLSA). Clinical and demographic characteristics were collected for all participants in the CLSA who died and had a completed decedent interview. We employed univariate logistic regression analyses to describe the association between demographic and clinical factors and the consideration and reception of PAD. Subsequently, we conducted LASSO regression to identify the most influential characteristics associated with PAD consideration and undergoing PAD.

Results: There was a total of 1,287 deceased participants with a completed decedent interview of which 183 considered PAD and 66 experienced PAD. In both groups, more participants were male, married, and died of cancer. For those who considered PAD, they were more likely to be younger (OR 1.76, CI 1.14-2.67), die in hospice or palliative care (OR 1.72, CI 1.16-2.54), and use the ED once or more in the last year of life (OR 1.66, CI 1.02-2.65). For those who had a PAD, they were more likely to die at home (OR 2.35, CI 1.27-4.27), and less likely to not die in place and not experience peace with dying (OR 0.45, CI 0.23-0.81; OR 0.24, CI 0.06-0.66). The male gender, advancing age, absence of religious affiliation, frequent use of emergency department services, and having end-of-life or healthcare arrangements significantly influence consideration of physician-assisted death (PAD), while the presence of life insurance has the most significant influence on undergoing PAD.

Conclusion: The findings suggest some incongruence between participants who consider PAD and the participants who undergo it. Future studies should explore the barriers between considering PAD and its implementation to ensure older adults have a comfortable end-of-life experience.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-13 Longitudinal clustering of liver cancer biomarkers

Bashayr Aldawsari, David Haughes
Health Data Science Department, University of Liverpool, Liverpool, United Kingdom

Introduction: Liver cancer, particularly hepatocellular carcinoma (HCC), causes a significant global health burden. Early detection of HCC is crucial for effective treatment and improved patient outcomes. Over the years, alpha-fetoprotein (AFP) has emerged as a key biomarker for the detection and monitoring of liver cancer. However, understanding the longitudinal patterns of AFP levels and their association with the risk of HCC remains a challenge. Our goal is to identify distinct trajectories of AFP levels and investigate their relationship with the risk of developing HCC.

Methods: There are several longitudinal clustering methods that can identify groups of individuals based on similar longitudinal patterns of AFP over time. We employ different advanced longitudinal clustering methods, including Latent Class Mixed Models and Mixtures of Generalized Linear Mixed Models. Also, K-means method.

Results: By applying various clustering techniques, we compared different longitudinal clustering methods and examined whether they yielded similar numbers of clusters, AFP patterns and also whether they assign same patient to the same group. Additionally, we evaluated the performance of these methods to determine their effectiveness in identifying meaningful AFP trajectories linked to HCC risk.

Conclusion: The findings from this study provide a better insight on the longitudinal clustering of AFP trajectories in liver cancer datasets. By comparing various clustering methods, we aimed to enhance our understanding of the dynamics of AFP levels and their implications for the risk of HCC. Our findings contribute to the growing body of research aimed at improving the early detection and management of liver cancer through advanced statistical methodologies. We also offer a through comparison of statistical approaches to longitudinal clustering.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-14 Examining heterogeneity in longitudinal data using latent class mixed effect modelling

Francesca Little, Noelle Van Biljon, Carlyle Mccready
University of Cape Town, Cape Town, South Africa

Introduction: The modelling of longitudinal data usually focuses on characterizing mean trajectories using generalized linear mixed effect models. However, the interest has increasingly focused not on identifying single mean trajectories but on examining the heterogeneity inherent in the data through unsupervised identification of clusters of similar longitudinal profiles. Latent class mixed effect models (LCMM) present an extension of mixed effect models to allow for the identification of latent classes within a longitudinal response. The application of LCMM allows for statistics to guide the determination of the optimal number of latent classes. Another approach is to view the identification of a stepwise increasing number of latent profiles as a method for identifying and illustrating the heterogeneity inherent within a response, and to explore this heterogeneity in increasingly finer detail as more classes are allowed. We illustrate this through the application of latent class mixed effect modelling to anthropometric growth and lung function profiles from a South African birth cohort.

Methods: Anthropometric growth measurements of weight and height for children aged 0 to 5 years were modelled using LCMM, both univariately and in a joint multivariate model. Profiles over time were characterized using piecewise-linear-spline models. Lung function measures were modelled using polynomial splines and an interrupted time series regression approach to cope with real and mechanically induced changes over time. Results based on a single mean profile were compared to results allowing for an increasing number of latent classes. Additionally, the movement of subjects across these classes given the increasing levels of heterogeneity was explored.

Results: Four latent profiles were identified for standardised height and weight measurements when modelled individually, this extended to five when latent classes were determined in a multivariate setting based on joint weight and height measurements. Additional latent classes captured heterogeneity in the period immediately after birth, after which they reconverged at fewer distinct levels. The additional heterogeneity due to changes in combined weight and height was also illustrated. For lung function measures, we were able to illustrate the division of latent classes into further sub classes that allow for characterization of more complex patterns.

Conclusion: LCMM is a powerful technique for data-driven identification and illustration of inherent heterogeneity in longitudinal profiles. Whilst various measures are useful to determine an optimal number of latent classes and to guard against spurious results, showing the transition from one mean profile to ever increasing latent profiles can also be informative.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-15 Cluster-based recurrent marked point process approach for longitudinal volume-outcome studies

Marta Spreafico[1,2], Marta Fiocco[1,2,3]
1 Mathematical Institute, Leiden University, Leiden, Netherlands
2 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
3 Trial and Data Center, Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands

Background/Introduction: High-volume healthcare providers, such as large hospitals or surgeons with a high procedure activity, may indicate more resources or experience. Volume outcome studies focus on evaluating whether patients receiving care from high-volume practitioners exhibit better post-treatment patient outcomes compared to those treated by low volume ones. Despite the great interest in volume-outcome studies in the literature, adopting the appropriate statistical methodology presents several challenges in daily practice, and the difficulties associated with it are often not recognised. There is no standard definition of the volume measure, as several specifications may be used based on the scientific question of interest. Volume is often analysed as a categorical variable, and the selection of cut-off points for the different categories may impact the statistical significance of the obtained volume-outcome associations. Alternatively, time-fixed aggregated measures are used, thereby neglecting the time-dependent nature of hospital/surgeon volume. In addition, volume-outcome analysis raises unique methodological issues because volume represents not only the time-dependent covariate of primary interest but also the cluster size. Many studies ignore bias that may occur in the estimation process when standard methods are used and certain assumptions are violated, as in the case of informative cluster size.

Methods: In this study, a novel approach combining within cluster resampling (WCR) and recurrent marked point process is developed to analyse longitudinal volume-outcome clustered data of patients who underwent oesophageal cancer surgery in Dutch hospitals. Both non-aggregate and aggregate measures for hospital-volume are considered, and various estimation methods are compared. A simulation study is performed to assess the performance of the proposed method and to evaluate whether the use of aggregate measures for volume introduces bias in the estimation of the volume-outcome association.

Results: Various estimation methods revealed a significant inverse relationship between hospital volume and post-treatment patient mortality. The estimated associations differed among the different estimation methods due to the bias introduced by informative cluster size. Simulations showed that when informative cluster size is present, the proposed method estimates the volume parameter with relatively small bias. Results also indicated that bias may be introduced in the estimation process when an aggregate measure is used.

Conclusion: Both the volume measure and the estimation method must be accurately selected based on the context of the analysis. Thanks to its cluster-based sampling scheme, the proposed methodology effectively analyses longitudinal volume-outcome data with informative cluster size.

POSTER SESSION B-10:
Longitudinal Data Analysis

P-B10-16 LB

Identifying clustering methods for longitudinal data with categorical features: A scoping review

Ottavio Khalifa[1], Viet-Thi Tran[2], Alan Balendran[3], François Petit[4]

1 Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), Paris, France
2 Centre for Clinical Epidemiology, Hôtel-Dieu Hospital, Assistance Publique-Hôpitaux de Paris, University of Paris, Centre of Research in Epidemiology and Statistics, Paris, France, Paris, France
3 Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), Paris, France
4 Université Paris Cité, Université Sorbonne Paris Nord, INSERM, INRAE, Centre for Research in Epidemiology and Statistics (CRESS), Paris, France, Paris, France

A central challenge in statistical epidemiology is longitudinal data analysis. These types of data consist of observations of one or several variables ordered over time. In the case of clinical pathways, these variables are often categorical, i.e., discrete and non-ordinal. In recent years, cluster analysis has received interest for analyzing such data, primarily through so-called 'State Sequence Analysis' (SSA) methods [1]. However, this task is common in various other fields. It remains to be seen which methods from the literature could be applied in epidemiology, to what extent they are applicable, and what their limitations are.

To address these questions, we conducted a methodological scoping review. We started with a qualitative search to identify key expressions used in various communities to designate this task, such as 'Sequence Analysis' in sociology, in order to build a search equation on Web of Science, PubMed, and Google Scholar. We also established a set of criteria to assess the underlying hypotheses, scalability, and reproducibility of these methods, such as their ability to handle missing data, time-invariant covariates, irregular sampling, large volumes of data, or the presence of a public implementation.

