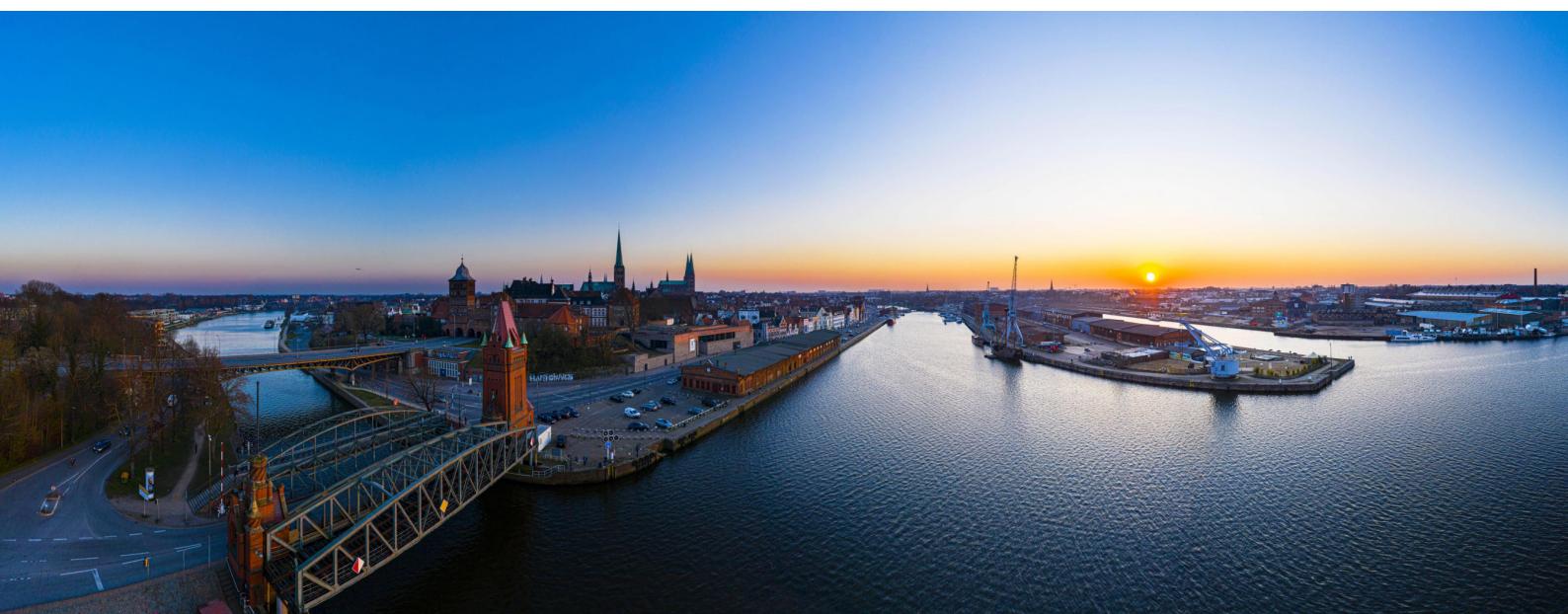


Navigating the Sea of Data: Biometrics Guides the Way

Conference & Abstract Book

70. Biometric Colloquium

28.02.-01.03.2024



UNIVERSITÄT ZU LÜBECK



The open-source L^AT_EX template, AMCOS_booklet, used to generate this booklet is available at https://github.com/maximelucas/AMCOS_booklet

THANK YOU

to our sponsors



**Boehringer
Ingelheim**



Solingen | Breda | Wien



perfood



Springer



CRC Press

Taylor & Francis Group

CHAPMAN & HALL



Die Gemeinnützige

Contents

About	6
About the Conference	6
Committees	7
Mitgliedschaft in der IBS-DR	8
General Information	9
Conference venue	9
Catering	10
WiFi	10
Floor plan	11
AG Meetings	12
Guidance for presenters	13
Social program	16
Town Hall reception	16
Get together with the Early Career Working Group (AG Nachwuchs) . . .	16
City tour with the night-watchman	17
Conference dinner	17
Restaurants	18
Bars and nightlife	19
Conference Schedule	20
Tutorials	38
Multiple endpoints and prioritized outcomes - Nonparametric analysis methods using generalized pairwise comparisons	38
Introduction to Causal Inference and Target Trial Emulation	39
An introduction to estimands and estimand-aligned estimation	40
Advanced data visualization in R: (Re)producing professional plots with ggplot2 and the tidyverse	41
Keynote Talks	42
Michael Love (University of North Carolina-Chapel Hill, USA)	42
Ina Rondak (European Medicines Agency, The Netherlands)	43
Maarten van Smeden (University Medical Center Utrecht, The Netherlands) . .	44

Invited Sessions	45
SES-22: Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics	45
SES-23: Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics	49
SES-24: Invited: Biometric Research Outside High-Income Settings	52
SES-25: Invited: Statistical Issues in Animal Testing	58
SES-31: Invited: Risk Prediction meets Causal Inference	63
SES-37: Invited: Synthetic Data with Privacy Guarantees: from Applied Statistics to Generative Machine Learning	66
SES-38: Invited: Copula Regression for Time-to-Event Data	69
SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update	72
SES-45: Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations	78
Special Sessions	84
SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data	84
SES-09: Special: Statistics in Toxicology	90
Statistics in Practice 1 and 2	97
IBS-DR: Nachwuchspreise	98
Feier zum 70. Biometrischen Kolloquium (in German only)	105
Young Statisticians	106
Panel Discussion (AG-Nachwuchs): Navigating the Academic Odyssey: Early Career Challenges	113
Workshops für Schulen: Künstliche Intelligenz - Spielerisch lernen	114
Vortrag für die Öffentlichkeit: Sich selbst mit dem Smartphone behandeln: Wie gut funktioniert das?	115
Abstracts of Contributed Talks	116
Abstracts of Contributed Posters	280
Job Advertisements from Sponsors	316

About

About the Conference

The 70th Biometric Colloquium will take place from February 28 to March 1, 2024 in the Audimax building of the University of Lübeck.

The conference slogan "Navigating the Sea of Data: Biometrics Guides the Way" emphasises the importance of biometrical knowledge and expertise for the advancement of medicine and life sciences, especially in the analysis of increasingly large and complex data sets.

The conference will give you the opportunity to learn about the latest developments in biometrics, to improve your biostatistical knowledge and skills and to strengthen and expand your scientific networks.

The conference program offers 3 keynote talks, 4 tutorials, 9 invited sessions, 13 contributed sessions, and a poster session including speed talks. Furthermore, special events include award sessions (Young Statisticians, IBS-DR Award Session), a panel discussion organized by the AG Nachwuchs, a birthday session for the 70th Biometric Colloquium, workshops for local school classes and a talk for the public.

Contact

University of Lübeck
Institute of Medical Biometry and Statistics
Ratzeburger Allee 160, V24
23562 Lübeck
Phone: +49 - 451 - 500 50610
E-mail: IMBS.bk24@lists.uni-luebeck.de

Committees

Program Committee

Chair: Anne-Laure Boulesteix (München)

Tim Friede (Göttingen)
Cornelia Frömke (Hannover)
Annika Hoyer (Bielefeld)
Klaus Jung (Hannover)
Inke König (Lübeck)
Frank Konietschke (Berlin)
Annette Kopp-Schneider (Heidelberg)
Andreas Mayr (Bonn)
Jörg Rahnenführer (Dortmund)
Andre Scherag (Jena)
Irene Schmidtmann (Mainz)
Silke Szymczak (Lübeck)
Richardus Vonk (Berlin)

Local Organizing Team

Inke König
Silke Szymczak

Members of the Institute of Medical Biometry and Statistics:

Christine Beneke
Sabine Brehm
Césaire Fouodo
Hanna Grube
Nicole Heßler
Kirsten Hinrichs
Björn-Hergen Laabs
Lisa-Marie Nuxoll
Lea Kronziel
Louis Macias
Tanja Rausch
Wolfgang Rudolph-Rothfeld
Pia Ullendahl
Maren Vens
Reinhard Vonthein

Mitgliedschaft in der IBS-DR

Die Biometrische Gesellschaft (IBS-DR) freut sich auf neue Mitglieder!

Ob Nachwuchswissenschaftler (auch schon vor dem Masterabschluss) oder erfahrene(r) Biometriker(in), ob in der öffentlichen Forschung Lehre oder in der Industrie tätig: wir freuen uns auf Sie als neues Mitglied! Zweck der Biometrischen Gesellschaft ist die Förderung der Biometrie in Forschung, Lehre und Anwendung. Sie ist ein gemeinnütziger Verein, der vom ehrenamtlichen Engagement von Biometriker/innen lebt.

Was haben Sie von einer Mitgliedschaft? (Stand: Februar 2024)

- Einfache Vernetzung mit anderen Mitgliedern aus verschiedenen Arbeitsbereichen
- Teilnahme an der Online-Seminarreihe "SIBSINAR" inkl. anschließender Diskussion und Zugang zu den Video-Aufnahmen der Präsentationen
- Teilnahme an unseren Sommerschulen, ggf. zu einem reduzierten Beitrag
- Möglichkeit zur aktiven Mitarbeit in unseren fachlichen Arbeitsgruppen (Teilnahme und Organisation von Workshops, Sessions, etc.)
- Online-Zugang zu den Zeitschriften Biometrical Journal, Biometrics und JABES
- Coming soon: Austausch-Platform (Discord)
- Reduzierte Tagungsgebühr für die Jahrestagungen (Biometrisches Kolloquium, DAGStat-Tagungen, CEN-Tagungen)
- Automatische Mitgliedschaft in der Muttergesellschaft "International Biometric Society" und assoziierte Vorteile (Teilnahme an Journal-Club, Distinguished Lecture, Mentoring, IBS Member Community)

Für Nachwuchs-Biometriker/innen:

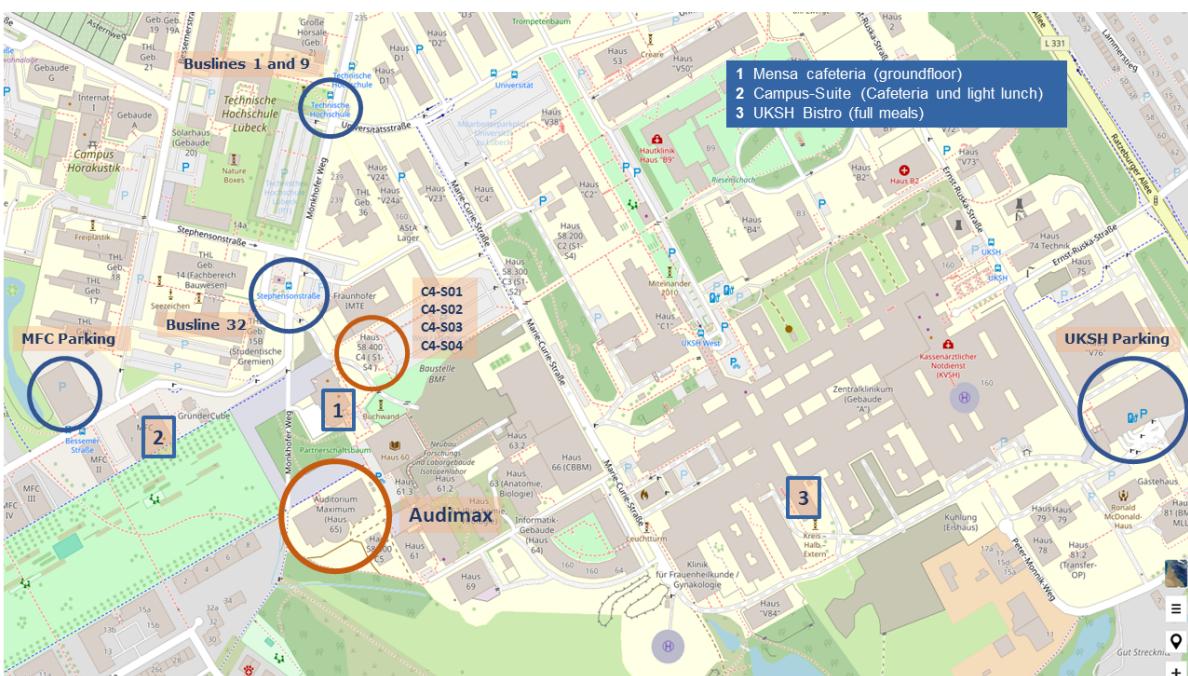
- Kostenlose Mitgliedschaft für immatrikulierte Studierende (inkl. Promotionsstudent(innen) mit einer höchstens 50%-Stelle)
- Mögliche Bewerbung um den Gustav-Adolf-Lienert-Preis
- Mögliche Mitarbeit in der AG-Nachwuchs

General Information

Conference venue

The venue of the conference is the Audimax building of the University of Lübeck:

Universität zu Lübeck
Ratzeburger Allee 160
23562 Lübeck
Audimax: Mönkhofer Weg 245
To Audimax Uni Lübeck: [Map](#)



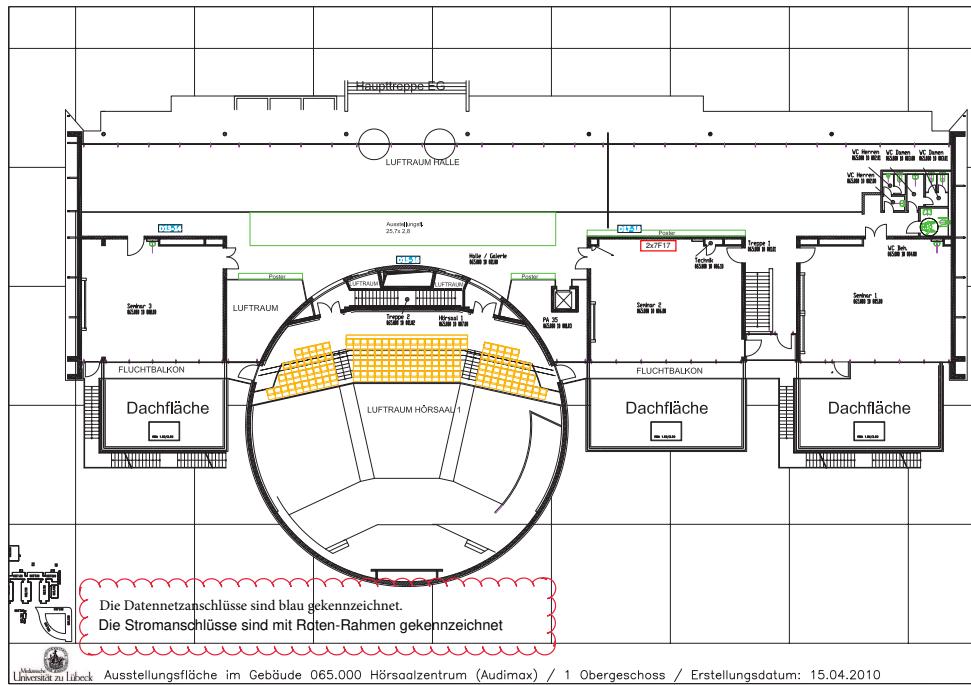
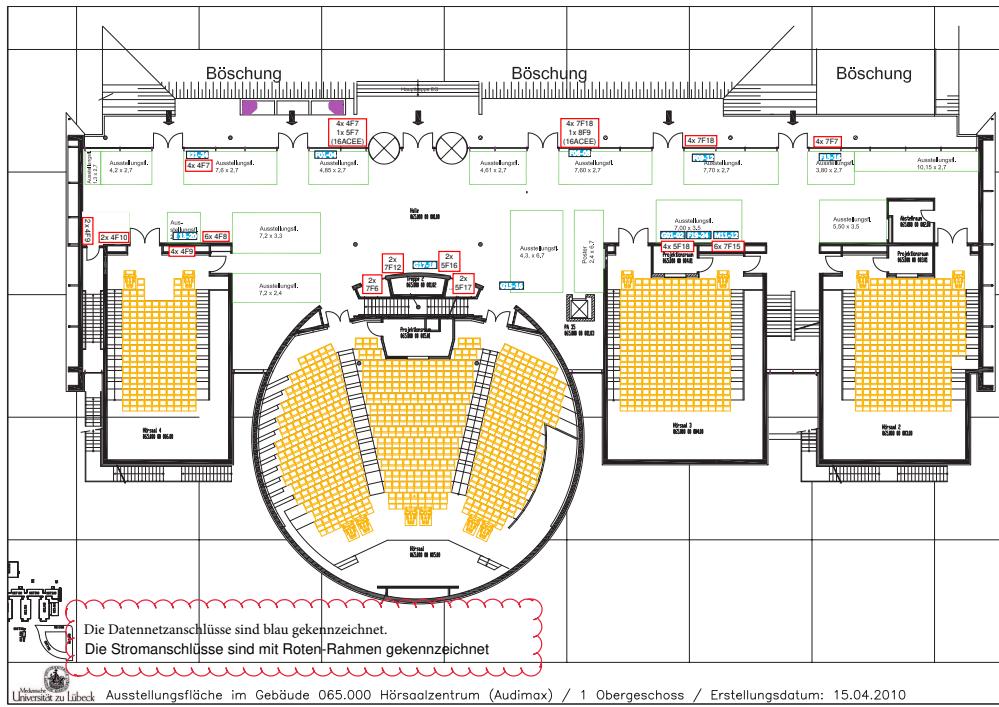
Catering

Catering for coffee breaks and lunch will be offered in the foyer of the Audimax building. Please note that no lunch will be provided on Wednesday, February 28th. You can eat in the cafeteria of the Mensa (ground floor), the Campus-Suite or the bistro of the University Hospital Schleswig-Holstein (UKSH) (see campus map).