We identified methods used by communities such as sociology, epidemiology, computer science, and bioinformatics. Our search query returned 5970 articles from all databases, from which we extracted around 100 methods. We classified them into three categories: distance-based (e.g., all variants of the so-called Optimal Matching distance [2]), model-based (e.g., mixture of Markov chains [3]), and featurization-based (e.g., methods based on Fourier transform). We identified subgroups in these categories to provide a relevant classification for each domain of application. We notably discovered that the majority of methods lack public implementations and face challenges with irregular sampling or covariates.

As cluster analysis is increasingly used to mine trajectories in EHR and administrative databases, our findings could serve as a valuable compass, guiding researchers in applying these methods in their future work.

References: *[1] Mathew et al. "Applying sequence analysis to uncover 'real-world' clinical pathways from routinely collected data: a systematic review." Journal of clinical epidemiology vol. 166 (2024). [2] Studer et al. « A Comparative Review of Sequence Dissimilarity Measures », Journal of the Royal Statistical Society Series A: Volume 179, Issue 2 (2016). [3] Zhang et al. « Model-based clustering of time-dependent categorical sequences with application to the analysis of major life event patterns ». Stat. Anal. Data Min. 14(3) (2021)*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-01 Estimating treatment effect in longitudinal trials with monotone missingness for mixed observations

Joydeep Basu, Nigel Stallard
University of Warwick, Coventry, United Kingdom

Introduction: We consider the setting of a clinical trial with a continuous primary endpoint with a number of missing observations when we have additional data on an early endpoint or baseline covariate for all patients and can assume that the primary endpoint is normally distributed with expected value linearly related to the early endpoint. Galbraith and Marschner[1] show how to improve the precision of an estimator for the mean of the primary endpoint by constructing the maximum likelihood estimate (MLE) utilising all available observations on both endpoints when the early endpoint is normally distributed. In this case the pair of observations for each patient has a multivariate normal distribution.

Methods: In this work, we propose imputation of the missing primary endpoint values with expected values from a linear regression relating the primary and early endpoint observations. We show that the estimator obtained is the maximum likelihood irrespective of the distribution of the early endpoint, allowing the approach of Galbraith and Marschner to be extended to settings in which the early endpoint is not normally distributed.

Results: Explicit expressions are provided for the MLE and its variance when the early endpoint has a Bernoulli distribution, with the latter demonstrating the gain in precision from using the early endpoint data for patients for whom primary endpoint data are not available. Focussing on this setting, we also compare our procedure with multiple imputation techniques implemented by mice package in R. Conclusions: Our proposed approach provides a simple imputation method by which we can gain precision in estimation of the mean of a primary endpoint using data on an early endpoint when some primary endpoint observations are missing and compares favourably with more computationally intensive multiple imputation approaches. Limitations and further possible extensions of the method are also discussed.

References: *[1] Galbraith, S., Marschner, I.C. Interim analysis of continuous long-term end- points in clinical trials with longitudinal outcomes, Statistics in Medicine, 22, 2003, 1787-1805.*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-02

Assessment and comparison of classical and machine learning approaches for imputing missing data in the framework of excess hazard analysis

Vineet Kumar Kamal[1], Roch Giorgi[2]
1 All India Institute of Medical Sciences, Kalyani, India
2 Aix-Marseille University, Marseille, France

Background: Missing data on covariates is a persistent problem in public health research. Population-based cancer registry data typically contain a non-negligible proportion of missing values on important prognostic factors, which can cause biased estimates and biased standard errors, resulting in incorrect p-values and confidence intervals; and inefficiency (loss of information from the available data). It would be interesting to compare and assess the performance of classical approaches and machine learning-based imputation approaches in a comparative mode in the framework of excess hazard analysis in order to appropriate analysis of cancer registry data.

Methods: We generated a simulation study of 1000 sample sizes, represented moderate-sized studies, with 1000 simulations in the framework of the basic (linear proportional effect of covariates at diagnosis) relative survival model with varying censoring rates (20%, 30% and 50%) to compare multiple imputation methods using traditional MICE (Multiple Imputations by Chained Equations), MICE-CART (Classification and Regression Tree) and MICE-RF (Random Forest). Before imputation, we created missingness in covariates, keeping an overall percentage of missing value as 50 percent, under missing at random (MAR) mechanism with monotone and non-monotone pattern. We assessed the performance of imputation methods using the bias, the relative bias, the empirical coverage rates, and the mean of the standard error estimates. Also, we used real data of a population-based study of French colorectal cancer of French cancer registry data to visualize the simulation results. Additionally, we compared missing data indicator, and single imputation techniques (naïve, KNN, missForest, spmm, famd, missranger, misscforest) in addition to above mentioned multiple imputations in French cancer registry data. We used MICE package in R for amputation and imputation.

Results: In simulation data, we found MICE has the least mean bias and relative bias, however, RF has a comparatively good coverage rate (>95%) with minimum mean standard error. In real data, we found that missing data indicators analysis has the most precise estimates, but MICE and RF had almost similar performance in our scenario. Single imputation methods do not seem to be very beneficial, however, they were not very inferior in comparison to multiple imputation in this scenario, with KNN as the best and Naive method as the most imprecise.

Conclusion: We recommend to use MICE or MICE-RF to impute covariates in the framework of basic excess hazard analysis.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-03 A data driven approach to address missing data in the 1970 British birth cohort

Michail Katsoulis[1], Martina Narayanan[2], Brian Dodgeon[2], George Ploubidis[2], Richard Silverwood[2]
1 Population Science & Experimental Medicine, Institute of Cardiovascular Science, UCL, London, United Kingdom
2 Centre for Longitudinal Studies, Institute of Education, UCL, London, United Kingdom

Background/Introduction: Missing data may induce bias when analysing longitudinal population surveys. We aimed to tackle this problem in the 1970 British Cohort Study (BCS70).

Methods: We utilised a data-driven approach to address missing data issues in BCS70. Our method consisted of a 3-step process to identify important predictors of non-response from a pool of ~20,000 variables from 9 sweeps in 18037 individuals. Briefly, this process involves Stage 1: Univariable modified Poisson regressions of non-response at sweep t on each individual potential predictor of non-response at sweep 0 up to sweep t –1 in complete-case analysis. Keep predictors with p-value<0.001. Stage 2: Multivariable modified Poisson regressions of non-response at sweep t on all predictors retained from stage 1, separately from sweep 0 up to sweep t –1, in complete-case analysis. Keep predictors with p-value<0.05. Stage 3: MI using all retained variables plus non-response at sweep t in the imputation model. MI multivariable modified Poisson regressions for all retained predictors at sweep 0, up to sweep t –1, adjusted for predictors at all previous (but not subsequent) sweeps. Keep predictors with p-value <0.001. Our aim was to identify a moderate set of variables (predictors of non-response) that can be used as auxiliary variables in principled methods of missing data handling to restore baseline sample representativeness.

Results: Individuals from disadvantaged socio-economic backgrounds, increased number of older siblings, non-response at previous sweeps and ethnic minority background were consistently associated with non-response in BCS70 at both early (ages 5-16) and later sweeps (ages 26-46). Country of birth, parents not being married and higher father's age at completion of education were additional consistent predictors of non-response only at early sweeps. Moreover, being male, greater number of household moves, low cognitive ability, and non-participation in the UK 1997 elections were additional consistent predictors of non-response only at later sweeps. Using this information, we were able to restore sample representativeness, as we could replicate the original sample distribution of father's social class and cognitive ability. Moreover, we presented an example from survival analysis where we showed that when we utilised for the set of auxiliary variables (obtained from our approach) in multiple imputation, we were able to reduce the bias due to missing data in the relationship between father's socioeconomic status and mortality.

Conclusions: We provide a set of variables that researchers can utilise as auxiliary variables to address missing data issues in BCS70 and restore sample representativeness.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-04 Biomarkers with limit of detection: Modelling composite event in patients with PAD

Natasa Kejzar[1], Ales Blinc[1,3], Anja Boc[1,3], Kevin Pelicon[3], Klemen Petek[3], Mojca Bozic Mijovski[2,3], Josko Osredkar[3], Vinko Boc[3]
1 University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia
2 University of Ljubljana, Faculty of Pharmacy, Ljubljana, Slovenia
3 University Medical Centre Ljubljana, Ljubljana, Slovenia

Intro/Background: Patents with peripheral arterial disease (PAD) admitted for endovascular revascularisation of the lower limb arteries were observed for all-cause death combined with major ischemic events within one year. The aim was to estimate the prognostic value of two biomarkers: high-sensitive cardiac troponin I (hsTnI) and ischemia-modified albumin (IMA).