WiFi

Free WiFi on the campus of the University of Lübeck is available via eduroam.

Floor plan



AG Meetings

2024-02-28 12:00 - 13:00

- AG Bayes-Methodik
C4-S02
- AG DAGStat
AM S3
- AG Ethik und Verantwortung
AM 2
[Zoom-Link](#)
- AG Nachwuchs
C4-S01
[Zoom-Link](#)
- AG Non-Clinical Statistics
AM S1
- AG Pharmazeutische Forschung
AM S2
- AG Statistik Stochastischer Prozesse
C4-S04
[Zoom-Link](#)
- AG Statistische Methoden in der Epidemiologie and AG Statistische Methoden in der Medizin
AM 1 (Empore)

2024-02-29 12:00 - 12:40

- AG Öffentlichkeitsarbeit
AM S2

2024-02-29 12:40 - 13:20

- AG Leitersitzung
AM S1

Guidance for presenters

Session time

The time allocated for a presentation is 20 minutes including QA (approx. 3 to 5 minutes)

Presentation format

- Your presentation should be in a PowerPoint (pptx) or PDF file format.
- No other formats (e.g. Excel, Word or Prezi) will be accepted.
- Keynote will not be supported because it cannot be played back on a Windows PC.
- Please export your presentation as MS Office 365, using filename extension ".pptx".
- All presentations will be projected in 16:9 format landscape (not 4:3) using the provided PowerPoint or PDF file on a Windows machine.
- All presentations must be in English followed by QA in English.
- Clearly identify your presentation with your presentation date and name in the filename (example: 29FEB24_John_Doe.pptx)

Presentation design recommendations

- Use high contrast colours
- Light text on dark background or vice versa
- Hyperlinks to external content such as websites cannot be supported

Submitting your presentation

Please submit your presentation within your conftool account no later than Sunday, 25th of February. We cannot accept USB sticks or files via email at the conference. If you want to update your presentation, you can upload a new file the same way the first one was uploaded, but no later than 25th of February. At the conference speakers should arrive in the session room 10 minutes before the session is due to start. Please introduce yourself to the session chairs. Be aware of the time for your presentation.

General Information

- A technician will be present in each session room to assist speakers in accessing the respective presentations
- All presentations will be controlled by the respective presenter from the lectern
- A microphone will be used for all presentations
- You will not be able to use a laser pointer since the conference is held hybrid and virtual attendees cannot see the laser pointer. If you need to point to anything in your presentation you have to use the pointer within PowerPoint or the computer's mouse in PDF files. Please get familiar with this before the conference.

Guidance for Posters

Poster session

We will have a session in which you can present your poster. This session starts with the 1-minute poster speed session. Afterwards, please come to your poster so that the other conference participants can come to you and ask questions. The sessions takes place on Thursday immediately after the keynote and before the lunch break. The posters will be exhibited the whole conference.

Poster Format: A0 portrait

Printing: Note that there is no printing service at the conference and you will need to bring your printed poster with you.

Speed Session Presentation Format

- Please submit your poster slide for the speed session within your conftool account no later than Sunday, 25th of February
- Please provide only one slide (no exceptions) in PowerPoint (.pptx) in 16:9 format
- Keynote will not be supported because it cannot be played back on a Windows PC. Please export presentation as MS Office 265, using filename extension ".pptx"
- We cannot accept any other slide format since we are combining all poster pitch slides into one presentations
- Do not use any animations

- The slide should include the following: poster title, author names, poster number and 3-4 headline messages to encourage attendees to visit your poster. You can also add your email address or LinkedIn so that attendees can contact you.
- You will find the poster number in conftool at the end of the title of your submitted abstract (number between 01 and 22)
- The slide must be in English
- Clearly identify your slide with your name, and poster number in the filename (example: John_Smith_poster_number.pptx)
- We cannot accept USB sticks or files via email at the conference

At the conference

- Come to the poster walls on the first day of the conference and hang up your poster.
- Please arrive at least 20 minutes before the poster session since the poster pitches are starting right after the keynote.
- Please introduce yourself to the session chairs and take a seat on the seat that is marked with your poster number.
- When starting the speed session after the keynote presentation, we will ask you to stand in line in front of the room for your poster pitch. The session will have a faster pace than other sessions. It is very important that you remember your poster number so that you are presenting the right slide.
- If you do not submit a slide, you can't present your poster. Let us know if you do not want to submit a slide.
- The presentation of the poster slides will be controlled by the session chair. You will not be able to use a laser pointer or any other pointers.
- A microphone will be used for all presentations.
- Remove your poster after the whole conference.

Social program

Town Hall reception

2024-02-28 18:30 - 20:30

Address: Breite Straße 62

Bus stop: Kohlmarkt

<https://www.visit-luebeck.com/old-town/town-hall>



The townhall reception will take place in the historical town hall of Lübeck and will start at 18:30. You will get access to the building by the main entrance in the "Breite Straße". Out of safety regulations it is not allowed to enter the building with large or wet items (e.g. rainwear, backpacks, carrier packs larger than A3). It will be possible to store some items at the entrance.

Get together with the Early Career Working Group (AG Nachwuchs)

2024-02-28 21:00

Join our Get Together with the Early Career Working Group (AG Nachwuchs) for an enjoyable evening! This event provides a relaxed opportunity to network with fellow researchers and working group members in a casual setting. We'll be meeting at a unique location that promises a fun atmosphere. More information will be provided during the Young Statisticians session on Wednesday. Don't miss this opportunity to mingle and network!

City tour with the night-watchman

2024-02-28 21:30 - 23:00

Meeting point: Holstentor, Holstentorplatz, 23552 Lübeck

Meeting time: Please arrive no later than 21:25

Tour language: English or German

Remaining tickets can be booked at the registration desk



Experience a historical journey through time

At night, the Hanseatic city of Lübeck appears in a completely new and mysterious splendor. Experience a historical journey through Lübeck's history by night with residents of the old town. Learn about life and work in the Middle Ages in a humorous way. Learn about customs and traditions, beliefs and superstitions.

Conference dinner

2024-02-29 19:00 - 22:00

Address: Breite Straße 2

Bus stop: Kohlmarkt (+ 700 m walk along Breite Strasse)



Location:

Traditional restaurant Schiffergesellschaft:

<https://schiffergesellschaft.de/>

Restaurants

Hours	Category	NAME	Address	Special
Lübsch (only in Lübeck)				
16-23		Im Alten Zolln	Mühlenstraße 93	Regional beer, fried potatoes
12-22		Schiffergesellschaft	Breite Straße 2	Medeval frescoes, shipt models, labskaus
12-22		Fangfrisch	An der Untertrave 51	Fish
17-23		Schabbelhaus	Mengstraße 48-50	Fine dining in historic surroundings
12-15		Suppentopf	Fleischhauerstraße 36	Only Tue-Wed 12-15 only soup
17-		Brauberger	Alfstraße 36	Pub with its own brewery
12-22		Kartoffelkeller	Koberg 8	Historic vault, rustic
Fine dining				
19-		Wullenwever	Beckergrube 71	Michelin Star, from Wed.
12-21		VAI	Hüxstraße 42	Special recommendation, not Wed, Thur
11:30-22		Miera	Hüxstraße 57	Upstairs fine, downstairs wine bar, pasta
17-23		Newport	Willy-Brandt-Allee 31	View on harbour
International				
17-21		Fermenti	Beckergrube 90	Pizza of 72-hours sour-dough
12-23		La Vigna	Hüxstraße 63	Pasta +
?		Sea Side	Kanalstraße 78	Italian on top of a riverboat with a view
11:30-22		Florentina	Pferdemarkt 19	Italian, centrally located
12-22		Onni	Mühlenstraße 54	Korean
17-21		Hana	Krähenstraße 34	Korean
12-23		Taj Mahal	Große Burgstraße 59	Indian
11:30-15, 17:30-22		Gandhi	Wakenitzufer 13	Indian with a view on Wakenitz
12-21:45		Namaste	Königstraße 26	Indian
11:30-22:30		Buddha Bowl	Schmiedestraße 24-26	Sushi, bowls
12-20		Nitsche Bowl	Fleischhauerstr. 67	Vegan
11:30-21:30		Ni-Vegan	Mühlentorstr. 68	Vietnamese, vegan
Café				
9-19		Niederegger	Breite Straße 89 u.a.	Not just marzipan
9-18		Uter	Fleischhauerstraße 62 u.a.	Vegan
9:30-18		Soulmade	Fleischhauerstraße 61	Scones, sorbets
8-19		Landwege	Kanalstraße 78 u. Am Brink 9	Organic vegetables

Streets with several restaurants

An der Obertrave
 An der Untertrave
 Beckergrube
 Große Burgstraße
 Fleischhauerstraße
 Hüxstraße
 Mühlenstraße

This list is purely subjective and guaranteed to be incomplete.
 Hours and addresses are from the internet.

Bars and nightlife

Pub	Description			Address
Allegro Bar S. Töbe	diversity	nice	casual	Ägidiensr. 27
Altstadt Bierhaus	traditional	original	cosy	Braunstr. 19
Angus	rockig	entertaining	rustic	Marlesgrube 53
Bar 112	casual	sociable	relaxed	Königstr. 112
Barcio	athmosphere	caribbean	big	An der Untertrave 106
Bierpub Kö 39	rustic	tiny	oldschool	Königstr. 39
Blauer Engel	studentic	casual	charming	Clemensstr. 8
Bolero	mexican	big	athmosphere	Breite Str. 1
Brauberger	original	rustic	traditional	Alfstr. 36
Buthmanns Bierstuben	cosy	long tradition	historic	Glockengießerstr. 3
Café & Bar Celona	big	casual	modern	Hafenstr. 1
Cafe Babel	entertaining	exotic	tiny	Marlesgrube 3-7
Cocoloco Bar	personal	tiny	cosy	Hüxstr. 121
Corner Bar	sports	moody	casual	An der Untertrave 88
Cosmos Pfeffermühle	tiny	sports	cosy	Engelsgrube 78
Dietrich's	original	diversity	sympathetic	An der Untertrave 108
Finnegan	irish	authentic	original	Mengstr. 42
Friends Café	modern	studentic	diverse	Mühlstr. 75
Funambules	athmosphere	relaxed	wassernah	An der Obertrave 18
Gang No. 56	charming	rustic	vielseitig	Dankwartsgrube 48
Heiners Bierstube	familiar	tiny	personal	Beckergrube 6
Hochwasser	sports	big	waterside	Dankwartsgrube 74
Im alten Zolln	traditional	big	historic	Mühlstr. 93-95
Jazz Café	modern	sports	big	Mühlstr. 62
Kandinsky	cosy	diversity	studentic	Fleischhauerstr. 89
Klostereck	music	casual	original	Dr.-Julius-Leber-Str. 35
La Havanna	caribbean	athmosphere	moody	Enelsgrube 56
Larry's Bar	charming	philippine	extraordinary	Marlesgrube 9-15
Lemmys Bier Pub	rustic	familiar	athmosphere	Dr.Julius-leber-Sr. 35
Location 25	original	cosy	loving	Hartengrube 25
Mac Thomas	irish	entertaining	cosy	An der Untertrave 95
MENGwirtschaft	tiny	familiar	authentic	Mengstr. 22
Morgana Pub	music	casual	sociable	Engelsgrube 94
My Way by Helena	authentic	personal	tiny	An der Obertrave 18
Nachtschwärmer	regional	original	athmosphere	Alfstr. 32
No. 12	studentic	moody	entertaining	Clemensstr. 12
Rauchfang	rustic	charming	casual	Hüxstr. 123
Sternschnuppe	entertaining	rustic	casual	Fleischhauerstr. 78
Tikiki Bar	hawaiianic	diversity	exotic	Wahmstr. 34
Tonfink	regional	music	cosy	Große Burgstr. 46
Torrios American Bar	big	stylish	diversity	Königstr. 36
Unklar	tiny	charming	casual	Clemensstr. 6
Weltwirtschaft	rustic	entertaining	tiny	Königstr. 16
Zur Drehbrücke 2	cosy	tiny	klassisch	An der Untertrave 62
Zur Loge 4	original	athmosphere	scecial	Mühlstr. 62
Zur Rose	cosy	personal	tiny	Rosenstr. 9-11

This list is subjective and incomplete. It may be outdated.

Conference Schedule

70th Biometric Colloquium

Date: Wednesday, 28/Feb/2024

Date: Thursday, 29/Feb/2024

9:00am -	Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update Location: AM 1 Chair: Matthias Schmid Chair: Carsten Oliver Schmidt	Panel Discussion (AG-Nachwuchs) Location: AM 2 Chair: Ina Dormuth Chair: Julia Christin Duda Navigating the Academic Odyssey: Early Career Challenges	Meta-Analysis 1 Location: AM S1 Chair: Tim Friede Chair: Gerta Rücker	Clinical Trials Location: AM S2 Chair: Thomas Asendorf Chair: Cornelia Ursula Kunz	High Dimensional Molecular Data Location: AM S3 Chair: Stefan Böhringer Chair: Michael Love
10:20am -	Break				
10:40am -	Keynote Ina Rondak Location: AM 1 Chair: Anne-Laure Boulesteix Chair: Anika Großhennig				
11:40am -	Poster Location: AM 1 Chair: Björn-Herken Laabs 11:40-12:05 h Speed Session				
12:00pm -	AG Öffentlichkeitsarbeit Location: AM S2				
12:40pm -	Lunchbreak		AG Leitersitzung Location: AM S1		
1:20pm -	Invited: Risk Prediction meets Causal Inference Location: AM 1 Chair: Vanessa Didelez Chair: Michael Schomaker	Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics Location: AM 2 Chair: Silvia Calderazzo Chair: Christian Röver	Machine Learning 1 Location: AM S1 Chair: Marvin N. Wright Chair: Inke Regina Koenig	Meta-Analysis 2 Location: AM S2 Chair: Tim Mathes Chair: Johannes Rauh	Application with Molecular Data Location: AM S3 Chair: Klaus Jung Chair: Silke Szymczak
2:40pm -	Break				
3:00pm -	IBS-DR: Nachwuchspreise Location: AM 1 Chair: Annette Kopp-Schneider Chair: Anne-Laure Boulesteix	Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics Location: AM 2 Chair: Dennis Dobler Chair: Jan Feifel	Machine Learning 2 Location: AM S1 Chair: Jörg Rahnenführer Chair: Björn-Herken Laabs	Special: IQWiG/IQTIG: Statistical analyses based on aggregated data Location: AM S2 Chair: Tim Friede Chair: Ralf Bender Chair: Jona Cederbaum	Genetic Epidemiology Location: AM S3 Chair: Silke Szymczak Chair: Amke Caliebe
4:20pm -	Break				
4:30pm -	IBS-DR: Mitgliederversammlung Location: AM 1				
6:00pm -	Break				
7:00pm -	Social Event: Conference Dinner Address: Breite Straße 2, 23552 Lübeck				
10:00pm					

Date: Friday, 01/Mar/2024

9:00am - 10:20am	Invited: Biometric Research Outside High-Income Settings Location: AM 1 Chair: Michael Schomaker Chair: Martje Rave	Statistics in Practice 1 Location: AM 2 Diagnostic Accuracy Studies: Basic and Advanced Statistical Methods	Time-to-Event 1 Location: AM S1 Chair: Annika Hoyer Chair: Sarah Friedrich	Agricultural and Biological Statistics 1 Location: AM S2 Chair: Klaus Jung Chair: Dörte Wittenburg	Multiple Testing Location: AM S3 Chair: Werner Brannath Chair: Edgar Brunner	School 1 (only for pupils) Location: AM S4 Workshop: Künstliche Intelligenz – Spielerisch lernen
10:20am - 10:40am	Break					
10:40am - 12:00pm	Invited: Statistical Issues in Animal Testing Location: AM 1 Chair: Bernd-Wolfgang Igl Chair: Michael Lauseker	Statistics in Practice 2 Location: AM 2 Diagnostic Accuracy Studies: Basic and Advanced Statistical Methods	Time-to-Event 2 Location: AM S1 Chair: Irene Schmidtmann Chair: Andreas Faldum	Agricultural and Biological Statistics 2 Location: AM S2 Chair: Hans-Peter Piepho Chair: Sabine K. Schnabel	Simulation Studies Location: AM S3 Chair: Maarten van Smeden Chair: Anne-Laure Boulesteix	School 2 (only for pupils) Location: AM S4 Workshop: Künstliche Intelligenz – Spielerisch lernen
12:00pm - 12:30pm	Group photo A group photo will be taken in front of the Audimax building.					
12:00pm - 1:00pm	Lunchbreak					
1:00pm - 2:20pm	Geburtstagsfeier (only German) Location: AM 1 Chair: Annette Kopp-Schneider Chair: Anne-Laure Boulesteix	Diagnostic Studies Location: AM 2 Chair: Annika Hoyer Chair: Maria Stark	Time-to-Event: Estimation Location: AM S1 Chair: Kathrin Möllenhoff Chair: Alexandra Erdmann	Time Series and Longitudinal Data Location: AM S2 Chair: Michael Lauseker Chair: Almond Stöcker	Nonparametric Methods Location: AM S3 Chair: Frank Konietzschke Chair: Ekkehard Glimm	
2:20pm - 2:40pm	Break					
2:40pm - 4:00pm	Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations Location: AM 1 Chair: Christian Staerk Chair: Andreas Mayr	Data Sharing and Reproducibility Location: AM 2 Chair: Matthias Schmid Chair: Max Westphal	Time-to-Event: Machine Learning Location: AM S1 Chair: Moritz Maximilian Berger Chair: Jan Beyersmann	Special: Statistics in Toxicology Location: AM S2 Chair: Bernd-Wolfgang Igl Chair: Tina Lang	Teaching Statistics Location: AM S3 Chair: André Scherag Chair: Cornelia Frömke	
4:00pm - 5:20pm	Closing Session & Keynote Maarten van Smeden Location: AM 1 Chair: Jan Beyersmann					
5:20pm - 6:00pm	Break					
6:00pm - 7:00pm	Vortrag für die Öffentlichkeit: "Sich selbst mit dem Smartphone behandeln: Wie gut funktioniert das?" Location: AM 1 Chair: Inke Regina Koenig Vortragender: Prof. Dr.med. Stefan Sauerland					

70th Biometric Colloquium

Date: Wednesday, 28/Feb/2024

	<p>generation via transformers</p> <p>Farhadyar, Kiana; Königs, Lukas; Binder, Harald</p>	<p>diagnostic accuracy studies</p> <p>Fierenz, Alexander; Rackow, Britta; Badpa, Mahnaz; Zapf, Antonia</p> <hr/> <p>3:40pm - 4:00pm</p> <p>Estimande in early phase studies, with an example in atopic dermatitis</p> <p>Klein, Stefan; Friedrichs, Frauke; Kunz, Michael</p>	<p>Berger, Moritz; Klein, Nadja; Schmid, Matthias</p> <hr/> <p>3:40pm - 4:00pm</p> <p>Comparative analysis of proportional odds models in a simulation study. A robust alternative to linear regression?</p> <p>Klinger, Andreas; Heinze, Georg; Dunkler, Daniela; Gregorich, Mariella; Kammer, Michael; Kraemmer, Daniel</p>	<p>Schürmeyer, Leonie; Schorning, Kirsten; Hengstler, Jan Georg</p> <hr/> <p>3:20pm - 3:40pm</p> <p>Probabilistic Approaches for Modeling Patient-Specific Effects of Antihypertensive Medication</p> <p>Hunsdieck, Berit; Elci, Eren; Ickstadt, Katja</p> <hr/> <p>3:40pm - 4:00pm</p> <p>The evaluation of clinical prediction models – methodological pitfalls illustrated with an application to peripartum depression</p> <p>Mayr, Andreas; Costa, Raquel; Martins, Rui; de Sousa, Bruno; Kneib, Thomas</p>
4:00pm - 4:20pm	Break			
4:20pm - 5:40pm	<p>Invited: Copula Regression for Time-to-Event Data</p> <p>Location: AM 1 Chair: Nadja Klein Chair: Andreas Mayr</p>	<p>Young Statisticians</p> <p>Location: AM 2 Chair: Stefanie Peschel Chair: Moritz Fabian Danzer</p>	<p>Missing Data</p> <p>Location: AM S1 Chair: Maren Vens Chair: Louis Rodrigue Macias</p>	<p>Preclinical Drug Development and Toxicology 2</p> <p>Location: AM S2 Chair: Richardus Vonk Chair: Katja Ickstadt</p>
	<p>4:20pm - 5:00pm</p> <p>Copula based Cox proportional hazards models for dependent censoring</p> <p>Van Keilegom, Ingrid; Deresa, Negera Wakgari</p>	<p>4:20pm - 4:40pm</p> <p>Sample size calculation and recalculation for non-inferiority trials in the ‘gold standard’ design using the studentized permutation test</p> <p>Schulz, Maxi; Mütze, Tobias; Friede, Tim</p>	<p>4:20pm - 4:40pm</p> <p>Recoverability of Causal Effects in a Longitudinal Study under Presence of Missing Data</p> <p>Holovchak, Anastasiia; Schomaker, Michael</p>	<p>4:20pm - 4:40pm</p> <p>Shift from Frequentist to Bayesian Dose-Response modelling</p> <p>Duda, Julia Christin; Wheeler, Matthew</p>
	<p>5:00pm - 5:20pm</p> <p>Boosting distributional copula regression models for bivariate time-to-event data</p> <p>Briseno Sanchez, Guillermo; Klein, Nadja; Mayr, Andreas; Groll, Andreas</p>	<p>4:40pm - 5:00pm</p> <p>A statistical deconvolution method with an application to secretomic and proteomic data</p> <p>Anarat, Akin; Krutmann, Jean; Schwender, Holger</p>	<p>4:40pm - 5:00pm</p> <p>Comparing propensity score methods combined with multiple imputation for controlling confounding: a case study on mantle cell lymphoma treatment regimens</p> <p>Gutmair, Katja; Cunningham, Nicholas; Silkenstedt, Elisabeth; Kluin-Nelemans, Hanneke; Dreyling, Martin; Villa, Diego; Hoster, Eva</p>	<p>4:40pm - 5:00pm</p> <p>Testing for similarity of multivariate mixed outcomes with application to efficacy-toxicity responses</p> <p>Hagemann, Niklas; Marra, Giampiero; Bretz, Frank; Möllenhoff, Kathrin</p>
	<p>5:20pm - 5:40pm</p> <p>A model-based boosting approach to deal with dependent censoring</p> <p>Strömer, Annika; Klein, Nadja; Mayr, Andreas</p>	<p>5:00pm - 5:20pm</p> <p>Evaluation of Time-To-Event-Endpoints in Oncology Biosimilar Trials – A Simulation Study</p> <p>Bohlken, Jan-Georg; Wright, Marvin N.; Hemmelmann, Claudia</p>	<p>5:00pm - 5:20pm</p> <p>Various approaches to deal with missing data when estimating causal effects with targeted</p>	<p>5:00pm - 5:20pm</p> <p>Prediction intervals for counted observations and their application in toxicological and medical</p>
				<p>5:00pm - 5:20pm</p> <p>When will I be cured from cancer? – An application of cure</p>

	5:20pm - 5:40pm Real world vs. controlled diagnosis: The example of response in oncology Miller, Carina; Beyersmann, Jan; Herpers, Matthias	maximum likelihood estimation Wiederkehr, Christoph Dominik	quality control charts Menssen, Max; Fneish, Firas; Schaarschmidt, Frank	models to cancer registry data from Schleswig-Holstein Baltus, Hannah; Schliemann, Antje; Schumann, Laura; Katalinic, Alexander; Eisemann, Nora
	5:20pm - 5:40pm Adaptive predictor-set linear model: an imputation-free method for linear regression prediction on datasets with missing values Planterose Jiménez, Benjamin; Kayser, Manfred; Vidaki, Athina; Caliebe, Amke	5:20pm - 5:40pm Methods of model selection for models with common parameters Gül, Onur; Schorning, Kirsten		

5:40pm Break

-
6:30pm

6:30pm **Social Event: Town Hall Reception**

Address: Breite Straße 62, 23552 Lübeck

-
8:30pm

9:30pm **Social Event: Night Watchman Tour**

Remaining tickets are only available at the conference registration desk in Lübeck.

-
11:00pm Meeting point: Holstentor, Holstentorplatz, 23552 Lübeck

Meeting time: Please arrive latest at 21:25.

Date: Thursday, 29/Feb/2024

9:00am - 10:20am	<p>Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update</p> <p>Location: AM 1 Chair: Matthias Schmid Chair: Carsten Oliver Schmidt</p>	<p>Panel Discussion (AG-Nachwuchs)</p> <p>Location: AM 2 Chair: Ina Dormuth Chair: Julia Christin Duda</p> <p>Navigating the Academic Odyssey: Early Career Challenges</p>	<p>Meta-Analysis 1</p> <p>Location: AM S1 Chair: Tim Friede Chair: Gerta Rücker</p>	<p>Clinical Trials</p> <p>Location: AM S2 Chair: Thomas Asendorf Chair: Cornelia Ursula Kunz</p>	<p>High Dimensional Molecular Data</p> <p>Location: AM S3 Chair: Stefan Böhringer Chair: Michael Love</p>
	<p>Evaluating biostatistical methods before use - each phase is important</p> <p>Heinze, Georg; Boulesteix, Anne-Laure; Kammer, Michael; Morris, Tim P.; White, Ian R.</p>		<p>Overall and landmark survival estimates by stage for patients with non-small cell lung cancer treated with either surgery alone or surgery plus adjuvant systemic anticancer treatment – an analysis based on German cancer registry data</p> <p>Baltus, Hannah; Labohm, Louisa; Katalinic, Alexander; Waldmann, Annika</p>	<p>"Randomize the first patient" - old, but still most important concept</p> <p>Großhennig, Anika; Koch, Armin; Beutel, Gernot; Theodor, Framke</p>	<p>Robust statistical detection of interaction effects in high-throughput sequencing data</p> <p>Stadler, Mara Stefanie; Müller, Christian L.</p>
	<p>Initial Data Analysis, Data Quality Assessments and proper Information Management to Improve the Transparency of Statistical Analyses</p> <p>Schmidt, Carsten Oliver; Lusa, Lara; Huebner, Marianne</p>		<p>A discrete time-to-event model for the meta-analysis of full ROC curves</p> <p>Stöve, Ferdinand Valentin; Tschammler, Claudia; Kuss, Oliver; Hoyer, Annika</p>	<p>A holistic approach to improve chronic kidney disease trials - unlocking the potential of hierarchical composite endpoints</p> <p>Tasto, Christoph</p>	<p>Using gene-set tests on expression data of mRNA targets to predict miRNAs involved during West Nile virus infections</p> <p>Boege, Franz Leonard; Ruff, Sergej; Selle, Michael; Hemandhar Kumar, Shamini; Jung, Klaus</p>
	<p>Recent developments in measurement error modelling</p> <p>Küchenhoff, Helmut</p>		<p>Identifying the risk of sample overlap in meta-analysis of registry-based studies</p> <p>Zhang, Zhentian; Mathes, Tim</p>	<p>Sample size calculation for cluster randomized trial with heterogeneous cluster size within cluster variances</p> <p>Franco Castiblanco, Ana Carolina; Brannath, Werner</p>	<p>Development of metabolomic risk scores for Alzheimer's Disease</p> <p>Tug, Timur; Liang, Donghai; Tan, Youran; Gearing, Marla; Levey, Allan I.; Lah, James J.; Wingo, Aliza P.; Wingo, Thomas S.; Ickstadt, Katja; Hüls, Anke</p>
	<p>The STRATOS Open Science panel</p> <p>Hoffmann, Sabine; Boulesteix, Anne-Laure; Dunkler, Daniela; Hornung, Roman; Kammer, Michael; Luijken, Kim</p>		<p>Addressing Challenges in Subgroup-Specific Treatment Effects and Aggregation Bias in Meta-Analysis</p> <p>Panaro, Renato Valladares; Röver, Christian; Friede, Tim</p>	<p>Optimal standardization as an alternative to matching using propensity scores</p> <p>Glimm, Ekkehard; Yau, Lillian</p>	<p>Evaluating deep learning models for cell detection and multi-class cell classification: a comparative analysis of metrics and solutions</p> <p>Pfrang, David; Ghete, Denisia-Tabita; Pontones, Martina; Metzler, Markus; Kock, Farina; Höfener, Henning; Westphal, Max</p>

10:20am	Break
10:40am	
10:40am	Keynote Ina Rondak Location: AM 1 Chair: Anne-Laure Boulesteix Chair: Anika Großhennig
11:40am	Wider access to more informative data and the key role of methodological experts in this endeavour
11:40am	Poster
	Location: AM 1
12:40pm	Chair: Björn-Herren Laabs 11:40-12:05 h Speed Session
	Comparison of genetic maps from different cattle breeds (Poster ID 01) <u>Wittenburg, Dörte; Ding, Xi; Melzer, Nina; Schwarzenbacher, Hermann; Seefried, Franz R.</u>
	<hr/>
	Bayesian borrowing using mixture prior: frequentist operating characteristics (Poster ID 02) <u>Weru, Vivien; Calderazzo, Silvia; Wiesenfarth, Manuel; Kopp-Schneider, Annette</u>
	<hr/>
	A multi-omics differentiation pattern analysis of CCI4-treated mice data (Poster ID 03) <u>Heiner, Jonas; Hengstler, Jan; Groll, Andreas</u>
	<hr/>
	Combining recurrent and terminal events into a composite endpoint may be problematic (Poster ID 04) <u>Liu, Xiaofei; Koch, Armin</u>
	<hr/>
	Quantification of prior impact in terms of effective sample size targeting test decisions (Poster ID 05) <u>Wiesenfarth, Manuel; Kopp-Schneider, Annette; Calderazzo, Silvia</u>
	<hr/>
	BACE2 polymorphisms are associated with memory impairment in a general population cohort SHIP-TREND (Poster ID 06) <u>Bonk, Sarah; Kirchner, Kevin; Garvert, Linda; Völzke, Henry; Grabe, Hans Jörgen; Van der Auwera, Sandra</u>
	<hr/>
	Paradoxes of Inter-Rater Reliability Measures of skewed ordinal Data (Poster ID 07) <u>Mönch, Maximilian; Grittner, Ulrike; Unger, Nina; Keller, Theresia; Breitenstein, Caterina; Schulze, Daniel; Pigorsch, Mareen</u>
	<hr/>
	Performance of different Interpolation Methods on self-reported symptoms in the context of Digital Allergology (Poster ID 08) <u>Hernandez-Toro, Camilo Jose; Grittner, Ulrike; Caminiti, Lucia; Charpin, Denis; Delgado, Luis; Dramburg, Stephanie; Kalpaklioglu, Fusun; Nieto, Antonio; Papadopoulos, Nikolaos G.; Pelosi, Simone; Potapova, Ekaterina; Priftanji, Alfred; Travaglini, Alessandro; Tripodi, Salvatore; Matricardi, Paolo Maria</u>
	<hr/>
	A MAP prior approach for piecewise constant hazards and competing risks (Poster ID 09) <u>Stemke, Alexander; Sailer, Oliver</u>
	<hr/>
	Using early or baseline data in a trial with missingness in a continuous primary endpoint (Poster ID 10) <u>Basu, Joydeep; Stallard, Nigel</u>
	<hr/>
	Simultaneous inference of multiple binary endpoints in biomedical research: small sample properties of multiple marginal models and a resampling approach (Poster ID 11) <u>Budig, Sören; Schaarschmidt, Frank</u>

Kirstine.jl: A Julia Package for Bayesian Optimal Design of Experiments (Poster ID 12)
Sandig, Ludger

{cases} - an R package for simultaneous evaluation of multiple diagnostic tests or prediction models regarding co-primary endpoints sensitivity and specificity (Poster ID 13)
Westphal, Max; Zapf, Antonia

{mldesign} - an R package to conduct meaningful data splitting in applied machine learning (Poster ID 14)
Westphal, Max

Multivariate modelling of water quality parameters in nigeria (Poster ID 15)
Dosumu, Ebun Adegbola

A unified parametric approach to the estimation of dependence and marginal distributions in bivariate competing risks survival data (Poster ID 16)
Zhang, Hyun-Soo; Jung, Inkyung; Nam, Chung Mo

Sample size re-examination for clinical trials with survival endpoints (Poster ID 17)
Dormuth, Ina; Liu, Tiantian; Chen, Zijian; Ditzhaus, Marc; Pauly, Markus; Xu, Jin

Improvement in population-based survival in cutaneous malignant melanoma after the introduction of new therapies (Poster ID 18)
Eisemann, Nora; Schumann, Laura; Baltus, Hannah; Labohm, Louisa; Kraywinkel, Klaus; Katalinic, Alexander

"Initiative Biokybernetik" - ten years later (Poster ID 19)
Mau, Jochen

Structured evaluation of drug prescription data from Schleswig-Holstein in a networked big data context (Poster ID 20)
Schuster, Reinhard; Emcke, Timo; Burmester, Mareike

High Degree of Agreement but Low Measures - Problems of Inter-Rater Reliability Measures for Unbalanced Ordinal Data (Poster ID 21)
Grittner, Ulrike; Schulze, Daniel; Unger, Nina; Keller, Theresa; Breitenstein, Caterina; Pigorsch, Mareen; Mönch, Maximilian

Time-dependent change in risk through an exposure and possible estimands of interest: A simulation study using two clinically motivated examples (Poster ID 23)
Meiszl, Katharina; Tokic, Marianne C.; Timmesfeld, Nina

12:00pm -	AG Öffentlichkeitsarbeit Location: AM S2
12:40pm	Lunchbreak
12:40pm -	AG Leitersitzung Location: AM S1
1:20pm	
1:20pm -	Invited: Risk Prediction meets Causal Inference
2:40pm	Invited: Innovative Clinical Trial Designs: Estimands and
	Machine Learning 1 Location: AM S1 Chair: Marvin N. Wright
	Meta-Analysis 2 Location: AM S2
	Application with Molecular Data Location: AM S3