Methods: Data from 487 patients with PAD from 2020 to 2022 was analysed. The biomarkers were measured at the time of admission. They were log-transformed. The limit of detection (LoD)for hsTnI was 10 ng/l and for IMA 0.48 U/l. The LoD was considered in different ways. Tobit regression was used to estimate the mean values of log biomarkers for patients with and without events. Multiple logistic and Cox survival models were fixed to account for this population's already-known predictor variables. Sensitivity analyses with two-step (propensity score) modeling were performed, with the first step balancing the patients, and the second step fixing the final regression model. The results were compared with two simplified methods of treating LoD variables: treating them as categorical variables (below/above the LoD) or assigning the values below the LoD half of the LoD value.

Results: In our analysis, 65% of the patients had hsTnI values below the LoD and 23% had IMA values below the LoD. Mean log biomarker values significantly differed between patients with and without event. Multiple regression models showed marginal significance for both biomarkers across all analyses.

Conclusion: The results were similar in the simplified analyses, although the difference between the(non-) event groups appeared to be greater. Furthermore, the change in estimate increased with the proportion of values below the LoD.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-05

The use of Joint and Cox models to assess the association between a longitudinal marker and a time-to-event: A simulation study under different missing mechanisms and applications in ICU setting

Matteo Petrosino, Stefania Galimberti, Maria Grazia Valsecchi, Paola Rebora
University of Milano-Bicocca, Milano, Italy

Introduction: Dynamic information plays a crucial role in the continuous monitoring and predictive assessment of patients' health status. In the Intensive Care Units (ICU) setting several markers are measured longitudinally, but longitudinal profiles frequently encounter missing data or are no longer observed owing to patient recovery or end-of-life considerations. Joint models (JM, [1]) were introduced to account for missing-not-at-random in longitudinal studies, addressing for random measurement error on the temporal covariate, and represent an interesting alternative to the time-dependent Cox model (TDCM) to evaluate the association between a marker and a time-to-event endpoint. Inspired by a study on the evaluation of the association between ICP and mortality in patients with Acute Brain Injury (ABI) [2], we aim to investigate by simulation the impact of missing values in the evaluation of the association between a longitudinal marker and mortality. The objective is to assess the ability of JMs and TDCM in estimating the true association, in terms of hazard ratio (HR), when longitudinal trajectories have different functional forms and exhibit intermittent observations or monitoring truncation.

Methodology: In R, we simulated individual longitudinal markers and time-to-event data from a JM, specifying two different functional forms for the covariate: linear and quadratic. With a probabilistic approach we recreated two missing data patterns, with varying missing rates, resulting in intermittent and truncated profiles. In the first case the covariate marker exhibits lack of measurements with higher probability when the marker value is high (or low), in the second case the profiles are no longer observed after the first missing value. In our motivating context [2], 514 patients with ABI were recruited from 13 ICUs. Among 318 patients with ICP monitoring, ICP value was recorded every 4 hours throughout the first week of ICU, while for a subgroup of about 100 patients a minute-by-minute measure of ICP was available, and we observed both intermittent and truncated ICP profiles.

Results: Preliminary simulations results suggest superior performances of JMs over TDCM in estimating the true HR, particularly when trajectories are intermittent or deviate from linearity. Results on the clinical motivating data indicate a significant association between ICP raise and the risk of death overtime with both approaches.

Conclusions: Summarizing, the study accentuates the call for employing JMs to discern associations between longitudinal markers and temporal outcomes, positing the TDCM as a viable alternative, underscoring the need for further research in environments characterized by missing data patterns and different marker trajectories.

References: *[1] Rizopoulos, D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R (1st ed.). Chapman and Hall/CRC (2012). [2] Oddo M., Taccone F., Galimberti S., Rebora P., Citerio G. Orange Study Group. Outcome Prognostication of Acute Brain Injury using the Neurological Pupil Index (ORANGE) study: protocol for a prospective, observational, multicentre, international cohort study. BMJ Open. (2021)*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-06

Investigating the performance of missing data methods when applied to time-series environmental exposure data: A simulation study

Bohan Zhang[1], Andy Vail[1], Craig Smith[2], Amit Kishore[2], Matthew Gittins[1]
1 University of Manchester, Manchester, United Kingdom
2 Salford Royal NHS Foundation Trust, Manchester, United Kingdom

Background/Introduction: Analysis of exposure data may be subject to bias and reduced statistical power due to missing observations. New Machine Learning techniques for missing data imputation are growing in popularity, but few studies have explored their reliability in time-series exposure data. This study aims to investigate the performance of new and emerging imputation methods within a time-series environmental exposure context. It will explore their performance under different missing data generation mechanisms, when applied and compared to an already complete exposure dataset that reflects our known truth.

Methodology: We will simulate missing data in a dataset that has complete data. We are interested in the performance of missing data imputation techniques within environmental exposure data measurements. This study will be conducted in several steps as follows: the simulation step, the application step and the validation step. Firstly, we will identify a complete empirical exposure dataset. Then, simulate missing data under different mechanisms and impute missing values using above methods. In the validation step, the imputed dataset will be compared against the known truth. To analysis the effect of different method.

Preliminary Results: Environmental exposure dataset identified from the DEFRA (Department for Environment, Food and Rural Affairs) website with 87648 hourly observations for three different pollutants, i.e PM2.5, PM10 and NO2 from multiple measurement sites from Manchester in the time interval between 2013 and 2023. These variables were found to possess different cases of missingness located between 3% and 7.5%, and the missing data were all at the MCAR. Afterwards, a complete dataset was picked with 29934 complete observations. MCAR and two missing ratios between 5% and 20% were simulated to get two different datasets. With the application of simple imputation methods, Last Observation Carried Forward (LOCF) has better performance than Mathematics Statistics Value Imputation (MSVI). by having smaller Normalized Rooted Mean Square Error (NRMSE) value. Where more complexed method and different missing ratio and missing mechanism still to be applied.

Conclusion: Between LOCF and MSVI, LOCF has a better imputation performance on MCAR dataset. But with the high Normalized Rooted Mean Square Error (NRMSE) value suggests these imputation are not good enough for the actual research and more complexed imputation methods need to be applied and analysed.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-07 Missing data imputation in the context of propensity score analysis: A systematic review

Sarah Friedrich, Saghar Garayemi
Augsburg University, Augsburg, Germany

Background and Introduction: Researchers often grapple with the common challenge of missing data in observational studies [1]. Conversely, propensity score analysis is a method within observational studies to emulate conditions reminiscent of clinical trials. Nowadays most of the studies just consider the complete cases for analysis but as we know this has many pitfalls [2]. Studying different methods for imputing data helps us to know how each method works in different situations. In the context of propensity score analysis understanding which methods of handling missing data lead to a better balance of covariate distribution between cases and controls could help us to make a better inference about the outcome.

Methods: To assess which methods are used for (multiple) imputation in the context of propensity score analysis, we conduct a systematic review. Therefore, we considered two important keywords: "propensity score" and "missing data" and we restricted ourselves to the time between 2010 and February 2024. We consider PubMed a comprehensive library of medical research. After excluding systematic reviews, pure theoretical papers, non-English publications, and animal studies our research resulted in 137 manuscripts. We extracted research parameters such as sample size, and study design, as well as considerations for missing data, such as imputation methods used, pre- and post-imputation outcome comparisons, and the use of sensitivity analyses. Publications are divided into two-year intervals to identify possible patterns in missing data imputation. Furthermore, the method used to estimate propensity scores is considered another aspect that may influence data imputation efficacy.

Results and Conclusion: Our literature review investigates the extent to which methods for handling missing data are employed in observational studies using propensity score analysis. We especially focus on the aspect of reproducibility i.e. detailed description of the methods used and aspect of sensitivity analysis. Investigating the effects of different imputation methods in the context of propensity scores in more detail will be a part of future simulation studies.

References: *[1] Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons; 2009. 297 p. [2] Barnard J, Rubin DB. Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika. 1999;86(4):948–55.*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-08

Scoping review of software implementations of quantitative bias analysis to informatively missing data

Emily Kawabata[1,2], Margarita Moreno-Betancur[3,4], Katherine J Lee[3,4], Kate Tilling[1,2], Rachael A Hughes[1,2]

1 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
3 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne, Australia
4 Department of Paediatrics, University of Melbourne, Melbourne, United Kingdom

Introduction: In health research, data can be informatively missing, where the likelihood of missing data depends on unobserved factors such as the missing values themselves or unmeasured variables. In these scenarios, standard missing data methods such as multiple imputation or inverse probability weighting may not be appropriate as they require the assumption that missingness is not informative. Since the analyst cannot determine if this assumption is met based on the observed data only, a quantitative bias analysis (QBA) is recommended to assess how conclusions change under different assumptions about the missingness. The adoption of QBA approaches by applied researchers has been slow, which could be due to lack of training in these more complex methods, but also the availability of accessible software. To assess this, we conducted a scoping review of current state-of-the art software implementations of QBA approaches to informative missingness.

Methods: We conducted three searches: (1) of R packages available on the Comprehensive R archive Network, (2) of Stata commands on the Boston College Statistical Software Components archive, and (3) of any software implementation (i.e., not limited to R and Stata) reported in the published literature. Our inclusion criteria were software that did not require adaption (i.e., code changes) before application, were made publicly available in the past decade, were still available in January 2024, and were accompanied by documentation. Key properties of each software program were extracted.