<p>Location: AM 1 Chair: Vanessa Didelez Chair: Michael Schomaker</p> <p>1:20pm - 2:00pm Causal blind spots in risk-based medical decision making <u>van Geloven, Nan</u></p>	<p>Operating Characteristics Location: AM 2 Chair: Silvia Calderazzo Chair: Christian Röver</p> <p>1:20pm - 2:00pm Estimands and Complex Innovative Designs <u>Bretz, Frank</u></p>	<p>Chair: Inke Regina Koenig</p> <p>1:20pm - 1:40pm Achieving explainable machine learning by functional decomposition of black-box models into explainable predictor effects <u>Köhler, David; Rügamer, David; Schmid, Matthias</u></p>	<p>Chair: Tim Mathes Chair: Johannes Rauh</p> <p>1:20pm - 1:40pm Trials and triangles - The geometric interpretation of multi-arm studies in network meta-analysis <u>Rücker, Gerta</u></p>	<p>Chair: Klaus Jung Chair: Silke Szymczak</p> <p>1:20pm - 1:40pm Companion diagnostics in oncology clinical trials: lessons from the practice <u>Weispfenning, Anke; Descamps, Tine</u></p>
<p>2:00pm - 2:20pm Methods for obtaining causal predictions and quantifying associated uncertainty <u>DiazOrdaz, Karla</u></p>	<p>2:00pm - 2:20pm Multiplicity Issues in Platform Trials – adjusting for what? Looking forward and benefiting from the past <u>Koenig, Franz; Posch, Martin; Zehetmayer, Sonja</u></p>	<p>1:40pm - 2:00pm Red-light crossing or bank robbery? On the bias in model performance estimates resulting from incorrect optimization of algorithm and preprocessing hyperparameters <u>Sauer, Christina; Hanßum, Luzia; Hodiamont, Farina; Bausewein, Claudia; Boulesteix, Anne-Laure; Ullmann, Theresa</u></p>	<p>1:40pm - 2:00pm A comparison of different software tools to support systematic reviews - considering the size of the training dataset and the machine learning methods used. <u>Schröder, Christin; Brewer, Madisen</u></p>	<p>1:40pm - 2:00pm Genome-wide association studies on hedonic eating behaviour <u>Schliemann, Antje; König, Inke R.; von Holt, Björn-Herzen</u></p>
<p>2:20pm - 2:40pm Causal effect estimation and risk prediction - the example of screening colonoscopy and colorectal cancer <u>Braitmaier, Malte; Didelez, Vanessa</u></p>	<p>2:20pm - 2:40pm Modeling strategies for analysing platform trials with non-concurrent controls <u>Bofill Roig, Marta</u></p>	<p>2:00pm - 2:20pm Using background knowledge from previous studies in model building: the good, the bad and the ugly <u>Hafermann, Lorena; Heiko, Becher; Herrmann, Carolin; Klein, Nadja; Kammer, Michael; Rauch, Geraldine; Heinze, Georg</u></p>	<p>2:00pm - 2:20pm Path-Based Approach for Detecting and Assessing Inconsistency in Network Meta-Analysis: A Novel Method <u>Rajabzadehtahmasebi, Noosheen; Papakonstantinou, Theodoros; Nikolakopoulou, Adriani</u></p>	<p>2:00pm - 2:20pm Partial interaction analysis in case-only defined clusters for high-dimensional biomarkers <u>Böhringer, Stefan; Maarseveen, Tjardo D.; Knevel, Rachel</u></p>
	<p>2:00pm - 2:20pm Random forests more data hungry than logistic regression models? A confirmatory, large-scale, real-data study on the link between the number of events per variable and <u>Schulz, Maxi; Kramer, Malte; Kuss, Oliver; Mathes, Tim</u></p>	<p>2:20pm - 2:40pm Is the early microbiome linked to childhood obesity? – A network perspective <u>Peschel, Stefanie; Depner, Martin; von Mutius, Erika; Boulesteix, Anne-Laure; Müller, Christian L.</u></p>		

			prediction performance	
			<u>Lange, F. Julian D.; Boulesteix, Anne-Laure</u>	
2:40pm - 3:00pm	Break			
3:00pm - 4:20pm	IBS-DR: Nachwuchspreise Location: AM 1 Chair: Annette Kopp-Schneider Chair: Anne-Laure Boulesteix Infusing structural assumptions into dimension reduction for single-cell RNA sequencing data to identify small gene sets <u>Brunn, Niklas</u>	Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics Location: AM 2 Chair: Dennis Dobler Chair: Jan Feifel 3:00pm - 3:40pm Counting processes in stochastic epidemic models: the link with survival analysis <u>Putter, Hein; Goeman, Jelle; Wallinga, Jacco</u>	Machine Learning 2 Location: AM S1 Chair: Jörg Rahnenführer Chair: Björn-Herren Laabs 3:00pm - 3:20pm Adversarial random forests for imputing missing values <u>Golchian, Pegah; Kapar, Jan; Blesch, Kristin; Watson, David S.; Wright, Marvin N.</u>	Special: IQWiG/IQTIG: Statistical analyses based on aggregated data Location: AM S2 Chair: Tim Friede Chair: Ralf Bender Chair: Jona Cederbaum 3:00pm - 3:20pm Performing subgroup analyses in HTA applications <u>Grouven, Ulrich; Skipka, Guido</u>
	Evaluation of Index-based Response-adaptive Randomization Procedures in Clinical Trials <u>Drescher, Sonja; Kunz, Cornelia Ursula; Walther, Andrea</u>	3:40pm - 4:00pm Identifying alert concentrations using a model-based bootstrap approach <u>Möllenhoff, Kathrin; Schorning, Kirsten; Kappenberg, Franziska</u>	3:20pm - 3:40pm Confidence intervals for tree-structured varying coefficients based on parametric bootstrap <u>Spuck, Nikolai; Schmid, Matthias; Berger, Moritz</u>	3:20pm - 3:40pm Using meta-regression for investigating subgroups in a meta-analysis <u>Röver, Christian; Kramer, Malte; Friede, Tim</u>
	Functional Additive Models on Manifolds of Planar Shapes and Forms <u>Stöcker, Almond</u>	4:00pm - 4:20pm Resampling-based inference for the average treatment effect in time-to-event data <u>Rühl, Jasmin; Friedrich, Sarah</u>	3:40pm - 4:00pm On the handling of method failure in comparison studies <u>Wünsch, Milena; Boulesteix, Anne-Laure</u>	3:40pm - 4:00pm Statistical analysis of aggregate results of health care providers <u>Rauh, Johannes</u>
	A connection between survival multistate models and causal inference for external treatment interruptions <u>Erdmann, Alexandra; Loos, Anja; Beyersmann, Jan</u>		4:00pm - 4:20pm An empirical study of the performance of semi-supervised machine learning methods in systematic review tools for abstract and title screening <u>Kutil, Raoul; Borgelt, Christian; Hirlaender, Simon; Zimmermann, Georg</u>	4:00pm - 4:20pm Volume-outcome analyses based on aggregated data <u>Gutzeit, Maurilio; Rauh, Johannes; Cederbaum, Jona</u>
4:20pm - 4:30pm	Break			
4:30pm - 6:00pm	IBS-DR: Mitgliederversammlung Location: AM 1			Genetic Epidemiology Location: AM S3 Chair: Silke Szymczak Chair: Amke Caliebe 3:00pm - 3:20pm Clinical utility of polygenic scores: A critical 2023 appraisal <u>Koch, Sebastian; Schmidtko, Jörg; Krawczak, Michael; Caliebe, Amke</u>
				3:20pm - 3:40pm Tools for predicting the effects of genetic variants: a systematic review and practical guide <u>Riccio, Cristian; Jansen, Max Louis; Ziegler, Andreas</u>
				3:40pm - 4:00pm Detecting interactions in High Dimensional Data using Cross Leverage Scores <u>Teschke, Sven; Ickstadt, Katja; Munteanu, Alexander; Schikowski, Tamara</u>
				4:00pm - 4:20pm Estimating sparse graphical models in high dimensions <u>Foraita, Ronja; Mose, Kristof; Hanke, Moritz; Didelez, Vanessa</u>

6:00pm Break

-

7:00pm

7:00pm Social Event: Conference Dinner

Address: Breite Straße 2, 23552 Lübeck

10:00pm

Date: Friday, 01/Mar/2024

<p>9:00am - 10:20am</p> <p>Invited: Biometric Research Outside High-Income Settings</p> <p>Location: AM 1 Chair: Michael Schomaker Chair: Martje Rave</p> <p>9:00am - 9:40am</p> <p>Assessment of the treat-all public health approach in Eswatini – challenges and opportunities</p> <p><u>Kerschberger, Bernhard;</u> <u>Ciglenecki, Iza</u></p> <hr/> <p>9:40am - 10:00am</p> <p>The South African HIV Cancer Match (SAM) Study: a unique resource for cancer research among people with HIV in South Africa</p> <p><u>Rohner, Eliane;</u> <u>Olago, Victor;</u> <u>Ruffieux, Yann;</u> <u>Metekoua, Carole;</u> <u>Tombe-Nyahuma,</u> <u>Tinashe;</u> <u>Mwansa-Kambafwile, Judith;</u> <u>Egger, Matthias;</u> <u>Muchengeti,</u> <u>Mazvita</u></p> <hr/> <p>10:00am - 10:20am</p> <p>Interplay between Applied Biostatistics and Public Health in Humanitarian Settings</p> <p><u>Luque Fernandez, Miguel Angel</u></p>	<p>Statistics in Practice 1</p> <p>Location: AM 2</p> <p>Diagnostic Accuracy Studies: Basic and Advanced Statistical Methods</p> <p>Time-to-Event 1</p> <p>Location: AM S1 Chair: Annika Hoyer Chair: Sarah Friedrich</p> <p>9:00am - 9:20am</p> <p>On historically controlled survival trials</p> <p><u>Feld, Jannik;</u> <u>Danzer, Moritz;</u> <u>Faldum, Andreas;</u> <u>Schmidt, Rene</u></p> <hr/> <p>9:20am - 9:40am</p> <p>Conditional survival of younger patients with mantle cell lymphoma: novel insights into disease course and dynamic prediction by baseline and time-dependent prognostic factors</p> <p><u>Jiang, Limmiao;</u> <u>Dreyling, Martin;</u> <u>Hermine, Olivier;</u> <u>Schumacher, Martin;</u> <u>Hoster, Eva</u></p> <hr/> <p>9:40am - 10:00am</p> <p>The challenge of time-to-event analysis for multiple events: Which method of analysis can we trust?</p> <p><u>Schmeller, Sandra;</u> <u>Erdmann, Alexandra;</u> <u>Beyermann, Jan;</u> <u>Ozga, Ann-Kathrin</u></p> <hr/> <p>10:00am - 10:20am</p> <p>Unleashing the power of adjusted survival curves: Introducing the</p>	<p>Agricultural and Biological Statistics 1</p> <p>Location: AM S2 Chair: Klaus Jung Chair: Dörte Wittenburg</p> <p>9:00am - 9:20am</p> <p>A random-regression model for analyzing the genetic variance of litter weight in mice</p> <p><u>Reinsch, Norbert;</u> <u>Ding, Xi;</u> <u>Jahnel, Ricarda</u></p> <hr/> <p>9:20am - 9:40am</p> <p>Optimising sampling efforts in hierarchical Lincoln-Petersen experiment: Towards precise population size estimation</p> <p><u>Chin, Su Na;</u> <u>Overstall, Antony;</u> <u>Böhning, Dankmar</u></p> <hr/> <p>9:40am - 10:00am</p> <p>Optimizing Genetic Gain and Haplotype Diversity in Genomic Selection through Mendelian Sampling-Based Similarity Matrices</p> <p><u>Musa, Abdulraheem Arome;</u> <u>Reinsch, Norbert</u></p> <hr/> <p>10:00am - 10:20am</p> <p>Asymptotic online familywise error rate control for dependent test statistics</p> <p><u>Jankovic, Vincent;</u> <u>Fischer, Lasse;</u> <u>Brannath, Werner</u></p>	<p>Multiple Testing</p> <p>Location: AM S3 Chair: Werner Brannath Chair: Edgar Brunner</p> <p>9:00am - 9:20am</p> <p>Surviving the multiple testing problem: RMST-based tests in general factorial designs</p> <p><u>Munko, Merle;</u> <u>Ditzhaus, Marc;</u> <u>Dobler, Dennis;</u> <u>Genuneit, Jon</u></p> <hr/> <p>9:20am - 9:40am</p> <p>Informative simultaneous confidence intervals for graphical test procedures</p> <p><u>Scharpenberg, Martin;</u> <u>Brannath, Werner</u></p> <hr/> <p>9:40am - 10:00am</p> <p>Multiple contrast testing procedures for semiparametric MANCOVA</p> <p><u>Baumeister, Marlène;</u> <u>Thiel, Konstantin Emil;</u> <u>Matits, Lynn;</u> <u>Kolassa, Iris-Tatjana;</u> <u>Pauly, Markus;</u> <u>Zimmermann, Georg</u></p> <hr/> <p>10:00am - 10:20am</p> <p>Asymptotic online familywise error rate control for dependent test statistics</p> <p><u>Jankovic, Vincent;</u> <u>Fischer, Lasse;</u> <u>Brannath, Werner</u></p>	<p>School 1 (only for pupils)</p> <p>Location: AM S4</p> <p>Workshop: Künstliche Intelligenz – Spielerisch lernen</p>
---	---	---	--	---

			adjustedCurves R-package Denz, Robin; Timmesfeld, Nina	10:00am - 10:20am	Challenges in statistical consulting for Animal Science Schnabel, Sabine K.	
10:20am	Break					
10:40am						
10:40am	Invited: Statistical Issues in Animal Testing Location: AM 1 Chair: Bernd-Wolfgang Igl Chair: Michael Lauseker	Statistics in Practice 2 Location: AM 2 Diagnostic Accuracy Studies: Basic and Advanced Statistical Methods	Time-to-Event 2 Location: AM S1 Chair: Irene Schmidtmann Chair: Andreas Faldum	10:40am - 11:00am	Agricultural and Biological Statistics 2 Location: AM S2 Chair: Hans-Peter Piepho Chair: Sabine K. Schnabel	Simulation Studies Location: AM S3 Chair: Maarten van Smeden Chair: Anne-Laure Boulesteix
12:00pm	10:40am - 11:20am Statistical Review of Animal trials in Ethics Committees <u>Piper, Sophie K.; Konietzschke, Frank; Zocholl, Dario; Röhle, Robert; Tölch, Ulf</u>		Adaptive redesigning of combination testing procedures in survival analysis <u>Danzer, Moritz Fabian; Dormuth, Ina</u>	10:40am - 11:00am	Promises and limitations of applying structural equation modelling techniques from psychology in other disciplines exemplified in geoecology <u>Rieger, Alex; de Haan, Jan; Schneider, Anne-Kathrin; Schibalski, Anett; Eggert, Frank; Schröder-Esselbach, Boris</u>	10:40am - 11:00am Statistical Plasmode Simulations - Potentials, Challenges and Recommendations <u>Schreck, Nicholas; Slyko, Alla; Saadati, Maral; Benner, Axel</u>
	11:20am - 11:40am Sample size justification in preclinical animal studies <u>Wilcke, Juliane C.; Boulesteix, Anne-Laure</u>		A new parametric accelerated failure time model for semi-competing risk data <u>Dineva, Antoniya; Kuss, Oliver; Hoyer, Annika</u>	11:00am - 11:20am		11:00am - 11:20am When is Plasmode simulation superior to parametric simulation? <u>Stolte, Marieke; Rahnenführer, Jörg; Bommert, Andrea</u>
	11:40am - 12:00pm 3R initiatives in preclinical toxicology and research <u>Lang, Tina; Vaas, Lea</u>		Dynamic prediction of the risk of preeclampsia – Landmarking with continuous time-dependent covariates in left-truncated competing risks data <u>Stegherr, Regina; Aigner, Annette; Verlohren, Stefan</u>	11:20am - 11:40am	Analysis of greenhouse gas emission and the effect on rainfall outcomes in nigeria <u>Dosumu, Ebun Adegbola</u>	11:20am - 11:40am On the role of benchmarking data sets and simulations in method comparison studies <u>Friedrich, Sarah; Friede, Tim</u>
					11:20am - 11:40am Antibiotic resistance in pigs - analysis of the VetAmUR data to <u>Callahan, Patrick; Boulesteix, Anne-Laure</u>	11:40am - 12:00pm Translating methodological simulation studies into practice: a reproducible application

				investigate the influence of antibiotic use on the temporal development of resistance in pig farming
			<u>Mers, Fiona;</u> Bonnelett, Clarissa; Kreienbrock, Lothar; Rehberg, Betty; Ickstadt, Katja; Tug, Timur	
12:00pm	Group photo	A group photo will be taken in front of the Audimax building.		
-				
12:30pm	Lunchbreak			
-				
12:00pm				
-				
1:00pm				
1:00pm	Geburtstagsfeier (only German)	Diagnostic Studies	Time-to-Event: Estimation	Time Series and Longitudinal Data
-	Location: AM 1	Location: AM 2	Location: AM S1	Nonparametric Methods
2:20pm	Chair: Annette Kopp-Schneider Chair: Anne-Laure Boulesteix	Chair: Annika Hoyer Chair: Maria Stark	Chair: Kathrin Möllenhoff Chair: Alexandra Erdmann	Location: AM S2
	Die wilden Achtziger und die goldenen Neunziger <u>Rücker, Gerta;</u> <u>Pigeot, Iris;</u> <u>Kreienbrock, Lothar</u>	Unblinded sample size re-estimation in diagnostic test accuracy studies <u>Köster, Denise;</u> Hoyer, Annika; Zapf, Antonia	1:00pm - 1:20pm Implication of the choice of time scales in survival analysis <u>Vilsmeier, Judith;</u> Büchele, Gisela; Rehm, Martin; Rothenbacher, Dietrich; Beyersmann, Jan	1:00pm - 1:20pm On detecting change points in unlabelled multivariate time series <u>Balestra,</u> <u>Chiara;</u> Li, Bin; Mayr, Andreas; Müller, Emmanuel
	AG Nachwuchs: Rückblicke, Einblicke, Ausblicke	1:20pm - 1:40pm Comparing methods to handle missing values in the index test in diagnostic studies - a simulation study Stahlmann, Katharina; Kellerhuis, Bas; Zapf, Antonia	1:20pm - 1:40pm Hazards: key quantities for analysis, interpretation and understanding of time-to-event data Beyersmann, Jan; Schmoor, Claudia; Schumacher, Martin	1:00pm - 1:20pm Inference for Random Effects in Nonparametric Repeated Measures Designs with Missing Data via Randomization Amro, Lubna; <u>Dobler, Dennis;</u> Kuhn, Jörg-Tobias
	AG Nachwuchs, <u>Verschiedene</u> <u>Mitglieder</u>	1:40pm - 2:00pm Covariate adjustment, factorial designs and clustered data in diagnostic accuracy studies Weber, Philipp; Kramer	1:40pm - 2:00pm Piecewise constant hazard estimation with the fused lasso Rosenbaum, Manuel; Beversmann	1:20pm - 1:40pm NANCOVA: Nonparametric Analysis of Covariance for Rare Disease Research <u>Thiel, Konstantin</u> Emil; Sattler, Paavo; Bathke, Arne C.; Zimmermann, Georg
				1:40pm - 2:00pm Sample size planning for rank-based multiple contrast tests <u>Pöhlmann, Anna;</u> Brunner, Edgar; Konietzschke, Frank
				2:00pm - 2:20pm A non-parametric proportional risk

	Katharina; Zapf, Antonia	Jan; Vogt, Michael	model specifications	model to assess a treatment effect in an application to long-term carcinogenicity assays	
	2:00pm - 2:20pm	2:00pm - 2:20pm	Häckl, Sebastian; Koch, Armin; Lasch, Florian	Ameis, Lucia; Kuß, Oliver; Hoyer, Annika; Möllenhoff, Kathrin	
	Power and sample size estimation for comparing diagnostic methods with imperfect reference standards	Estimation Within The Responder Stratified Exponential Survival Model	Kilian, Samuel; Kieser, Meinhard	Function-on-Scalar Regression (FoSR) with Wavelet Basis Functions for the analysis of Periodic Time-Series	
	Paul, Roman Harald; Othman, Ahmed; Altmann, Sebastian; Schmidtmann, Irene		Neumann, Konrad		
2:20pm	Break				
2:40pm					
2:40pm - 4:00pm	Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations Location: AM 1 Chair: Christian Staerk Chair: Andreas Mayr 2:40pm - 3:20pm Power of inclusion: Enhancing polygenic prediction with admixed individuals Tanigawa, Yosuke; Kellis, Manolis	Data Sharing and Reproducibility Location: AM 2 Chair: Matthias Schmid Chair: Max Westphal 2:40pm - 3:00pm Federated Generalized Additive Models for Location, Scale and Shape in DataSHIELD Swenne, Annika; Intemann, Timm; Pigeot, Iris	Time-to-Event: Machine Learning Location: AM S1 Chair: Moritz Maximilian Berger Chair: Jan Beyermann 2:40pm - 3:00pm A Large-Scale Neutral Comparison Study of Survival Models Burk, Lukas; Bischl, Bernd; Bender, Andreas; Wright, Marvin; Lang, Michel; Sonabend, Raphael	Special: Statistics in Toxicology Location: AM S2 Chair: Bernd-Wolfgang Igl Chair: Tina Lang 2:40pm - 3:00pm How to benefit from statistics in toxicology Rahnenführer, Jörg; Kappenberg, Franziska	Teaching Statistics Location: AM S3 Chair: André Scherag Chair: Cornelia Frömke 2:40pm - 3:00pm The Power of Data: A Story by global Biostatistics and Data Sciences (gBDS) Kunz, Cornelia Ursula
3:20pm - 3:40pm	Bridging the performance gap for underrepresented populations: How to account for population structure in polygenic risk modelling? Klinkhammer, Hannah; Maj, Carlo; Staerk, Christian; Krawitz, Peter; Mayr, Andreas	3:00pm - 3:20pm The CodeClub at the MPI of Psychiatry – better code and better reproducibility Hagenberg, Jonas; Karlbauer, Vera N.; Dieckmann, Linda	3:00pm - 3:20pm Integrating Fine & Gray's Subdistribution Weights into Random Survival Forests for Competing Event Analysis Behning, Charlotte; Bigerl, Alexander; Wright, Marvin; Berger, Moritz; Schmid, Matthias	3:00pm - 3:20pm Virtual Control Groups in Toxicity Studies Vaas, Lea A.I.; Gurjanov, Alexander; Ulbrich, Hannes-Friedrich; Kreuchwig, Annika; Steger-Hartmann, Thomas	3:00pm - 3:20pm The Alienator App: Unveiling Data Science in Clinical Trials to a Lay Audience Andersen, Lars
3:20pm - 3:40pm	Simple tips for writing and publishing clear code to ensure reproducible results Hornung, Roman		3:20pm - 3:40pm Robust and simple experimental designs for fitting dose response A random forest pseudo-	3:20pm - 3:40pm If four programs do the same thing, it's still not the same... (A model building catalog) Sahlmann, Jörg	3:40pm - 4:00pm Overview over 5 years of Academia meets Industry Workshop Lehn, Annette; Schulte-Göbel, Marlene; Jahn, Antje; Stucke-

	3:40pm - 4:00pm Boosting European and Multi-ancestry polygenic models: an analysis across different phenotypes of the UK Biobank <u>Mai, Carlo;</u> Klinkhammer, Hannah; Staerk, Christian; Krawitz, Peter; Mayr, Andreas	3:40pm - 4:00pm Addressing researcher degrees of freedom through adjustment for the multiplicity of analysis strategies <u>Mandl,</u> <u>Maximilian M;</u> Becker-Pennrich, Andrea S; Hinske, Ludwig C; Hoffmann, Sabine; Boulesteix, Anne-Laure	value approach for modeling restricted mean survival times <u>Schenk, Alina;</u> Basten, Vanessa; Schmid, Matthias	curves in toxicology <u>Holland-Letz,</u> <u>Tim</u>	3:40pm - 4:00pm How to benefit from high-dimensional expression data in toxicology <u>Kappenberg,</u> <u>Franziska;</u> Schorning, Kirsten; Rahnenführer, Jörg	Straub, Kathrin; Beyersmann, Jan; Lanius, Vivian; Lang, Tina; Scharpenberg, Martin; Brannath, Werner; Kunz, Cornelia Ursula
4:00pm -	Closing Session & Keynote Maarten van Smeden Location: AM 1 Chair: Jan Beyersmann					
5:20pm	Clinical prediction modeling in the era of AI: a blessing and a curse					
5:20pm -	Break					
6:00pm						
6:00pm -	Vortrag für die Öffentlichkeit: "Sich selbst mit dem Smartphone behandeln: Wie gut funktioniert das?" Location: AM 1 Chair: Inke Regina Koenig					
7:00pm	Vortragendener: Prof. Dr.med. Stefan Sauerland					

Tutorials

Multiple endpoints and prioritized outcomes - Nonparametric analysis methods using generalized pairwise comparisons

2024-02-28 09:00 - 12:00, AM S3

Lecturers

Edgar Brunner (University Medical Center Göttingen) and Werner Brannath (University of Bremen)

Learning objectives

- The role of multiple endpoints in clinical trials and medical research
- Understanding Generalized Pairwise Comparisons (GPC) and related effect measures
- Relation of GPC to the Mann-Whitney test and other non-parametric effects

Target group

Experienced and early career biostatisticians, clinical trial experts

Prerequisites

Some knowledge in clinical trials and non-parametric statistics

Format

Hybrid

Introduction to Causal Inference and Target Trial Emulation

2024-02-28 09:00 - 12:00, AM S1

Lecturers

Vanessa Didelez and Malte Braithmaier (both BIPS, Bremen)

Learning objectives

- Be able to recognise avoidable sources of bias in naive studies using observational data
- Become aware of principled techniques to address the above issues
- Acquire a basic understanding of TTE that will facilitate studying the more advanced literature

Target group

PhD students, Post-docs (and more seniors) in statistics and related fields who want to learn about the causal analysis of observational (real-world) data, or even about addressing intercurrent events in RCTs

Prerequisites

A willingness to actively participate in discussions

Format

Hybrid

An introduction to estimands and estimand-aligned estimation

2024-02-28 09:00 - 12:00, AM S2

Lecturers Tobias Mütze (Novartis, Basel) and Tim Friede (University Medical Center Göttingen)

Learning objectives

- Understand the basic concepts related to the ICH E9 (R1) addendum including estimands and their attributes as well as intercurrent events and their handling strategies
- Identify an appropriate primary analysis method that targets the estimand of interest, fully alignes with the ICH E9 addendum
- Understand how to change assumptions made for the primary analysis in sensitivity analyses
- Implement appropriate analyses and sensitivity analyses

Target group

Statisticians, in particular those who have limited experience with the estimand framework

Prerequisites

- R version 4.1 or higher
- R packages: tidyverse, rbmi

Format Presence only

Advanced data visualization in R: (Re)producing professional plots with ggplot2 and the tidyverse

2024-02-28 09:00 - 12:00, C4-S04

Lecturer

Paul Schmidt (BioMath GmbH, Hamburg)

Learning objectives

- Deep dive into ggplot2: Understand the intricacies of ggplot2 for top-tier data visualization
- Data manipulation with tidyverse: Utilize packages like dplyr, tidyr, and forcats to shape data optimally for creating graphs
- Aesthetic fine-tuning: Master the skills of axis formatting, theme detailing, and color selection
- Reproducing Exercise: Recreate published plots to understand the application of ggplot2 techniques in real-world scenarios
- Exporting excellence: Learn best practices for exporting plots in publication-ready formats

Target group

Anyone wanting to produce or present high-quality plots. Conveying complex data insights through refined visualizations is a relevant part of statistical analysis

Prerequisites

- Basic understanding of R programming is required
- Basic knowledge of ggplot2 and the tidyverse is beneficial, but not required
- R-packages: tidyverse, cowplot, ggrepel, ggtext, viridis

Format Hybrid

Keynote Talks

Michael Love (University of North Carolina-Chapel Hill, USA)



Michael Love's research focuses on statistical and computational methods for the analysis of high-dimensional genetic data with the goal of facilitating biomedical and biological research. He has developed several widely used open source software packages for the analysis of RNA sequencing data, which also feature highly detailed step-by-step instructions.

2024-02-28 13:00 - 14:20, AM 1

Keynote 1: Opening Session & Keynote Michael Love

Pragmatic Biometrics for Transcriptomics: Rigor, Reproducibility, and Readability

Major advances in sequencing and other biotechnologies have propelled the state of the art in transcriptomic measurement, to the current state of profiling transcriptomes of individual cells, as well as the ability to directly sequence full RNA transcripts. Throughout changes in technology, accurate biometric analysis requires pragmatic choices in the processing and statistical modeling of transcriptomic measurements, guided by exploratory data analysis. I will discuss lessons learned from the past decade of transcriptomics, from rigorous bias correction, to automated mechanisms of ensuring reproducible analysis, and current efforts at facilitating code readability for data processing and analysis. I will conclude by suggesting how these lessons may be applied to data from new transcriptomic technologies.

Ina Rondak (European Medicines Agency, The Netherlands)



Ina Rondak is a statistician for the European Medicines Agency at the Methodology Division of the Data Analytics and Methods Task Force. Her focus is on providing scientific support for the development and evaluation of medical devices, especially using complex and innovative study designs and statistical methods.

2024-02-29 10:40 - 11:40, AM 1

Keynote 2: Keynote Ina Rondak

Wider access to more informative data and the key role of methodological experts in this endeavour

The European regulatory network is undertaking various activities to increase access to data and to improve the quality of data that underpin decision-making on the benefits and risks of medicines in the EU. At the same time, it is also ensuring that the best expertise is available to address methodological challenges in the evolving landscape of drug-regulation.

In this talk we will look at the aim and progress of launched activities to increase data accessibility and data quality. We will reflect on key documents published by the European Medicines Agency (EMA), feedback from stakeholder interactions on related topics and what the future might hold in relation to access to data.

We will take a deep-dive into the emerging landscape of methodological topics where biostatistical expertise is needed more than ever and will outline how the restructuring of EMA's working parties with the newly established Methodology Working Party and Methodology European Specialised Expert Community (ESEC) can increase the European regulatory network's capacity to tackle the challenges ahead.

Maarten van Smeden (University Medical Center Utrecht, The Netherlands)



Maarten van Smeden focuses on the development, validation and implementation of predictive models. Through collaborations, he has contributed to the implementation of complex methodology in various disciplines and has been involved in the development and validation of a variety of diagnostic and prognostic prediction models.

2024-03-01 16:00 - 17:20, AM 1

Keynote 3: Closing Session & Keynote Maarten van Smeden

Clinical prediction modeling in the era of AI: a blessing and a curse

Medicine has a long history of using clinical prediction models to guide medical decision making and inform patients about prognosis and diagnosis. In the current era of AI, accessible high-performance computing, and large datasets, the possibilities for developing better clinical prediction models seem endless. The reality, however, is different. In this talk, I will reflect some of the blessings and some of the curses that come with the new era of AI in clinical prediction modeling.

Invited Sessions

SES-22: Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics

2024-02-29 13:20 - 14:40, AM 2

Organizer

Silvia Calderazzo (DKFZ, Heidelberg)

2024-02-29 13:20 - 14:00, AM 2

SES-22: Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics

Estimands and Complex Innovative Designs

Bretz, Frank

Novartis, Switzerland

Since the release of the ICH E9(R1) document in 2019, the estimand framework has become a fundamental part of clinical trial protocols. In parallel, complex innovative designs have gained increased popularity in drug development. It is currently unclear, however, to which degree the estimand framework applies to these novel designs. For example, should a different estimand be specified for each sub-population (defined, for example, by cancer site) in a basket trial? Or can a single estimand focusing on the general population (defined, for example, by the positivity to a certain biomarker) be used? In the case of platform trials, should a different estimand be proposed for each drug investigated? We discuss relevant estimand considerations pertaining to different types of complex innovative designs. We consider trials that allow adding or selecting experimental treatment arms, modifying the control arm, and selecting or pooling populations. We also address the potentially data-driven, adaptive selection of estimands in an ongoing trial and disentangle certain statistical issues that pertain to estimation rather than to estimands, such as the borrowing of non-concurrent information.

Invited Sessions

References

Collignon O, Schiel A, Burman CF, Rufibach K, Posch M, Bretz F. Estimands and Complex Innovative Designs. *Clinical Pharmacology & Therapeutics*. 2022;112(6):1183–90.

2024-02-29 14:00 - 14:20, AM 2

SES-22: Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics

Multiplicity Issues in Platform Trials – adjusting for what? Looking forward and benefiting from the past

Koenig, Franz; Posch, Martin; Zehetmayer, Sonja

Medical University of Vienna, Österreich

Platform trials have been proposed where several randomized clinical trials with related objectives are combined to a single trial with a joint master protocol to improve efficiency by reducing costs and saving time. Treatment arms can enter and leave the study at different times during its conduct, possibly depending on previous results or available resources and the total number of treatment arms in a platform trial is not fixed in advance. One big advantage of platform trials is the sharing of one or several control arms.

As many hypotheses will be eventually tested, we will discuss whether an adjustment for multiplicity is indeed needed in the context of platform trials. In addition to the two extreme positions of no adjustment at all or requiring strict control of the familywise error rate, control of other error rates such as the False Discovery Rate (FDR) will be scrutinized. Particular attention has to be given as the total number of hypotheses being tested is usually unknown in the planning phase and interim analyses might complicate matters further. Another issue is how the information of already collected data might impact the planning of treatments to be added.

We compare the impact on sample sizes needed and power also depending on which error rate should be controlled, e.g., the experiment-wise error rate or control of the FDR.