Results: Our review identified 14 programs of which three were Bayesian approaches, eight were imputation-based, two were maximum likelihood approaches, and one was a weighting-based approach. Most (64%) were only applicable for informative missingness in the outcome of the substantive analysis. However, programs covered diverse types of incomplete variables (continuous, binary, ordinal, nominal) and types of data (aggregate, individual-level, longitudinal). Only five programs provided specialist features to conduct a QBA under multiple missingness assumptions simultaneously (i.e., a probabilistic or multidimensional bias analysis). The remaining programs required the user to repeatedly apply the program for different missingness assumptions and to summarise and display the multiple sets of results.

Conclusions: Software is now available to implement a QBA for informative missingness. However, current software is still limited in terms of the range of missingness scenarios covered and provision of specialist QBA features. Software implementations should include inbuilt features to aid applied researchers to conduct a QBA as routine practice. Provision of detailed QBA guidelines for informative missingness would also be highly beneficial.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-09

Accounting for missing data when calculating conditional/predictive power

Neža Dvoršak[1], Thomas Burnett[1], Christopher Jennison[1], Robin Mitra[2]
1 University of Bath, Bath, United Kingdom
2 University College London, London, United Kingdom

Introduction: Missing data is a problem in any research area, and it occurs more often than desired. Clinical trial research is no exception. Missing data in clinical trials often arise for various reasons and especially in trials with an adaptive design, the missing data problem takes on added complexity due to the evolving nature of this study design.

Methods: In our setting of an adaptive confirmatory Phase 3 trial, we have the option to stop the trial early for futility. To do so we consider the calculation of conditional/predictive power for continuous endpoints. One of the challenges when making the decision is the problem of not having all the data available, whether that be due to the recruitment process, timing of analyses, or patient behaviour (loss to follow-up). This can cause additional uncertainty in the estimated probabilities that guide the decision about the continuation of the trial. In our approach we treat this as a missing data problem. By implementing different techniques to deal with missing data, particularly emphasizing multiple imputation (MI), we address the issue of incomplete information with the aim to enhance decision making regarding futility stopping at an interim analysis.

Results: We consider the different sources of missingness as well as the amount of missing data in the dataset and present our findings in a simulation study.

Conclusion: We demonstrate the problems introduced through missing data while making interim decisions in adaptive clinical trials. By having this in place, we are then able to show the potential benefits of implementing missing data methods in futility stopping decisions, showcasing how they may be used in practice to improve decision making.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-10 Quantifying the benefits of double measurements for anthropometric estimates in paediatric clinical trials

Dominik Grathwohl
Societé des Produids Nestlé, Lausanne, Switzerland

Introduction: Anthropometric measurements, such as weight, height, and head-circumference, are typically taken as single measurements in pediatric practices. However, in pediatric clinical trials, protocols often require double measurements. The purpose of the double measurement is to enhance precision. In this research work, we investigate the error of double measurements, and we quantify the improvement compared to single measurement using the concept of design-effect. We make mainly use of published data by the WHO and we will provide a different view on the WHO-data.

Methods: In "Reliability of anthropometric measurements in the WHO multicentre growth reference study" the coefficients of reliability R is defined as R=1-TEM(inter)2/SD2, where TEM(Inter) is the inter-observer technical error of measurement and SD2 the total variance. To quantify the advantages of double measurements versus single measurements, we suggest the concept of the design-effect. We use the following definition of the design-effect DE=VAR'/VAR where VAR' is the total variance of a trial where anthropometry was estimated by double measurements and VAR is the total variance of a trial where anthropometry was estimated by a single measurement. One can show that VAR'/VAR equals n'/n, the ratio of the sample sizes needed in both scenarios. This provides a new interpretation of the design-effect, in terms of sample size gain. We proof that the coefficient of reliability is equal to the design-effect: DE=R.

Results: The MGRS reliability paper reports on the coefficients of reliability for the double measurements R: 0.99, 0.99, 0.97, 0.97, for length, height, head-, arm-circumference, respectively. Considering that R=DE this would mean that employing double measurements for these anthropometry parameters instead of a single measurement, would lead to a reduction in the sample size of one subject for length and height and three subjects for head and arm circumference if the trial would be powered for 100 subjects.

Conclusion: The WHO reliability paper focuses on reporting the technical error of measurements to demonstrate high reliability and comparable performance of clinical teams with an anthropometry expert. It does not specifically compare single measurements with double measurements. However, we provide a different view on the data of the WHO, suggesting that it can be used to quantify the gain in precision achieved through double measurements compared to single measurements, in terms of sample size. We prefer to communicate on sample size rather than variance, because sample size directly impacts the budget and timelines of a clinical trial.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-11

Methods of estimating parameters of skewed or truncated normal distribution in the presence of observations outside of measurable range

Markéta Janošová[1], Stanislav Katina[1], Jozef Hanes[2]
1 Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, Czech Republic
2 Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia

Introduction: Every laboratory equipment has limits to what it can accurately measure. Generally, for every laboratory apparatus three types of limits should be defined – limit of blank, limit of detection and limit of quantitation. If an observation falls outside of the measurable range, there is an issue of estimating parameters of the distribution. Several methods can be used to address the issue, but their performance needs to be considered before selecting, which method to use.

Methods: In this contribution we look at four different methods – ignoring censored observations, replacing censored observations, using a truncated version of target distribution, and using target distribution with censored observations. To compare these methods we designed a simulation study, where generated samples were trimmed from the left at selected quantiles, mimicking situations with certain percentages of the population falling below the limit. Parameters' estimates were then compared to the original values. Simulation study was run separately on skew-normal distribution and truncated normal distribution.

Results: Based on the results of the simulation study for truncated normal distribution, using truncated normal distribution with censored observations provides the best estimates of the parameters. For left-skewed skew-normal distribution, replacement by 0.5-times the limit provides the best estimates of the parameters. For right-skewed skew-normal distribution, using skew-normal distribution with censored observations provides the best estimates of the parameters.

Conclusion: It may be tempting to use methods that are easy to implement, however selecting an incorrect method may lead to biased estimates and change the research outcomes, e.g. incorrect result of two-sample test about means when comparing two populations or biased estimation of regression line.

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-12 Maximum likelihood estimation of the correlation coefficient for a trivariate normal distribution with missing data

Tamae Kawasaki[1], Takashi Seo[2]
1 Aoyama Gakuin University, Shibuya-ku, Japan
2 Tokyo University of Science, Shinjuku-ku, Japan

Introduction: Several studies have been carried out on the estimation of correlation coefficients when data are missing. In particular, Minami and Shimizu (1997), Minami and Shimizu (1998) and Garren (1998) discuss the bivariate case. Minami and Shimizu (1997) and Minami and Shimizu (1998) consider the maximum likelihood estimate and the restricted or residual maximum likelihood estimate for the common intraclass correlation coefficient, and also discuss the asymptotic variance. Garren (1998) presents the theoretical asymptotic nature of the MLE of correlation coefficients. This study discusses the estimation of correlation coefficients for missing data in the three variables.

Methods: The maximum likelihood estimator is derived from a trivariate normal distribution. We consider the correlation coefficients between each of the three variables when the observation of one of the variables is missing.

Results: The estimates are given as solutions of polynomial equations. Therefore, in simulations and numerical examples, the evaluation is through the approximate solution of polynomial equations by computational methods. Conclusion: This study considers the estimation of correlation coefficients for observations where one variable is missing. An example and some simulation results are given to illustrate the characteristics of the estimates.

References: *[1] Garren, S.T. (1998). Maximum likelihood estimation of the correlation coefficient in a bivariate normal model with missing data, Statistics & Probability Letters, 38, 281-288. [2] Minami, M. and Shimizu, K. (1997). Estimation for a common intraclass correlation in bivariate normal distributions with missing observations, American Journal of Mathematical and Management Sciences, 17, 3-14. [3] Minami, M. and Shimizu, K. (1998). Estimation for a common correlation coefficient in bivariate normal distributions with missing observations, Biometrics, 54, 1136-1146.*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B11-13 Can we estimate a risk without observing the relevant number of cases?

Toshiro Tango
Center For Medical Statistics, Tokyo, Japan

Introduction: The relevant important data cannot be always collected due to the difficulty in doing so. In this presentation, we shall consider the following situation: We would like to investigate the association of some adverse reproductive outcomes such as infant deaths with mothers living within 10km from 63 municipal solid waste (MSW) incinerators with high dioxin emission levels in Japan. The study area was defined as circles of radius 10 km from MSW incinerators. To estimate the "risk profile" around the MSW incinerators, each study area was divided into ten sub-areas (called "zone") delimited by ten circles of a radii of 1,2,…,10 km. For each zone, population size or live births is available but the relevant number of deaths is difficult to obtain. But we can obtain the total number of deaths in each study area. In this situation, we would like to estimate the profile of crude or standardized infant death rate.

Method and Results: Taking advantage of the total observed number of deaths in each study area and the population size in each zone, a non-parametric estimator is proposed to get the mean profile of risk assuming the homogeneity among different MSW incinerators. Regarding the estimation of standard error of estimates, the Bootstrap method can be used. The basic idea is similar to that described in a different context elsewhere [1].

Conclusion: The proposed non-parametric estimator might potentially have a wide applicability to some estimation problem where we cannot observe the relevant data.

Reference: *[1] Tango T. Linear equations with random variables. Statistics in Medicine 24, 2005, 3213 - 3222.*

POSTER SESSION B-12:
Rare Data Analysis or Case Studies in Medical Statistics

P-B12-01 Rare variants statistical tests for analysing genetic changes in 15Q13.3

Audrone Jakaitiene[1], Tadas Zvirblis[1], Christian P. Schaaf[2]
1 Vilnius University, Vilnius, Lithuania
2 Heidelberg University, Heidelberg, Germany

Background: Genetic changes affecting the copy number of chromosome 15q13.3 have been associated with a group of rare neurodevelopmental conditions (autism spectrum disorder, epilepsy, schizophrenia, and others). The critical region contains approximately 10 genes. Treatments are limited and are restricted to targeting the main symptoms rather than the underlying aetiology. Not each person harboring a 15q13.3 copy number change will manifest disease, and the severity and clinical diagnosis is difficult to predict. This represents a significant challenge in modelling and determining health outcomes. We will review recent statistical models applicable for the 15q13.3 CNV rare disorder analysis.

Methods: Advances in high-throughput next-generation sequencing and biostatistics have greatly increased the number of association studies. Classical single-variant based association test will result in low statistical power for low-frequency and rare variants unless sample sizes or effect sizes are very large. Therefore, recently, scientists have developed new methods tailored for rare-variant association analysis to boost power [2]. We will cover recent aggregation tests for rare variants, such as burden, variance-component, and omnibus tests. Results and Conclusion: Aggregation tests have been developed to test for disease association with sets of rare variants. There is no definitive evidence which model to use. Selection and relative performance of these methods will depend on prior information available [1, 2]. The methods are new, and limited information exists regarding relative performance, especially when using real data. One method was developed using real data from neurodegenerative diseases [1]. Funding. This work is part of the EJP RD project 'Resolving complex outcomes in 15q13.3 copy number variants using emerging diagnostic and biomarker tools (Resolve 15q13)' No. DLR 01GM2307 and has received funding from EJP RD partner the Research Council of Lithuania (LMTLT) under grant agreement No. S-EJPRD-23-1.

References: *[1] Boutry, Simon, et al. "Rare variant associa9on on unrelated individuals in case–control studies using aggregation tests: existing methods and current limita9ons." Briefings in Bioinformatics 24.6 (2023): bbad412. [2] Lee, Seunggeung, et al. "Rare-variant association analysis: study designs and statistical tests." The American Journal of Human Genetics 95.1 (2014): 5-23. Keywords. rare variant, aggregation test, statistical power, real data*

POSTER SESSION B-11:
Analysis of Imperfect Data (including Missing Data & Measurement errors)

P-B12-02 On the comparison of Classical, Bayesian and Jackknife binary logistic regression models in paediatric epilepsy patients

Adnan Karaibrahimoglu[1], Mutlu Altuntas[2]
1 Suleyman Demirel University, Isparta, Turkey
2 Sinop University, Sinop, Turkey

Introduction: Jackknife resampling is a method in which the analysis is repeated by removing a single observation from the data set and this process is repeated for each observation in the data set in turn. This method is particularly effective in situations based on small sample sizes and prior knowledge. Bayesian statistics is a statistical approach based on probability theory to obtain information about uncertainty. Epilepsy is a neurological disorder caused by abnormal electrical activity of brain cells. Epileptic seizures in children are characterized by symptoms such as loss of consciousness, muscle twitching and confusion, which usually start suddenly and last for a short time. The aim of this study is to compare the performances of various binary logistic regression models to obtain more significant Odds Ratios in rare diseases such as epilepsy in infants.

Methods: This study was enrolled with a total of 262 infant epilepsy patients who admitted to pediatric neurology clinics. The data were collected within five years. The seizure status of infants were classified as relapse and remission. The gender, diagnose age, cranial MRI, EEG, number of antiepileptic drug usage and 5-year diagnose age were independent variables for patients. The models for seizure status were established as classical, Jackknife, Bayesian and Bayesian Jackknife Binary Logistic Regression. The Bayesian logistic regression model allows us to analyze in a more flexible and broad framework. Jackknife logistic regression is a method that allows the model to be re-estimated by removing an observation from the data set.

Results: In binary logistic regression models, the area under curve (AUC) in ROC curve of the models was found as 0.857, 0.855, 0.857 and 0.855 for classical, Jackknife, Bayesian and Bayesian Jackknife, respectively. Since the sample size was big enough for Jackknife method, a smaller sample size was randomly determined (n=70) and the models were established with same methods. Therefore, the AUC's of the models were 0.886, 0.915, 0.891 and 0.912, respectively.

Conclusion: It can be concluded that the Jackknife methods give us better results than classical and Bayesian models when the sample size is small. It can be useful, especially, for rare diseases.

POSTER SESSION B-12:
Rare Data Analysis or Case Studies in Medical Statistics

P-B12-03

When the size matters: A simulation study evaluating stability of variable importance driving pancreas re-transplant graft survival in a very small dataset using machine learning techniques

Cinzia Anna Maria Papappicco, Giulia Lorenzoni, Maria Vittoria Chiaruttini, Dario Gregori
Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy, Padua, Italy

Introduction: Addressing the challenge of very small datasets is crucial, as they hold huge potential in providing meaningful insights in scenarios where diseases are so rare that large-scale data collection is impractical. Inherent limitations of small datasets can hinder machine learning algorithms from effectively learning models. The use of such methods, like Random Survival Forest (RSF), raises doubts about reliability of variable importance measure (VIMP) estimates, because they might not be robust for few statistical units and correlation among predictors.

Methods: Starting from a motivating real-case study comprised 10 patients who underwent a pancreas re-transplant, we developed a RSF model for exploring potential factors driving graft survival and we designed a simulation study with synthetic very small datasets to evaluate the stability of VIMP scores estimated by the RSF model. We set four different scenarios of beta coefficients to mimic varying degrees of variable importance. We constructed two simulation designs, generating 7 vectors (predictors), each consisting of 10 random observations. In simulation design (A) we generated random variables normally distributed and, in simulation design (B), random variables that follow a Bernoulli distribution. The simulation was repeated for 1,000 iterations, through different scenarios. Subsequently we recorded the percentage of the variables ranked in order of importance to assess which variables were consistently considered the most important across the 1,000 simulations.

Results: Preliminary results allowed us to quantify the stability of the VIMP estimates. In 1,000 simulations across the four variable importance degree scenarios, the effect of the coefficient improved the identifying ability of the informative variables and thus the stability of the RSF model estimates. In particular, the results showed the variables with the highest beta ranked at most in the third position 70% of the time across 1,000 simulations, suggesting potential of RSF in providing reliable estimates by mitigating uncertainties related to analysis of small samples.

Conclusion: Our results highlight the potential of RSF in the analysis of very small datasets, while emphasizing the importance of validation procedures such as simulation to confirm the reliability of estimates. Our simulation methodology offers a replicable approach to evaluate the stability in VIMPs in similar contexts representing a starting point that solicits an intensified research effort to extend machine learning applications also in clinical contexts where we can rely exclusively on few statistical units.

MINI SYMPOSIUM 1

Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development

MS1-1 Beyond conventional randomised controlled trials: patient reported outcomes in single arm studies

Saskia le Cessie
Leiden University, Netherlands

Single arm trials (SAT) can serve as alternatives to randomized control trials, for example to accelerate drug approval for new promising treatments in seriously ill patients, to study treatments for rare diseases, and in situations where randomization is not feasible or ethical. However, the lack of a randomized control group may compromise the validity of conclusions, particularly when assessing the effect of a specific treatment on Patient Reported Outcomes (PROs), such as quality of life, in a single arm study.

In this talk we will discuss several issues which arise in the design and analysis of SATs with PROs. Research questions on PROs in SATs may be descriptive or confirmatory. In both cases, the ICH-E9 estimand framework for clinical studies can help define relevant research questions. We demonstrate that different summary measures and different approaches to handle intercurrent events may give different results and conclusions, even in a descriptive setting. Specifically, addressing death should be carefully considered in advance, as patient reported outcomes after death are not defined. Furthermore, missing data are a potential source of bias in the results, as reasons for missing data may depend on the patient's medical condition. We address some new methods to deal with it.

Making statements on treatments in single arm studies is challenging as changes over time in PROs cannot be solely attributed to the treatment. Various factors such as natural changes over time (e.g., due to disease worsening), response shift and the effects of concomitant therapies and comorbidities may also contribute to observed changes.

An alternative is to compare the results of a SAT to external data. This also possesses challenges: from defining a relevant estimand for the treatment effect, to accounting for confounding and different study drop out.