2024-02-29 14:20 - 14:40, AM 2

SES-22: Invited: Innovative Clinical Trial Designs: Estimands and Operating Characteristics

Modeling strategies for analysing platform trials with non-concurrent controls

Bofill Roig, Marta

Medical University of Vienna, Austria

Platform trials evaluate the efficacy of multiple treatments, allowing for late entry of the experimental arms and enabling efficiency gains by sharing controls. The control data is divided into concurrent (CC) and non-concurrent controls (NCC) for arms that join the trial later. Using NCC for treatment-control comparisons can improve the power but might cause biased estimates if there are time trends. Several approaches have been proposed to utilise NCC while aiming to maintain the integrity of the trial. Frequentist model-based approaches adjust for potential bias by adding time as a covariate to the regression model. The Time Machine considers a Bayesian generalised linear model that uses a smoothed estimate for the control response over time. The Meta-Analytic-Predictive prior approach estimates the control response by combining the CC data with a prior distribution derived from the NCC data.

In this talk, we review the analysis approaches proposed for incorporating NCC in the treatment-control comparisons of platform trials. We investigate the operating characteristics of the considered approaches by means of a simulation study, focusing on assessing the impact of the overlap between treatment arms and the strength of the time trend on the performance of the evaluated models. We furthermore present the R-package "NCC" for the design and analysis of platform trials. We illustrate the use of the above-mentioned approaches and show how to perform simulations in various settings through the "NCC" package.

References

Bofill Roig, M., Krotka, P., Burman, C.F., Glimm, E., Gold, S.M., Hees, K., Jacko, P., Koenig, F., Magirr, D., Mesenbrink, P., Viele, K., Posch M. On model-based time trend adjustments in platform trials with non-concurrent controls. BMC medical research methodology, 22(1), 1–16, (2022).

Krotka, P., Hees, H., Jacko, P., Magirr, D., Posch, M., and Bofill Roig, M. NCC: An R-package for analysis and simulation of platform trials with non-concurrent controls. SoftwareX, 23 (2023): 101437.

SES-23: Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics

2024-02-29 15:00 - 16:20, AM 2

Organizers

Dennis Dobler (Vrije Universiteit Amsterdam, The Netherlands) and Jan Feifel (Merck Healthcare KGaA, Darmstadt)

2024-02-29 15:00 - 15:40, AM 2

SES-23: Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics

Counting processes in stochastic epidemic models: the link with survival analysis

Putter, Hein¹; Goeman, Jelle²; Wallinga, Jacco³

¹Leiden University Medical Center, The Netherlands

²Leiden University Medical Center, The Netherlands

³National Institute for Public Health and the Environment (RIVM), The Netherlands

Mathematical models based on ordinary differential equations (ODE's) quantifying the interactions between susceptible, infectious, and recovered individuals within a population have played an important role in infectious disease modeling. The aim of this presentation is to explore the link between stochastic epidemic models based on the susceptible-infectious-recovered (SIR) model, and methods from survival analysis. We illustrate how standard software for survival analysis in the statistical language R can be used to estimate pivotal parameters in the stochastic SIR model in the very much idealized situation where the epidemic is completely observed. Extensions incorporating interventions, age structure and heterogeneity are explored and illustrated.

2024-02-29 15:40 - 16:00, AM 2

SES-23: Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics

Identifying alert concentrations using a model-based bootstrap approach

Möllenhoff, Kathrin^{1,2}; Schorning, Kirsten³; Kappenberg, Franziska³

¹University of Cologne, Germany

²Heinrich Heine University Düsseldorf, Germany

³TU Dortmund University, Germany

The determination of alert concentrations, where a pre-specified threshold of the response variable is exceeded, is an important goal of concentration–response studies. The traditional approach is based on investigating the measured concentrations and attaining statistical significance of the alert concentration by using a multiple t-test procedure.

In this talk, we propose a new model-based method to identify alert concentrations, based on first fitting a concentration–response curve and second constructing a simultaneous confidence band for the difference of the response of a concentration compared to the control. In order to obtain these confidence bands without the need of deriving the asymptotic behaviour of the underlying process, we use a bootstrap approach which can be applied to any functional form of the concentration–response curve. This particularly offers the possibility to investigate also those situations where the concentration–response relationship is not monotone and, moreover, allows to detect alerts at concentrations which were not measured during the study, providing a highly flexible framework for the problem at hand.

We demonstrate the validity of the method by means of a simulation study and present an application to a real dataset investigating the effect of different concentrations of the compound VPA on the development of hESC to neuroectoderm.

2024-02-29 16:00 - 16:20, AM 2

SES-23: Invited: Stochastic Processes in Time-to-Event Analysis and Biostatistics

Resampling-based inference for the average treatment effect in time-to-event data

Rühl, Jasmin; Friedrich, Sarah

Augsburg University, Deutschland

The g-formula can be used to estimate treatment effects while accounting for confounding bias in observational studies. For time-to-event endpoints, statisticians need to take additional difficulties into account. It is for example not advisable to answer causal questions by hazard ratios, which is why we consider the risk difference instead. This way, a competing risks framework is accommodated on top. The distribution of the associated stochastic process is rather complicated, and hence, confidence intervals are commonly constructed by means of the classical non-parametric bootstrap. In certain situations, e.g., when the data lack independence, the classical bootstrap suffers from limitations, though. Furthermore, its execution can be rather time-consuming.

This work investigates alternative resampling methods proceeding from the underlying stochastic process of the average treatment effect. Apart from Efron's classical bootstrap, we consider an approach that is based on the influence function [1] and, since counting processes are inherent to time-to-event analysis, a bootstrap version based on the martingale representation of the process. It is shown that the bootstrap methods approximate the distribution of the stochastic process at hand.

We further compare the precision of the different techniques in a simulation study considering confidence intervals and bands for the average treatment effect. Our simulations imply that the wild bootstrap generally achieves accurate coverage levels if the sample size is small and sufficient data on the event of interest have been accrued.

References

- [1] Ozenne, B. M. H. and Scheike, T. H. and Stærk, L. and Gerdts, T. A. (2020). On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biometrical Journal*, 62: 751-763.

SES-24: Invited: Biometric Research Outside High-Income Settings

2024-03-01 09:00 - 10:20, AM 1

Organizer

Michael Schomaker (LMU München)

2024-03-01 09:00 - 09:40, AM 1

SES-24: Invited: Biometric Research Outside High-Income Settings

Assessment of the treat-all public health approach in Eswatini – challenges and opportunities

Kerschberger, Bernhard¹; Ciglenecki, Iza²

¹Médecins sans Frontières, Mbabane, Eswatini

²Médecins sans Frontières, Geneva, Switzerland

Assessing public health interventions under real-world conditions necessitates a holistic research approach, going beyond only methodological considerations. Considerations entail local engagement and institutional collaboration, ethical scrutiny, acquisition and quality of data, and statistical methodologies. We adopted a mixed-method research approach and embraced adaptability in study designs and methodologies when expanding HIV antiretroviral therapy in a resource-limited setting in Eswatini.

In 2011, the Ministry of Health (MOH) decided to pilot the HIV treat-all strategy several years ahead of the WHO recommendations. Under treat-all, all HIV-diagnosed patients become eligible for therapy immediately upon diagnosis, irrespective of immunological criteria. After years of preparation including the introduction of routine viral load monitoring, enhanced adherence support and expansion of HIV testing, MOH and Médecins Sans Frontières implemented the pilot intervention within one health zone from October 2014 to March 2016. A prospective cohort was established to assess the feasibility of treat-all with respect to intervention acceptance and uptake, and health outcomes. The neighbouring health zone continued applying the then national standard treatment recommendations based on CD4 cell thresholds. Approvals were obtained from MOH, and the local and institutional ethics committees. Data sources encompassed routine facility-based paper registers, a new electronic database for routine monitoring, and a research database maintained by MSF, with data clerks conducting data cleansing.

During analysis, data gaps emerged, necessitating multiple imputation techniques. The two patient cohorts exhibited disparities in baseline characteristics. Crude estimates and multivariate adjusted flexible parametric survival analysis unveiled substantial treatment uptake, yet with variations across patient sub-groups. Despite achieving favourable health outcomes (e.g. high viral suppression) and acceptable levels of program retention, patients commencing therapy on the same day as their HIV diagnosis appeared more prone to disengage from care. A follow-up investigation employing causal modelling, directed acyclic graphs under target trial assumptions, flexible parametric survival analysis, and targeted maximum likelihood estimation confirmed potential risks associated with same-day treatment initiation. Concurrently, qualitative research suggested high patient acceptance of treat-all, yet highlighted challenges of same-day treatment. A population-based survey conducted by another NGO substantiated the potential public health benefits of treat-all, demonstrating higher population-level viral suppression in this region compared to others.

Understanding feasibility and impact of public health interventions in real-world settings presents a challenge, considering issues pertaining to study methodology, data availability, and local procedures. Flexible approaches to study design, coupled with triangulation of findings, constitutes a pragmatic approach for affirming the effectiveness of such interventions.

2024-03-01 09:40 - 10:00, AM 1

SES-24: Invited: Biometric Research Outside High-Income Settings

The South African HIV Cancer Match (SAM) Study: a unique resource for cancer research among people with HIV in South Africa

Rohner, Eliane¹; Olago, Victor²; Ruffieux, Yann¹; Metekoua, Carole^{2,3}; Tombe-Nyahuma, Tinashe²; Mwansa-Kambafwile, Judith^{2,4}; Egger, Matthias^{1,5,6}; Muchengeti, Mazvita^{2,4}

¹Institute of Social and Preventive Medicine, University of Bern, Switzerland

²National Cancer Registry, National Health Laboratory Service, South Africa

³Graduate School for Health Sciences, University of Bern, Switzerland

⁴School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

⁵Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

⁶Centre for Infectious Disease Epidemiology and Research (CIDER), School of Public Health and Family Medicine, University of Cape Town, South Africa

As people with HIV (PWH) live longer due to the roll-out of antiretroviral therapy, cancer becomes an increasingly common comorbidity in this population. Most PWH live in sub-Saharan Africa, but studies on cancer risk among PWH in this region are scarce. The South African HIV Cancer Match (SAM) study aims to assess the cancer burden among PWH in the context of the evolving HIV epidemic. It is a nationwide cohort of PWH built through a probabilistic record linkage of HIV-related laboratory records from the National Health Laboratory Service (NHLs) and cancer diagnoses from the National Cancer Registry (NCR) in South Africa [1]. The NHLs is the largest diagnostic pathology service provider in South Africa. It serves approximately 80% of the population through a nationwide network of laboratories, and all electronic laboratory records are stored at a corporate data warehouse in Johannesburg. The NCR was established in 1986, and it collects information on pathologically confirmed cancer diagnoses. All names were encrypted using a privacy-preserving probabilistic record linkage (P3RL) encryption tool [2]. A probabilistic record linkage approach was used to identify HIV-related laboratory records (HIV tests, CD4 cell count and HIV RNA viral load measurements) belonging to the same person based on episode number, first name(s), surnames(s), sex, date of birth/age, and the province, district, and sub-district of the health facility. In a second step, the cohort was probabilistically linked to cancer diagnoses from the NCR based on first name(s), surname(s), sex, and date of birth/age. All record linkages were performed using G-Link (V.3.3 Rel V.5.2). For the period 2004-2014, 52,836,496 HIV-related laboratory records were retrieved and after deduplication linked to 664,869 cancer records from the NCR. The current SAM cohort includes 5,248,648 PWH who had 2 laboratory records on separate dates (69% female) with a median age at cohort entry of 33 years (interquartile range 26-40). The most common incident cancers were Kaposi sarcoma,

cervical cancer, breast cancer, non-Hodgkin's lymphoma, and ocular cancer. To date, the SAM database has been used for various analyses. For example, we found that most infection-related cancers were associated with low CD4 cell counts [3] and that the incidence rates of conjunctival squamous cell carcinoma decreased over time [4]. The SAM provides a unique resource for cancer surveillance and research among PWH. We are updating the SAM database to include data until 2021 and plan to construct an HIV-negative control cohort using diabetes related laboratory records.

References

1. Muchengeti M, Bartels L, Olago V, et al. Cohort profile: the South African HIV Cancer Match (SAM) Study, a national population-based cohort. *BMJ Open* 2022; 12:e053460.
2. Schmidlin K, Clough-Gorr KM, Spoerri A, SNC Study Group. Privacy preserving probabilistic record linkage (P3RL): a novel method for linking existing health-related data and maintaining participant confidentiality. *BMC Med Res Methodol*. 2015 May 30;15:46.
3. Ruffieux Y, Muchengeti M, Egger M, et al. Immunodeficiency and Cancer in 3.5 Million People Living with Human Immunodeficiency Virus: the South African HIV Cancer Match Study. *Clin Infect Dis*. 2021 Aug 2;73(3):e735-e744.
4. Metekoua C, Ruffieux Y, Olago V, et al. Decreasing Incidence of Conjunctival Squamous Cell Carcinoma in People with HIV in South Africa. *J Natl Cancer Inst*. 2023 Jun 28;djad119.

2024-03-01 10:00 - 10:20, AM 1

SES-24: Invited: Biometric Research Outside High-Income Settings

Interplay between Applied Biostatistics and Public Health in Humanitarian Settings

Luque Fernandez, Miguel Angel

University of Granada, Spain

Biometrical work in resource-constraint settings comes with several challenges, for example with respect to using the right techniques to evaluate the feasibility of interventions, but also in terms of engaging with local partners. I would like to share various empirical examples where advanced biostatistical methods were essential to answering complex population-based research questions in resource-constrained settings and humanitarian settings. First, we will share the “club of patients” research project where we used inverse probability weighting of marginal structural models to evaluate the effectiveness of a non-randomized intervention.[Luque-Fernandez et al., 2013] Second, we will describe the reemergence of cholera epidemics linked to climatic change, evaluated by using auto-regressive Poisson time series analysis.[Luque Fernandez et al., 2009, 2012, 2011] Finally, we will show how to use classical biostatistical methods to improve the diagnostic accuracy of the mid-upper-circumference-arm (MUAC), a device tool, used to identify severe malnutrition in humanitarian crises.[Fernandez et al., 2010]

References

- M. A. Fernandez, P. Delchevalerie, and M. Van Herp. Accuracy of MUAC in the detection of severe wasting with the new WHO growth standards. *Pediatrics*, 126(1):e195–201, 2010. ISSN 1098-4275 (Electronic) 0031-4005 (Linking). doi: 10.1542/peds.2009-2175. URL <https://www.ncbi.nlm.nih.gov/pubmed/20587675>.
- Fernandez, Miguel Angel Luque Delchevalerie, Pascale Van Herp, Michel eng Comparative Study 2010/07/01 *Pediatrics*. 2010 Jul;126(1):e195-201. doi: 10.1542/peds.2009-2175. Epub 2010 Jun 29.
- M. A. Luque Fernandez, A. Bauernfeind, J. D. Jimenez, C. L. Gil, N. El Omeiri, and D. H. Guibert. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Trans R Soc Trop Med Hyg*, 103(2):137–43, 2009. ISSN 1878-3503 (Electronic) 0035-9203 (Linking). doi: 10.1016/j.trstmh.2008.07.017. URL <https://www.ncbi.nlm.nih.gov/pubmed/18783808>.
- Luque Fernandez, Miguel Angel Bauernfeind, Ariane Jimenez, Julio Diaz Gil, Cristina Linares El Omeiri, Nathalie Guibert, Dionisio Herrera eng England 2008/09/12 *Trans R Soc Trop Med Hyg*. 2009 Feb;103(2):137-43. doi: 10.1016/j.trstmh.2008.07.017. Epub 2008 Sep 9.
- M. A. Luque Fernandez, P. R. Mason, H. Gray, A. Bauernfeind, J. F. Fesselet, and P.

Maes. Descriptive spatial analysis of the cholera epidemic 2008-2009 in harare, zimbabwe: a secondary data analysis. *Trans R Soc Trop Med Hyg*, 105(1):38–45, 2011. ISSN 1878-3503 (Electronic) 0035-9203 (Linking). doi: 10.1016/j.trstmh.2010.10.001. URL <https://www.ncbi.nlm.nih.gov/pubmed/21075411>.

Luque Fernandez, Miguel Angel Mason, Peter R Gray, Henry Bauernfeind, Ariane Fesselet, Jean Francois Maes, Peter eng England 2010/11/16 *Trans R Soc Trop Med Hyg*. 2011 Jan;105(1):38-45. doi: 10.1016/j.trstmh.2010.10.001. Epub 2010 Nov 13.

M. A. Luque Fernandez, M. Schomaker, P. R. Mason, J. F. Fesselet, Y. Baudot, A. Boulle, and P. Maes. Elevation and cholera: an epidemiological spatial analysis of the cholera epidemic in harare, zimbabwe, 2008-2009. *BMC Public Health*, 12:442, 2012. ISSN 1471-2458 (Electronic) 1471-2458 (Linking). doi: 10.1186/1471-2458-12-442. URL <https://www.ncbi.nlm.nih.gov/pubmed/22708576>.

Luque Fernandez, Miguel A Schomaker, Michael Mason, Peter R Fesselet, Jean F Baudot, Yves Boulle, Andrew Maes, Peter eng England 2012/06/20 *BMC Public Health*. 2012 Jun 18;12:442. doi: 10.1186/14712458-12-442.

M. A. Luque-Fernandez, G. Van Cutsem, E. Goemaere, K. Hilderbrand, M. Schomaker, N. Mantangana, S. Mathee, V. Dubula, N. Ford, M. A. Hernan, and A. Boulle. Effectiveness of patient adherence groups as a model of care for stable patients on antiretroviral therapy in khayelitsha, cape town, south africa. *PLoS One*, 8(2):e56088, 2013. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0056088. URL <https://www.ncbi.nlm.nih.gov/pubmed/23418518>.