The work presented here is part of the European IMI-SISAQOL project. SISAQOL-IMI is an international project, led by the European Organisation for Research and Treatment of Cancer and Boehringer Ingelheim. The aim of this four year project is to establish international standards in the analysis of patient reported outcomes and health-related quality of life data in cancer clinical trials.

References: *[1] ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. [2] Pe M, et al.; SISAQOL-IMI Consortium. Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. Lancet Oncol. 2023 Jun;24(6):e270-e283. doi: 10.1016/S1470-2045(23)00157-2. PMID: 37269858.*

MINI SYMPOSIUM 1

Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development

MS1-2 Applications of designs with external controls / hybrid designs in drug development

Marc Vandemeulebroecke
UCB, Basel, Switzerland

In drug development, randomized controlled trials (RCTs) are the gold standard for generating evidence of a treatment effect. Yet, in specific situations there may be reasons for deviating from this gold standard. Alternative design options have been developed that come with more or less operational ease and statistical rigor. Trial designs without concurrent control and/or with a purely external control are on the extreme of the spectrum. They are prone to biases and apply mainly to niche situations, such as when the untreated counterfactual is severe and certain (e.g., rapid death). Hybrid designs, however, offer an alternative, more tempered option that can be tuned to a wider range of situations. In these designs, a concurrent control is augmented with external information in a controlled way. In this talk, we explore opportunities and challenges on this spectrum of trial design options. We share our experiences with implementing hybrid designs at scale in a mid-size pharmaceutical company, highlighting the value of generating experience and accumulating efficiency gains. We explore application areas for such designs from exploratory to pivotal development phases.

MS1-3 Voilà: The European collaborative project INVENTS

Sarah Zohar
INSERM, Paris, France

The evaluation of new medicines for rare diseases (RD) including paediatric RDs is challenging for several reasons, among which are the small patient sample sizes, heterogeneity of patients and diseases and heterogeneity in disease knowledge. Due to these difficulties, access to effective treatments and the number of treatment options are often limited in RDs.

INVENTS is a European collaborative under Horizon Health Cluster aiming at providing clinical trial trialists, researchers and regulators with a global framework encompassing methods, workflows and evidence assessment tools to be implemented in RD drug development. The project kicked-off in January 2024 and it ambition is to significantly improve the evaluation of evidence and regulatory decision-making through the development and validation of: refined longitudinal model-based diseases trajectories and treatment effect, improved extrapolation models, in silico trials (e.g., virtual patient cohorts), optimised model-based clinical trial designs and evidence synthesis methods. These methods and models will be evaluated through simulation studies and challenged on extensive data from a range of use cases provided by our industrial partners Roche and Novartis and Real World data from the French National RD registry, they will also contribute to the scientific aspect of this project. The INVENTS framework will improve consistency and efficiency of the drug evaluation process for RD by augmenting clinical evidence without compromising its scientific integrity and providing regulators assessment credibility criteria.

At the end of this 5 years project, the RD trialist community will be able to exploit novel and improved clinical trial designs, in silico trials and RWD analysis approaches supporting drug development in RD. The European Medicine Agency and European national regulators (including Health Technology Assessment bodies) will be supplied with a general framework allowing better informed decision making. Most importantly, RD patients will benefit from an increased and faster access to efficacious and safe treatments.

INVENTS is funded by the European Union's Horizon Europe Framework programme under grant agreement 101136365. INVENTS is also funded by the Swiss State Secretariat for Education Research and Innovation (SERI).

MINI SYMPOSIUM 1

Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development

MS1-4 Regulatory aspects

Kit Roes
Radboud UMC, Netherlands

The confirmatory nature of the evidence generated to answer the primary research question of efficacy for new drugs is a cornerstone in regulatory decision making. The randomised clinical trial – in many of its shapes and forms – is the proven successful design to address this. Regulatory perspectives on novel design options and modelling approaches typically would, and possibly should, start from this proven quality of the RCT. Regulatory approval and subsequent reimbursement decisions are of huge importance to large populations of patients and have large societal impact. Ensuring a high standard of (confirmatory) evidence for such decision making is thus a shared obligation across all stakeholders involved. At the same time: (1) the RCT as often employed has known limitations in some settings, (2) decision making involves assessing a totality of evidence for benefit – risk and overall drug characteristics and (3) advancement of science and in-depth knowledge of disease (from molecule to man) and methodology & modelling have the potential for improving research designs for the (broad) range of research questions relevant for drug development. In specific settings such as rare diseases and pediatric development and novel treatment modalities (e.g. anti-sense oligo nucleotides almost tailored to individual patients) the classical confirmatory RCT may not be possible or will actually not be able to answer the primary research questions within a reasonable time frame. Modelling underpinning research design, data collection and analysis is in some sense at the very core of statistics (following Sir D.R. Cox): statistical inference is at its core based on an assumed model for the data generating mechanism. The usual (regulatory) statistical standards for assessing the level of evidence for confirmatory trials (type 1 error control, power, p=values etc.) may not be sufficient to establish the strength of (confirmatory) evidence of more complex designs, of results that substantially build on modelling results with limited data, or real world data studies. For modeling and simulation (complex) models, a credibility framework was proposed, that takes a broader perspective, specifically including the considerations of the size of impact and risk on the regulatory decision. In this presentation, regulatory aspects for typical examples - inspired by the presentations in this session - will be discussed based on the broader context of regulatory decision making and the principles underlying such a credibility framework.

MINI SYMPOSIUM 1

Beyond Conventional RCTs: Exploring Design Options and Modelling in Drug Development

MS1-5

Advanced trial designs and analysis methods to improve confirmatory trials in rare diseases

Martin Posch, Franz König
Medical University of Vienna, Austria

Clinical trials for rare diseases face unique challenges due to the limited number of participants, often resulting in underpowered studies and treatment effect estimates with high uncertainty. The adoption of innovative trial designs – such as platform trials, hybrid trials that incorporate external control data, and adaptive trials – can enhance trial efficiency, particularly if they are combined with modelling approaches. An example is platform trials, where model-based analyses allow for the incorporation of non-concurrent controls while maintaining robustness under time trends. Other approaches to improve the precision of treatment effect estimates include the integration of multiple endpoints into a single outcome measure for increased statistical power, the utilisation of longitudinal data, and appropriate accounting for prognostic or predictive baseline covariates in the analyses.

However, the use of modelling in the analysis and design of clinical trials can affect their robustness, primarily due to the additional assumptions inherent in the models. This is of particular concern in a confirmatory setting. Therefore, understanding the underlying explicit and implicit assumptions and assessing the consequences of their violation is essential.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future

MS2-1 for STRATOS

MS2-1 On six foci for the future of STRATOS

Willi Sauerbrei, J Carpenter, M Abrahamowicz, N van Geloven, P Gustafson, M Huebner, R Keogh, P Shaw, Els Goetghebeur for the STRATOS initiative

In this talk we discuss foci for the next 3 years of STRATOS research with the hope that further colleagues may support the initiative. More details were recently published (Carpenter et al, 2023, Biometric Bulletin 40(4), 7-9; available on the website)

Future foci

1. Simulation studies. Simulation studies are key tools for validating and comparing statistical methods, and hence critical to the development of evidence-based statistical guidance. STRATOS will maintain a focus on simulation studies and prioritize improving their methodology over the coming years.

2. Open science. The importance of open science is evident, but it is an extremely broad topic, and still in its infancy. For some challenges we will work on accessible guidance for making research more transparent, reproducible and hence credible.

3. Initial Data analysis (IDA). The 'Initial data analysis' TG3 aims to improve awareness of IDA as a critical component of the research process, and develop guidance on conducting IDA in a systematic, reproducible manner. Some issues will be discussed in the talk by CO Schmidt.

4. Machine learning (ML) enhanced statistical methods. While ML methodologies promise quick automated data driven answers to many questions, it is obvious that both ML and established statistical methodologies have their specific strengths and weaknesses. Each could benefit from the insights offered by the other. How to do that best and when is not obvious. We plan to identify the ML enhanced statistical methods that are most important for different TG's, and systematically assess their properties in realistic settings.

5. Estimands in observational data analysis. The term 'estimand' essentially refers to what is being estimated and for whom. In the trials context, the ICHE9 addendum (ICH, 2019) formally defines it in terms of five components which make for clear targets and more transparent reporting. The insights and benefits which the estimands framework is bringing to trials research are equally needed in observational studies, where much of the relevant methodological expertise was originally developed. This topic is discussed in a parallel mini-symposium with a contribution from a STRATOS project.