Luque-Fernandez, Miguel Angel Van Cutsem, Gilles Goemaere, Eric Hilderbrand, Katherine Schomaker, Michael Mantangana, Nompumelelo Mathee, Shaheed Dubula, Vuyiseka Ford, Nathan Hernan, Miguel A Boulle, Andrew eng R01 AI102634/AI/NIAID NIH HHS/U01 AI069924/AI/NIAID NIH HHS/ U01AI069924-01/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2013/02/19 *PLoS One*. 2013;8(2):e56088. doi: 10.1371/journal.pone.0056088. Epub 2013 Feb 13.

SES-25: Invited: Statistical Issues in Animal Testing

2024-03-01 10:40 - 12:00, AM 1

Organizers

Bernd-Wolfgang Igl (Boehringer-Ingelheim) and Michael Lauseker (LMU München)

2024-03-01 10:40 - 11:20, AM 1

SES-25: Invited: Statistical Issues in Animal Testing

Statistical Review of Animal trials in Ethics Committees

Piper, Sophie K.¹; Konietzschke, Frank¹; Zocholl, Dario¹; Röhle, Robert¹; Tölch, Ulf²

¹Charité Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology, Germany

²Berlin Institute of Health at Charité - Universitätsmedizin Berlin, QUEST Center for Responsible Research, Germany

Any experiment or trial involving living organism requires ethical review and agreements. Beyond reviewing medical need and goals of the trial, statistical planning of the design and sample size computations are key review criteria. Errors made in the statistical planning phase can have severe consequences on both the results and conclusions drawn from a trial. Moreover, wrong conclusions therof might proliferate and impact future trials—a rather unethical outcome of any research. Therefore, any trial must be efficient in both a medical and statistical way in answering the questions of interests to be considered as “ethically approvable”.

For clinical trials, ethical review boards are well established. This is, however, not the case for pre-clinical and especially animal trials. While ethical review boards are established within each local authority of animal welfare, most of them do not have an appointed statistician. Moreover, unified standards on statistical planning and reporting thereof are currently missing for pre-clinical trials.

It is the aim of our presentation to introduce and discuss

- i) the need for proper statistical reviews of animal trials,
- ii) a guideline for applicants of animal trials and statistical reviewers in animal welfare

act committees that describes important biometrical design aspects such as sample size planning, data analysis, blinding and randomization [Piper et al., Biom J. 2022], and

iii) the need to distinguish the planning of exploratory studies from confirmatory studies in pre-clinical research.

Our statistical criteria for ethical reviews of animal trials have been implemented in a form sheet that has been used from the Landesamt für Gesundheit und Soziales (local authority of animal welfare) in Berlin since 2019. It is online available at

<https://www.berlin.de/lageso/gesundheit/veterinaerwesen/tierversuche/>.

Moreover, we published a Userguide for applicants (<https://zenodo.org/records/7359565>).

References

Piper SK, Zocholl D, Toelch U, Roehle R, Stroux A, Hoessler J, et al. Statistical review of animal trials-A guideline. Biom J. 2022. Epub 20220907. doi: 10.1002/bimj.202200061. PubMed PMID: 36071025.

2024-03-01 11:20 - 11:40, AM 1

SES-25: Invited: Statistical Issues in Animal Testing

Sample size justification in preclinical animal studies

Wilcke, Juliane C.; Boulesteix, Anne-Laure

LMU Munich, Germany

There is no question that the number of animals in a preclinical study is worth some consideration. The sample size needs to be large enough to ensure sufficiently reliable knowledge claims about the research question of interest. Otherwise, the used resources, including animal lives, are wasted and potentially misleading results disseminated. Worthwhile results should, however, be created with the minimum sensible number of animals, in accordance with the reduction principle of the 3R. Interestingly, while all animal studies require approval from the competent authorities for the number of animals included in the experiments, sample sizes are often not even mentioned in the resulting journal articles and hardly ever justified there. This undermines the informational value of most preclinical animal studies.

We distinguish five major approaches to justifying sample sizes in preclinical animal studies: power analysis, accuracy analysis, resource and regulatory constraints, heuristics, and no justification. At least the first two approaches require information on the effect of interest before data collection. This input to sample size determinations is the focus of our talk. It involves decisions regarding (a) the type of effect size that best fits the research question, such as the smallest effect size of interest or the expected effect size, (b) the source of the required information about the effect, its variability and other relevant parameters, such as own pilot data, expert judgement, or a combination of published findings, (c) probable uncertainties in this information deriving from a multitude of sources, including the transition from in vitro experiments and the ultimate interest in human clinical effects, (d) likely biases in this information, such as publication bias or the inflation of effect sizes in small-sample studies, (e) possibly also the desired accuracy of own findings, and (f) suitable mechanisms for taking all these aspects into account.

After giving an overview of the options available to preclinical animal researchers and their statistical consultants regarding the input to sample size determinations, we evaluate the impact of the corresponding choices on both the planned sample size and the informational value of the study. We illustrate these points with simulations and examples from preclinical animal research. For the quality of research as well as the lives of animals and ultimately patients, preclinical sample sizes need to closely match the research question and be well justified and reported according to the ARRIVE guidelines.

2024-03-01 11:40 - 12:00, AM 1

SES-25: Invited: Statistical Issues in Animal Testing

3R initiatives in preclinical toxicology and research

Lang, Tina; Vaas, Lea

Bayer AG, Germany

What are the 3Rs?

Replacement: Avoiding or replacing animal use e.g. using cell cultures, computer modelling, or human tissue or volunteers.

Reduction: Where animal use is necessary, keeping numbers to the minimum e.g. using statistical methods to determine the smallest number of animals that can be used in an experiment.

Refinement: Where animal use is necessary, minimising pain and suffering and improving welfare e.g. using pain relief and providing housing that allows animals to perform their natural behaviours

(National Centre for the Replacement, Refinement & Reduction of Animals in Research).

As statisticians, we directly impact the Reduction with proper and sound statistical sample size calculations (Pieper et al, 2023). The Refinement is oftentimes out of our hands, even though there are exceptions (e.g., body weight monitoring as indicator for animal wellbeing).

The Replacement, however, is a very active field of research. With concepts as organ-on-a-chip (Ingber 2022) or organoids (Tang et al, 2022) becoming more mature, the statisticians enter the field. Proper planning & analysis methods have to be adapted to new technologies, and the gained expertise needs to be shared with experimental scientists.

Then, the design of future experiments and the translation of results into humans can be early supported by statistics for high quality research. As the new technologies will lead to revision and extension of existing or even compilation of new guidelines, this is an opportunity to strengthen the role of statistical support by making it mandatory for preclinical experiments.

Invited Sessions

This talk will summarize the current situation and outline the discussion results regarding the topic from the IBS-DR Working Group “Non-Clinical Statistics”.

References

- Ingber DE. Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nat Rev Genet.* 2022;23:467–491. <https://doi.org/10.1038/s41576-022-00466-9>[1]
- National Centre for the Replacement, Refinement & Reduction of Animals in Research. Available from: <https://www.nc3rs.org.uk/3rs-public>. Published November 9th, 2023
- Piper SK, Zocholl D, Toelch U, Roehle R, Stroux A, Hoessler J, Zinke A, Konietzschke F. Statistical review of animal trials—A guideline. *Biometrical Journal.* 2023;65:2200061. <https://doi.org/10.1002/bimj.202200061>[5]
- Tang XY, Wu S, Wang D, et al. Human organoids in basic research and clinical applications. *Sig Transduct Target Ther.* 2022;7:168. <https://doi.org/10.1038/s41392-022-01024-9>[2]

SES-31: Invited: Risk Prediction meets Causal Inference

2024-02-29 13:20 - 14:40, AM 1

Organizer

Vanessa Didelez (BIPS, Bremen)

2024-02-29 13:20 - 14:00, AM 1

SES-31: Invited: Risk Prediction meets Causal Inference

Causal blind spots in risk-based medical decision making

van Geloven, Nan

Leiden University Medical Center, The Netherlands

In this presentation, I challenge the prevailing belief that accurate prediction algorithms lead to better treatment decisions. I highlight situations where using prediction models for individual decision support can actually cause more harm than good. I will illustrate these failure modes in real life medical examples using causal directed acyclic graphs (DAGs). While DAGs are typically used in settings where a causal effect is explicitly of interest, I here use DAG-based explanations in a predictive context.

To effectively integrate prediction models into decision-support frameworks, they must incorporate the causal structure inherent in the data. This presentation explores various approaches to augment prediction models with causal inference techniques, enabling the generation of predictions under different intervention scenarios. Development of predictions under interventions has different requirements than development of standard prediction models. I will also present recently proposed methods for assessing predictive performance of such causal predictions.

2024-02-29 14:00 - 14:20, AM 1

SES-31: Invited: Risk Prediction meets Causal Inference

Methods for obtaining causal predictions and quantifying associated uncertainty

DiazOrdaz, Karla

University College London, United Kingdom

Prediction models, whether statistical or AI, are often used to help decision making. However, these approaches should not be used to answer ‘what if’ questions. Failure to recognise when the prediction estimand is causal leads to incorrect risk predictions and suboptimal treatment or policy decisions.

Thus, we focus on causal (or counterfactual) predictions, where for each individual we predict what their outcome would be under a hypothetical policy or treatment, assuming the causal structure is known, and there are we have measure a sufficient set of variable to control for confounding.

Targeting a causal prediction estimand brings new challenges, because we can only use the observed (factual) treated sample to develop the model, but we must make predictions for the entire population. In the presence of confounding, the distribution of the factual treated may substantially differ from the target population (i.e. covariate shift). Further, we also consider situations where relevant variables are available at the model-building stage but are not available at deployment.

We review some existing methods for causal prediction under covariate shift, which allow the use of machine learning (e.g. DR-learner). We also make a simpler (but less robust) proposal, based on inverse weighting. Additionally, we implement distribution-free prediction intervals using conformal inference.

We compare the methods in a simulation study and illustrate them in a real example using electronic health records to obtain counterfactual predictions for type 2 diabetes patients under different Hba1c lowering drugs choices.

2024-02-29 14:20 - 14:40, AM 1

SES-31: Invited: Risk Prediction meets Causal Inference

Causal effect estimation and risk prediction - the example of screening colonoscopy and colorectal cancer

Braitmaier, Malte; Didelez, Vanessa

Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany

In this work, our primary objective was to assess the causal effect of screening colonoscopy on the incidence of CRC within the eligible German population using observational data (health claims database, see Braitmaier et al. (1)). Following the principles of target trial emulation, we estimated the effect of baseline screening versus no baseline screening using inverse probability of treatment weighting, supplemented by the g-formula in a sensitivity analysis, to account for confounding. The talk will also discuss methods to model different treatment paths (e.g. defining “never screening” instead of baseline screening as comparator strategy). As we targeted the full CRC incidence curves over available follow-up, these can in fact also be seen as predicting an eligible person’s (within age and sex subgroups) risk of CRC cancer at different points in time under the hypothetical scenarios in which they do / do not undergo a screening colonoscopy with all else corresponding to the current standard of care (e.g. the current use of diagnostic colonoscopies in Germany). While not originally intended as a risk prediction tool, we make the case that the statistical methodology and structural assumptions are very similar to the ones required for those types of prediction tasks that have to take different scenarios of possible future treatment paths into account.

References

1. Braitmaier M, Schwarz S, Kollhorst B, Senore C, Didelez V, Haug U (2022) Screening colonoscopy similarly prevented distal and proximal colorectal cancer: a prospective study among 55-69-year-olds; *Journal of Clinical Epidemiology*; 149: 118-126

SES-37: Invited: Synthetic Data with Privacy Guarantees: from Applied Statistics to Generative Machine Learning

2024-02-28 14:40 - 16:00, AM 1

Organizer

Andreas Ziegler (Medicine Campus Davos, Switzerland)

2024-02-28 14:40 - 15:20, AM 1

SES-37: Invited: Synthetic Data with Privacy Guarantees: from Applied Statistics to Generative Machine Learning

Generative machine learning in biostatistics

Wright, Marvin N.^{1,2,3}

¹Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany

²University of Bremen, Germany

³University of Copenhagen, Denmark

With the recent excitement around chat bots and image generation, generative machine learning is everywhere. Beyond this hype, generative modeling holds substantial promise for biostatisticians, offering the unique ability to generate synthetic data that, in some cases, come with privacy guarantees. In this presentation, I will provide an introduction into generative modeling and synthetic data with a focus on tabular health data. Deep learning methods such as generative adversarial networks (GANs) and a tree-based method called adversarial random forests (ARFs) are introduced and compared in terms of quality and privacy of the synthetic data. The practicality and utility of generative models in biomedical research are exemplified through real-world use cases, notably in the analysis of data from the German National Cohort (NAKO).

2024-02-28 15:20 - 15:40, AM 1

SES-37: Invited: Synthetic Data with Privacy Guarantees: from Applied Statistics to Generative Machine Learning

A simple-to-use R package for mimicking study data by simulations

Koliopanos, Georgios¹; Ojeda, Francisco²; Ziegler, Andreas^{1,2,3}

¹Cardio-CARE, Medizincampus Davos, Switzerland

²Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

³School 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 which 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 modgo which 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.

Conclusions: 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.

2024-02-28 15:40 - 16:00, AM 1

SES-37: Invited: Synthetic Data with Privacy Guarantees: from Applied Statistics to Generative Machine Learning

A representation of longitudinal data for enabling synthetic data generation via transformers

Farhadyar, Kiana; Königs, Lukas; Binder, Harald

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Germany, Germany

Longitudinal cohort data play a crucial role in biomedical research, as they offer valuable insight into the dynamics of patient health. When individual privacy should be protected in the analysis of such data, synthetic data generation provides a solution that preserves statistical relationships in the data. Yet, this requires approaches that can learn sequential patterns. We consider transformer architectures, which are based on attention mechanisms, i.e., weighting schemes for evidence in the course of time. However, transformers have mainly been developed for natural language processing, i.e., application to biomedical cohort data requires adaptation. Specifically, we propose a method for converting longitudinal datasets into a discrete event format, which is more amenable to transformers. As an example, we take a binary variable from a dataset and record a discrete event if the value of interest occurs at least three times in a row. We then treat events in a specific variable or group of variables as a word, corresponding to natural language processing applications of transformers, and only keep an event in our vocabulary set if it is frequent. To obtain a word embedding, i.e., a dimension reduction, we train an autoencoder at the level of the original binary variables. The corresponding encoder then can also encode the event vectors to an embedding. We illustrate the advantages of such an event coding and embedding with data that comprises self-reported stressors, collected in a longitudinal resilience assessment study. Words are defined via the repeated experience of a stressor. There, we find that the word embeddings (at the event level) and attention scores (at the sequence level) align well with subject matter knowledge on stressors. In particular, the attention scores seem to be more meaningful, i.e., not just equally distributed over the past, when using the proposed embedding as compared to directly applying a transformer to the data. Thus, such embeddings can provide a promising foundation for subsequently generating synthetic data from transformers.

SES-38: Invited: Copula Regression for Time-to-Event Data

2024-02-28 16:20 - 17:40, AM 1

Organizers

Nadja Klein (TU Dortmund) and Andreas Mayr (University of Bonn)

2024-02-28 16:20 - 17:00, AM 1

SES-38: Invited: Copula Regression for Time-to-Event Data

Copula based Cox proportional hazards models for dependent censoring

Van Keilegom, Ingrid¹; Deresa, Negera Wakgari²

¹KU Leuven, Belgium

²Universiteit Hasselt, Belgium

Most existing copula models for dependent censoring in the literature assume that the parameter defining the copula is known. However, prior knowledge on this dependence parameter is often unavailable. In this article we propose a novel model under which the copula parameter does not need to be known. The model is based on a parametric copula model for the relation between the survival time (T) and the censoring time (C), whereas the marginal distributions of T and C follow a semiparametric Cox proportional hazards model and a parametric model, respectively. We show that this model is identified, and propose estimators of the nonparametric cumulative hazard and the finite-dimensional parameters. It is shown that the estimators of the model parameters and the cumulative hazard function are consistent and asymptotically normal. We also investigate the performance of the proposed method using finite-sample simulations. Finally, we apply our model and estimation procedure to a follicular cell lymphoma dataset.