6. More guidance for researchers with limited statistical knowledge and experience from the beginning, STRATOS highlighted that many methodological developments are not implemented in practice. Lack of guidance on practical issues is presumed to be an important hurdle. Researchers with only basic statistical knowledge and limited experience in using statistical methodology need much more help.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future MS2-1 for STRATOS

MS2-2 for TG3

Building blocks of Efficient Initial Data Analysis and Data Quality Assessments – Best practice examples

Carsten Oliver Schmidt, Lara Lusa, Marianne Huebners

Rigorous statistical analyses require an adequate understanding of the underlying data. Gaining such an understanding is the main goal of Initial Data Analysis (IDA) [1] and data quality assessments (DQA) [2]. IDA and DQA overlap strongly, but differ in that the former being more focused on assessing the fulfillment of prerequisites for the intended substantive analysis, whereas the latter has a more generic focus on data properties. Several works provide guidance on the building blocks for comprehensive and efficient implementation of IDA and DQA. These building blocks range from the setup of metadata to the assessment algorithms used for IDA and DQA. This talk provides best practice examples on the conduct of IDA and DQA in the context of observational health studies, using data from the Study of Health in Pomerania (SHIP) and the Survey of Health, Ageing and Retirement in Europe (SHARE). It will be illustrated, how a comprehensive information management supports automated assessments to increase the scope and quality of IDA and DQA related analyses.

References: [1] Huebner M, le Cessie S, Schmidt CO, Vach W. A. Contemporary Conceptual Framework for Initial Data Analysis. Observational Studies. 2018;4(1):171-92. doi:10.1353/obs.2018.0014. [2] Schmidt CO, Struckmann S, Enzenbach C, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Med.

MS2-3 for TG8

An Overview and Recent Developments in the Analysis of Multistate Processes

Malka Gorfine

Multistate models offer a powerful framework for studying disease processes and can be used to formulate intensity-based and more descriptive marginal regression models. They also represent a natural foundation for the construction of joint models for disease processes and dynamic marker processes, as well as joint models incorporating random censoring and intermittent observation times. This article reviews the ways multistate models can be formed and fitted to life history data. Recent works on pseudo-values and accommodation of random effects as a method of incorporating a dependence on the process history and between-process heterogeneity are also discussed. The software available to facilitate such analyses is listed.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future MS2-1 for STRATOS

MS2-4 for TG6

Assessing performance when developing or validating clinical risk prediction models in the era of machine learning

Ben van Calster, Ewout Steyerberg

An abundance of performance measures for clinical risk prediction models have been proposed in the statistical and machine learning literature. We aim to provide an overview of contemporary performance measures for models with binary outcomes, motivated by the assessment of the value of the previously developed ADNEX model to predict whether an ovarian tumor is malignant in external validation data (n=894, 49% malignant tumors). We consider five domains of model performance. These include overall measures (e.g. Brier score), measures for discrimination (e.g. AUROC), and measures of calibration (e.g. expected calibration error). When supporting a clinical decision for the patient, a decision threshold on the estimated risk is required to define classification as high versus low risk. The 2x2 table of classification versus outcomes can be described with classification measures (e.g. F1) and clinical utility measures (e.g. net benefit). We discuss 32 common performance measures (9 overall, 3 discrimination, 6 calibration, 11 classification, 3 utility). For each performance domain, matching graphical assessments are available. We define three key desirable characteristics for performance measures: properness (i.e. whether the value of the measure is optimal when the correct risks are used); having an understandable interpretation; and having a clear focus by targeting only one of the five domains. The majority of measures fail for at least one characteristic, while the F1 score fails at all three. All considered classification measures at a given threshold t are improper. A natural requirement is that a performance measure should match the intended use of the model. We discern three common situations. First, when externally validating models that aim to support clinical decision making, it makes sense to assess performance in the following order: discrimination (AUROC), calibration (calibration plot) and clinical utility (net benefit). Second, if a model is merely used for informing/counseling patients about their risk, external validation should focus on calibration. Third, when methodologically comparing multiple models, overall measures are useful. Other measures may be added, if they meet the three key characteristics. In conclusion, we recommend to consider a limited set of key measures to assess performance aspects in relation to the intended use of a prediction model, focusing on (semi-)proper measures with a clear interpretation and focus.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future MS2-1 for STRATOS

MS2-5 for TG2/ TG4

Adjusting for covariate measurement error on functional form estimation: design and early results from a blinded, collaborative STRATOS project

Aris Perperoglou, Paul Gustafson, Michal Abrahamowicz, Victor Kipnis, Mohammed Sedki, Anne Thiébaut, Lawrence Freedman

Introduction: The evaluation and estimation of relationships between outcome variables and covariates measured with error remains a challenge in observational studies. Challenges may be amplified when the true functional form between the covariate and the outcome is suspected to be non-linear. This work outlines the collaboration between two topic groups of the STRATOS initiative: TG2, specializing in the selection of variables and functional forms in multivariable analysis, and TG4, which focuses on addressing issues related to measurement error and misclassification. The project investigates the performance of methods for estimating complex functional relationships in observational data, where covariates are prone to measurement inaccuracies, specifically targeting the accurate estimation of non-linear relationships between outcome variables and covariates.

Project Design: The project adopts a blinded, multi-stage design to rigorously compare methodological approaches. A Data Generation and Evaluation team produces datasets simulating various functional relationships and measurement error scenarios, but withholds the true underlying model from the Methods teams. Three distinct Methods teams implement different analytic approaches, based on Bayesian methods, Imputation/Regression Calibration methods and SIMEX methods, respectively. For each method, estimation of the functional form based on (i) B-splines, (ii) P-splines and (iii) Fractional Polynomials are investigated, with pre-specified hyperparameters for each approach. The initial phase of the project involved generating 5 datasets, each comprising realizations of a binary outcome Y and a continuous covariate measured with independent error, alongside pairs of repeat covariate observations for validation purposes in a random subset. The relationship between Y and X was specified as a logistic regression, but the error structure and functional relationship between logit(P(Y=1) and X remained undisclosed to prevent bias in analysis and thus enhance the integrity of the study. Each methods team created their code and returned estimated functional relationships without knowledge of the generating model or the results from any other team's methods.

Early Results and Implications: The Evaluation Team assed the performance of the methods through a comparison of predicted values against undisclosed true values, using mean squared error and other relevant metrics to gauge performance. Initial findings reveal performance disparities across methods (blinded for the study's integrity). In Phase 2 of the project a further 75 datasets have been simulated, varying sample size, functional form and size of measurement error in a systematic manner. These datasets are now being analyzed. The results will offer insights into how factors like sample size, validation study size and spline and covariate functions influence method accuracy This analysis will be important as it will guide the selection and refinement of statistical methods for the final phase of the project that will include several hundred more simulated datasets. This blinded, collaborative structure fosters an unbiased and efficient evaluation of statistical techniques. Results will contribute to the STRATOS Initiative's broader goal of providing guidance for analyzing observational data. Notably, this project design showcases a model for collaborative, transparent, and rigorous statistical research to address challenges in real-world settings.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future MS2-1 for STRATOS

MS2-6 for TG7

Causal inference moving forward – embracing joint (dis)appearances

Els Goetghebeur, Saskia le Cessie

(https://stratos-initiative.org/en/group_7)

Since its start, TG7 has presented estimands as a needed focus for any causal effect estimation. `What are we actually estimating' is surprisingly often absent from applied publications [2023, Lancet Oncol., DOI10.1016/S1470-2045(23)00110-9]. It is not straightforwardly derived, however, from how we estimate our target but also depends on what plausible causal assumptions we make a priori. From the various principled answers developed in a setting with sequential point exposures and subsequent continuous outcome, we are now moving to guidance on more complex outcomes. These include right censored survival times and repeated outcome measures while patients are alive. Other intercurrent events may then appear. Already in randomized trials there is much controversy about what constitutes a meaningful causal effect in that case. These issues and much more play in observational studies, for which many causal methods were first developed. We elaborate in this talk on ongoing and planned work, which looks to collaborate with other topic groups next. We are working on the following:

1. To clarify various causal estimands and estimators we have introduced counterfactual cross world simulations for continuous outcomes as a learning tool. We are finalizing a similar plasmode like effort, starting from an observed case study, for right censored survival outcomes.

2. In the context of the European IMI-SISAQOL project (sisaqol-imi.org) we collaborate with a large consortium to develop guidance for causal effect analysis in (late stage) oncology trials. There, quality of life while alive, as well as survival must be jointly evaluated, often in single arm studies. We define relevant and feasible estimands in that setting and develop corresponding estimators.

3. As a rule, we value causal effects in terms of this joint outcome. While intercurrent events can sometimes be handled by defining composite or (not too) hypothetical outcomes, evaluating treatment policy and outcomes until death is often preferred. When implementing analyses for the joint outcome, many analysis choices must be made. We explain how issues inherent to (single arm) trials as well as cohorts can be more rigorously approached by causal inference methods to allow for target effect estimation under transparent assumptions.

In this talk we describe pitfalls and progress made. We refer to a forthcoming collaboration with TG1 to further analysis in the presence of common missing data patterns in our setting. We look forward to discussions in Thessaloniki and further cross topics work.

MINI SYMPOSIUM 2

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future MS2-1 for STRATOS

MS2-7 for TG1

Current and future initiatives in missing data

Katherine J Lee, Els Goetghebeur, James Carpenter

The aim of TG1 is to describe the principles for the analysis of partially observed observational data, illustrate potential methods for handling missing data and their application, and provide general guidance on how best to handle missing data across a range of settings. We have previously developed a framework for the handling and reporting of missing data. We are currently expanding this framework to the context of when data are missing dependent on unobserved data. This is exemplified through a worked case study. Future initiatives of TG1 include:

1. conducting a review of journal guidelines for handling missing data in top ranked medical journals with the aim of highlighting key misunderstandings, outlining the key components which we believe are useful to include in author guidance for missing data, and suggesting a template for author guidelines,

2. providing an overview of methods for handling missing data including a discussion of plausibility of needed assumptions, pros and cons of the various approaches and example code for conducting each using a single case study, and

3. evaluating methods to handle missing data in the context of informative drop out and nonpositivity, where (nearly) all further data are missing for some categories of participants. We consider a case study on missing quality of life data in a cancer trial with substantial treatment discontinuation and drop-out due to disease progression.

In this talk we describe these initiatives of TG1, in particular also how we are developing collaborations with TG7 to address questions regarding the handling of missing data in causal inference.

EARLY CAREER BIOSTATISTICIANS DAY

ECB-1 AI, small data, and me

Harald Binder
Institute of Medical Biometry and Statistics, Freiburg, Germany

Chatbots based on large language models (LLMs) can now easily generate study protocols and guide analysts through data analysis, including refining research questions and proposing modeling approaches, suggesting that some of the work done by statisticians could be taken over. At the same time, the techniques underlying LLMs are closely related to statistical modeling techniques in statistics. This is one of many examples where statisticians may feel threatened by artificial intelligence (AI) techniques, but at the same time have valuable knowledge that could be essential for establishing AI techniques in a responsible way. Modeling in small data settings is another such example, where AI techniques might currently be perceived as invading what used to be the home turf of statisticians, but could actually benefit greatly from biostatistical modeling expertise. Using these examples, I will discuss the current capabilities of AI techniques, including success stories, but also pointing out current shortcomings that indicate a potentially fertile ground for biostatisticians. Based on the experience of organizing an interdisciplinary AI research center that comprises statisticians as well as computer scientists and mathematicians, including a graduate program, I will also discuss career paths, including potential hurdles and opportunities.

ECB-2 Design and Analysis of Co-enrolment Trials

Abraham Contreras
University of Warwick, UK

Research sites that run several studies at one time can come across participants that are eligible for more than one trial. Where compatible, participants can be given the opportunity to co-enrol, either concurrently or sequentially. Co-enrolment is growing in its occurrence throughout the world of clinical research especially as research sites increase their capacity in running several trials at a time. However, implementing co-enrolment poses organisational and logistical hurdles. With adequate preparations, studies can drastically improve efficiency in trial delivery and maximise collaboration between healthcare researchers. Design challenges include data collection for multiple trials, consent strategies for patients and research site staff, and report documentation for data monitoring and interim analysis. Statistical challenges include management of interactions between combinations of interventions, additional variables with analytical methods to account for this, and modelling patient health outcomes and movement across a clinical pathway.

ECB-3 How can we leverage the power of Machine Learning and Artificial Intelligence when we design Clinical Trials? Insights and Lesson Learned

Ajsi Kanapari
University of Padova, Padova, Italy

The popularity of Machine Learning (ML) and Artificial Intelligence (AI) has led to the discussion and opportunities that they may have in the design stage of Clinical Trials. Several reviews and discussion papers have encouraged its usage however there is a huge gap between the opportunities that such methods can create and their actual usage in trials. This talk will discuss the state of art of AI and ML in the stages of site selection, participant recruitment, risk stratification, data collection and monitoring, and finally in the analysis of Clinical Trials. It will be highlighted when the process has started, where why and if we are ready for it. Some insight will be given on the main challenges encountered during the review process related to reporting limitations about on how GPT models have been leveraged during the screening steps.

EARLY CAREER BIOSTATISTICIANS DAY

ECB-4 Working with Healthcare Systems Data in Clinical Trials

Alice-Maria Toader
University of Liverpool, UK

Randomised clinical trials (RCTs) are the gold standard for evaluating healthcare interventions. Healthcare systems data (HSD) has the potential to optimise the efficiency of RCTs, by decreasing trial-specific data demands. The use of HSD in RCTs comes with several challenges. The COMORANT-UK study has established a prioritised list of these challenges to be addressed regarding HSD. The domains of the questions included data access, data collection, and outcome selection. Within my work, I aim to understand the extent and nature of the current use of HSD and its evolution over time, alongside its opportunities and challenges.

ECB-5 Analysis of aggregated data via generation of pseudo–Individual Participant Data at Danone Nutricia Research

Luca Carlini, M Gholam
Danone Research & Innovation, Utrecht, Netherlands

In clinical research, most of the meta-analyses are based on pooled/aggregated data (two-stage approach), while Individual Participant Data (IPD) has always been considered the gold standard. Sometimes this is not feasible due to lack of available and complete data, e.g., owing to privacy reasons. Nowadays, especially in industry, it is common to encounter restrictions that grant access only to summary statistics, rather than allowing access to the complete dataset containing all relevant information; thus, it is crucial to process data coming from different studies efficiently, while adhering to legal regulations and keeping high quality standards of data collection and analysis. Our proposed solution involves the creation of pseudo-Individual Participant Data derived from summary statistics and the application of mixed models (for example to evaluate how the product effect differs across sites). Considering the method's inherent simplicity, efficiency, and flexibility, as outlined in Evidence Synthesis Methods for Continuous Outcomes, it can become useful for Early Career Biostatisticians to apply and replicate this approach. Doing so allows them to circumvent the challenge of handling incomplete information within a dataset.

ECB-6 Developing clinical prediction models in maternal health; experiences from a systematic review

Mae Chester Jones
University of Oxford, United Kingdom

Abstract: Early warning scores (EWSs) are simple scores used by healthcare professionals to spot declining health. However, until recently, there was no national maternal EWS derived on statistical modelling for use in the UK. There are almost 150 maternal EWSs currently in-use in UK hospitals which are not evidence based, validated or have even been published in peer-reviewed journals. With maternal death rate and near-miss morbidities on the rise, early recognition and intervention is crucial. In this talk, I will discuss my experience of appraising the methods for developing maternal EWSs that ranged from simple colour-coded systems to more advanced scores such as automated clinical prediction models. I will focus on the challenges of developing clinical prediction models in maternity such as handling of repeated measures, validating a model in subgroups linked to life threatening events (i.e., age, ethnicity), physiological changes during and throughout pregnancy, and missing data. I will discuss how these models need to be developed and reported in the future and what an advanced maternal early warning score would look like in the future.

EARLY CAREER BIOSTATISTICIANS DAY

ECB-7

Child Malnutrition at Upazila Level in Bangladesh: A Small-Area Estimation Approach

Mynul Islam
University of Dhaka, Bangladesh

Background: The importance of promoting good nutrition to eliminate all forms of malnutrition is clearly mentioned in SDG 2. Bangladesh is one of many countries that have made significant progress in reducing malnutrition [1]. The dramatic progress made by Bangladesh means that the country must strengthen its efforts to achieve SDG 2. For effective distribution of food aid, it is useful to assess malnutrition indicators at a geographical level so that unusually high levels of malnutrition can be targeted. The objective of this study is to estimate the prevalence of stunting, wasting and underweight among children under five years at the Upazila level and prepare a map to assist in planning development assistance programs.

Method: This study uses the small area estimation technique (ELL method) developed by Elbers, Lanjouw and Lanjouw in 2003. This study uses the 2019 Multiple Indicator Survey [2] and the 2011 Population and Housing Census of Bangladesh (10%).

Results: The small estimates generated in this report have a low standard error up to the Upazila level. Department and district level estimates are consistent with direct estimates from MICS 2019. However, some exceptions are noted. In case of small area estimate using the developed model, the district estimates of stunting, wasting and underweight for the two or three districts are found to be over-estimated or under-estimated.

Conclusion: It is believed that this study will help decision makers gain insight into Upazilalevel estimates of under-five malnutrition.

Reference: *[1] Press release, Bangladesh sees sharp decline in child malnutrition, while violent disciplining of children rises, new survey reveals UNICEF, 2020, https://www.unicef.org/bangladesh/en/press-releases/bangladesh-sees-sharpdecline- child-malnutrition-while-violent-disciplining-children*

Authors Index

A.

B.

Authors Index

C.

Authors Index

D.

Authors Index

E.

F.

G.

Authors Index

H.

Authors Index

I.

J.

K.

Authors Index

L.

M.

Authors Index

Authors Index

N.

O.

P.

Authors Index

Q.

R.

Authors Index

S.

Authors Index

T.

Authors Index

U.

V.

W.

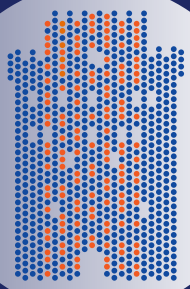
Authors Index

X.

Y.

Z.

ISCB International Society for Clinical Biostatistics



ISCB45

THESSALONIKI 2024