References

Negera Wakgari Deresa and Ingrid Van Keilegom (2023). Copula based Cox proportional hazards models for dependent censoring, Journal of the American Statistical Association, DOI: 10.1080/01621459.2022.2161387

2024-02-28 17:00 - 17:20, AM 1

SES-38: Invited: Copula Regression for Time-to-Event Data

Boosting distributional copula regression models for bivariate time-to-event data

Briseno Sanchez, Guillermo^{1,2}; Klein, Nadja^{1,2}; Mayr, Andreas³; Groll, Andreas²

¹Research Center Trustworthy Data Science and Security

²TU Dortmund University

³Department of Medical Biometrics, Informatics and Epidemiology, University Hospital Bonn

We propose a highly flexible distributional copula regression model for bivariate time-to-event data in the presence of right-censoring. The joint survival function of the response is constructed using parametric copulas, allowing for a separate specification of the dependence structure between the time-to-event variables and their respective marginal survival distributions. The latter are specified using well-known parametric distributions such as the log-Normal, log-Logistic (proportional odds model), or Weibull (proportional hazards model) distribution, which result in parametric (also known as Accelerated Failure Time, AFT) models for the respective univariate event times. Embedding our model into the class of generalized additive models for location, scale and shape (GAMLSS), possibly all parameters of the joint distribution may be modelled in an additive fashion via semi-parametric predictors, allowing for a wide range of covariate effects (e.g. linear, non-linear, spatial, etc.). We propose estimation of all model coefficients simultaneously by means of component-wise gradient-based boosting. This way, our approach is able to conduct data-driven variable selection, a feature that is extremely helpful in such a complex model class. To the best of our knowledge, this is the first implementation of multivariate AFT models via distributional copula regression and automatic variable selection via statistical boosting. A special merit of our model class is that it works for high-dimensional ($p >> n$) settings. We illustrate the practical potential of our approach along different biomedical bivariate examples. All of our methods are implemented in the open source statistical software R as add-on functions of the package gamboostLSS, to ensure transparent and reproducible research.

2024-02-28 17:20 - 17:40, AM 1

SES-38: Invited: Copula Regression for Time-to-Event Data

A model-based boosting approach to deal with dependent censoring

Strömer, Annika¹; Klein, Nadja²; Mayr, Andreas¹

¹Department of Medical Biometrics, Informatics and Epidemiology, University of Bonn, Germany

²Department of Statistics, Technische Universität Dortmund, Germany

A popular model to study the effect of some covariates on the survival time T is the semiparametric proportional hazards model (or Cox model). Estimation in this model is well-established for commonly observed right-censored data assuming that the survival time T and the (right) censoring time C are stochastically independent given the covariates. This assumption mainly holds when censoring occurs at the end of the study.

However, in medical studies, there is often a dependency between event time and censoring time. For example, if a patient's health deteriorates and they choose to withdraw from the trial due to poor prognosis, the time of censoring depends on their health status. This leads to dependent censoring as patients with poorer health are more likely to be censored earlier. In this work, we propose a model-based boosting approach to deal with dependent censoring using distributional copula regression. This approach allows to model the joint distribution of survival and censoring times by linking appropriately marginal distributions for T and C through a parametric copula. Rather than assuming the marginals are known, all distribution parameters (including the copula parameter) are estimated simultaneously as functions of (potentially different) covariates. A key merit of the boosting approach is that estimation is even feasible for high-dimensional data with $p >> n$, when classical estimation frameworks easily meet their limits. In addition, the boosting algorithm includes data-driven variable selection. To investigate the performance of our approach, we conduct an extensive simulation study and illustrate its practical application with a biomedical example.

SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update

2024-02-29 09:00 - 10:20, AM 1

Organizers

Matthias Schmid (University of Bonn) and Carsten Schmidt (University Medicine Greifswald)

2024-02-29 09:00 - 09:20, AM 1

SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update

Evaluating biostatistical methods before use - each phase is important

Heinze, Georg¹; Boulesteix, Anne-Laure²; Kammer, Michael³; Morris, Tim P.⁴; White, Ian R.⁴

¹Medical University of Vienna, Center for Medical Data Science, Vienna, Austria

²Ludwig-Maximilians University of Munich, Institute for Medical Information Processing, Biometry and Epidemiology, Munich, Germany

³Medical University of Vienna, Center for Medical Data Science & Division of Nephrology, Vienna, Austria

⁴UCL, Institute of Clinical Trials & Methodology, MRC Clinical Trials Unit, London, United Kingdom

New biostatistical methods are developed at an ever-increasing pace. Data analysts, with or without statistical training, have a hard time to keep pace with these developments. Therefore, This leads to a rapid uptake of some questionable methods in medical papers, but in the majority of applications, data analysts, medical collaborators and reviewers resort to conventional methods with which they are familiar. Some of these methods are suboptimal but general knowledge of their pitfalls is scarce. On the other hand, the lack of robust evaluation may lead to the uptake of methods not-yet well understood in medical papers.

We proposed classifying the development and evaluation of biostatistical methods into four phases similar to the phases of clinical drug development and evaluation (Heinze et al, 2023). In brief, the four phases cover (1) the introduction and mathematical description

of a new method to address a given problem, (2) its evaluation in limited scenarios including comparisons with competing methods, (3) a broad and neutral evaluation of its properties in an extended range of scenarios, and (4) further investigations after extensive experience with the method leading to new areas of application, the detection of possible pitfalls, and conclusions about the range of scenarios in which a method is or is not preferred.

In the biostatistical literature, studies of phase 2 seem to prevail. They typically illustrate a problem for which no satisfying solution exists, mathematically describe some methods that address the problem, including a new method introduced in the study, and report on a comparative simulation study and a real data analysis to discuss aspects of application. Such studies are important but often biased towards the new method (Pawel et al, 2023).

We highlight the necessity of methodological studies that contribute to all phases of research. In particular, researchers should be encouraged to conduct studies that take steps to minimise 'inventor-bias' and studies that investigate breakdown scenarios for a method. These studies help in building the scientific evidence base about methods, but still need more acceptance from reviewers, editors and funding agencies. We also emphasize that it is unrealistic that a study covers more than two phases for a single method. Authors of methodological studies should always clarify the scope of their work with respect to the phase of evaluation. This is helpful for readers to evaluate whether a method is ready for use, or still needs specific research to make it trustworthy.

References

- Heinze, G., Boulesteix, A., Kammer, M., Morris, T. P., & White, I. R. (2023). Phases of methodological research in biostatistics—Building the evidence base for new methods. *Biometrical Journal*, early view. <https://doi.org/10.1002/bimj.202200222>
- Pawel, S., Kook, L., & Reeve, K. (2023). Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biometrical Journal*, early view.
<https://doi.org/10.1002/bimj.202200091>

2024-02-29 09:20 - 09:40, AM 1

SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update

Initial Data Analysis, Data Quality Assessments and proper Information Management to Improve the Transparency of Statistical Analyses

Schmidt, Carsten Oliver¹; Lusa, Lara²; Huebner, Marianne³

¹University Medicine Greifswald, Deutschland

²University of Primorska, Slovenia

³Michigan State University, US

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, 3). While IDA and DQA overlap, the latter checks data against requirements, whereas the former has a strong descriptive focus on data properties to inform about appropriate statistical analysis options. In practice, conducting IDA and DQA all too often proves to be a complex and time-consuming task, mainly for two reasons, deficiencies in the underlying information management and complexity.

Information management refers to the proper documentation of the semantics of the data, its provenance, as well as requirements and expectations about the data. Provenance refers to how exactly the data was collected and processed until it reaches the hands of a statistician. Requirements and expectations include, for example, distributional assumptions, admissibility ranges, or contradiction rules. The entirety of such information can be referred to as metadata. Metadata may be missing, erroneous, incomplete, or not universally accessible due to inappropriate formats such as pdf reports or reliance on implicit knowledge of those conducting a study. Particularly in the latter case, it takes extensive communication time to extract relevant information. Therefore, metadata should be available in machine-readable formats. Not only does this enable transparent communication about the targeted data body. In addition, machine-readability means direct usability to carry out IDA and DQA related analyses. Metadata is available to a certain extent when using tools, standards, and data models such as OpenClinica®, REDCap®, FHIR, OMOP CDM, amongst others.

The second obstacle, complexity, refers to the large number of possible checks and views on a data body and its variables in order to gain a comprehensive understanding of its problems and properties. Without a highly structured approach, omissions are likely to occur.

This talk will provide an overview of challenges to efficient and transparent IDA and DQA

as a basis for substantive statistical analyses. Using applied examples, we will illustrate how a comprehensive information management can serve as a basis for automated assessments to substantially speed up and increase the scope and quality of related activities. This may help to overcome the current lack of transparency with regards to the reporting of IDA and DQA findings.(4)

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.Liaw ST, Guo JGN, Ansari S, et al. Quality assessment of real-world data repositories across the data life cycle: A literature review. *J. Am. Med. Inform. Assoc.* 2021.
doi:10.1093/jamia/ocaa340
- 3.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. Res. Methodol.* 2021;21(1):63.
doi:10.1186/s12874-021-01252-7
- 4.Huebner M, Vach W, le Cessie S, Schmidt CO, Lusa L, Topic Group "Initial Data Analysis" of the SI. Hidden analyses: a review of reporting practice and recommendations for more transparent reporting of initial data analyses. *BMC Med. Res. Methodol.* 2020;20(1):61. doi:10.1186/s12874-020-00942-y

2024-02-29 09:40 - 10:00, AM 1

SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update

Recent developments in measurement error modelling

Küchenhoff, Helmut

LMU München, Deutschland

I will present some work of the STRATOS Group on complex measurement error modelling (see also Shaw et al. 2020). This includes

- 1) The use of predicted values producing Berkson error in epidemiology. The effect of this type of errors and its correction on estimating summary statistics is presented.
- 2) Time-varying exposures prone to measurement error in survival analyses. Methods to correct for measurement error in survival models are compared.
- 3) Methods for handling misclassification in variables which are an outcome of Machine learning algorithms or latent class analyses. Since machine learning algorithms usually give information about misclassification error, methods for correction in corresponding epidemiological models are available.

References

Shaw, PA, Gustafson, P, Carroll, RJ, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 2—More complex methods of adjustment and advanced topics. *Statistics in Medicine*. 2020; 39: 2232–2263. <https://doi.org/10.1002/sim.8531>

2024-02-29 10:00 - 10:20, AM 1

SES-39: Invited: Developing guidance for statistical analysis in observational research - a STRATOS Initiative update

The STRATOS Open Science panel

**Hoffmann, Sabine¹; Boulesteix, Anne-Laure¹; Dunkler, Daniela²; Hornung, Roman¹;
Kammer, Michael²; Luijken, Kim³**

¹LMU Munich, Germany

²Medical University of Vienna, Austria

³University Medical Center Utrecht, Netherlands

This talk introduces the newly formed Open Science Panel of the STRATOS initiative. The aim of this panel is to promote open science practices, both within STRATOS and more broadly, by providing accessible guidance for the scientific community on ways to make research more transparent, reproducible and credible. Concerning open science practices within the STRATOS initiative, there is already a consensus that STRATOS publications should be open access, and that open access data sets (or a synthetic resemblance of it) should be used. The panel plans to develop more specific open science recommendations for STRATOS.

Beyond these important steps towards more open practices within STRATOS, the panel is currently working on two projects aiming at elaborating guidance documents for a level 1 to 2 audience. The first project aims at giving practical guidance on how to make studies “computationally reproducible” by providing code and data that allow to re-generate the results. The second project aims at giving preliminary guidance on how to deal with the multiplicity of possible analysis strategies (and resulting versions of the findings) that usually exist for a given data set and a given research question. These projects will be briefly presented in the second part of the talk.

SES-45: Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations

2024-03-01 14:40 - 16:00, AM 1

Organizers

Christian Stärk and Andreas Mayr (both University of Bonn)

2024-03-01 14:40 - 15:20, AM 1

SES-45: Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations

Power of inclusion: Enhancing polygenic prediction with admixed individuals

Tanigawa, Yosuke^{1,2}; Kellis, Manolis^{1,2}

¹Massachusetts Institute of Technology, United States of America

²Broad Institute of MIT and Harvard, United States of America

Admixed individuals offer unique opportunities for addressing limited transferability in polygenic scores (PGSs), given the substantial trans-ancestry genetic correlation in many complex traits. However, they are rarely considered in PGS training, given the challenges in representing ancestry-matched linkage-disequilibrium reference panels for admixed individuals. Here we present inclusive PGS (iPGS), which captures ancestry-shared genetic effects by finding the exact solution for penalized regression on individual-level data and is thus naturally applicable to admixed individuals. We validate our approach in a simulation study across 33 configurations with varying heritability, polygenicity, and ancestry composition in the training set. When iPGS is applied to $n = 237,055$ ancestry-diverse individuals in the UK Biobank, it shows the greatest improvements in Africans by 48.9% on average across 60 quantitative traits and up to 50-fold improvements for some traits (neutrophil count, $R^2 = 0.058$) over the baseline model trained on the same number of European individuals. When we allowed iPGS to use $n = 284,661$ individuals, we observed an average improvement of 60.8% for African, 11.6% for South Asian, 7.3% for non-British White, 4.8% for White British, and 17.8% for the other individuals. We further developed iPGS+refit to jointly model the ancestry-shared and -dependent genetic effects when heterogeneous genetic associations were present. For neutrophil count, for example, iPGS+refit showed the highest predictive performance in the African group ($R^2 = 0.115$), which exceeds the best predictive performance for the White British group

($R^2 = 0.090$ in the iP GS model), even though only 1.49% of individuals used in the iP GS training are of African ancestry. Our results indicate the power of including diverse individuals for developing more equitable PGS models.

References

Tanigawa Y, Kellis M. Power of inclusion: Enhancing polygenic prediction with admixed individuals. Am J Hum Genet. 2023 Nov;110(11):1888-1902.

2024-03-01 15:20 - 15:40, AM 1

SES-45: Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations

Bridging the performance gap for underrepresented populations: How to account for population structure in polygenic risk modelling?

Klinkhammer, Hannah^{1,2}; Maj, Carlo³; Staerk, Christian¹; Krawitz, Peter²; Mayr, Andreas¹

¹Institute for Medical Biometry, Informatics and Epidemiology, Medical Faculty, Bonn University, Germany

²Institute for Genomic Statistics and Bioinformatics, Medical Faculty, Bonn University, Germany

³Center for Human Genetics, Philipps University Marburg, Germany

In the field of clinical genetics, polygenic risk scores (PRS) offer the possibility of capturing the genetic predisposition to specific traits. PRS are derived from common genetic variants with low to medium individual effects. A challenge is the handling the high-dimensionality of genotype data, especially in large-scale datasets. This entails addressing both large sample sizes (denoted as n) and the vast number of genetic markers used as potential predictors (denoted as p). A wide range of PRS methods focus on summary statistics from genome-wide association studies based on univariate effect estimates and combine them to a single score (e.g. PRScs, LDpred2). More recently, methods have been developed that can be applied directly on individual-level genotype data to model the variants' effects simultaneously (e.g. BayesR, snpnet). In this context, we introduced snpboost, a framework that applies statistical boosting on individual-level genotype data to estimate PRS directly via multivariable regression models. By iteratively working on batches of variants, snpboost can deal with large-scale cohort data, e.g. from the UK Biobank. The choice of loss function in statistical boosting dictates the type of regression problem that is optimized. The snpboost framework implements a large range of loss functions making it applicable to various outcomes, e.g. continuous, binary, count and time-to-event data. Furthermore, the extension of the framework to quantile regression allows us to focus not only on the mean of the conditional distribution but also on other aspects crucial for risk stratification of individual patients.

On the practical front, we address the challenge of PRS transferability. Most databases used for PRS derivation are dominated by individuals of European ancestry, leading to decreased predictive accuracy in underrepresented populations. Genetic ancestry can be approximated using principal components (PCs) derived from the genotype matrix. Considering different techniques, we investigate the impact of incorporating PCs as covariates in the PRS estimation process, in dependence of the proportion of individuals with non-

European ancestry in the data. In this investigation, we assess PRS for standing height in individuals from the UK Biobank with European or South Asian ancestry. Our findings reveal that integrating PCs during the PRS derivation enhances prediction performance in underrepresented populations. This enhancement shows a linear relationship with the proportion of individuals with non-European ancestry in the dataset. Notably, while the performance of the PRS substantially improves within the underrepresented population, its performance in the major population experiences only a marginal decrease.

2024-03-01 15:40 - 16:00, AM 1

SES-45: Invited: Advanced Statistical Modelling for Polygenic Risk Scores to Enhance their Transferability to Underrepresented Populations

Boosting European and Multi-ancestry polygenic models: an analysis across different phenotypes of the UK Biobank

Maj, Carlo; Klinkhammer, Hannah; Staerk, Christian; Krawitz, Peter; Mayr, Andreas

Center for Human Genetic, University of Marburg, Germany

We employed the snpboost algorithm to derive sparse polygenic risk scores (PRS) for multiple traits within the UK Biobank dataset. We had three primary objectives: 1. exploring the extent to which PRS models prediction scales in accordance with trait heritability; 2. checking if model sparsity correlates with trait polygenicity; 3. investigating how the population heterogeneity in the training data influences prediction performance across different target populations.

We performed a genome-wide analysis to evaluate snpboost-derived PRS models for traits characterized by heterogeneous genetic architectures, prioritizing quantitative traits for statistical power. In order to analyze population structure, the UK Biobank samples (around half a millions individuals), were mapped onto the 1K Genomes principal component space and classified into four populations (AFR, EAS, EUR, SAS) according to genetic proximity to the 1K Genomes superpopulations centroids.

For each phenotype, we constructed two PRS models: a 'eur' model using only European samples for training (40%) and validation (40%), and a 'multi-ancestry' model inclusive of all populations, reflecting the original superpopulation ratios. Performance was measured in a test set comprising 20% of the samples from all superpopulations (i.e., AFR, EAS, EUR and SAS).

The analysis demonstrated snpboost's efficacy in predicting according to trait heritability while matching the genetic signal's sparsity (SNP-h² was computed via LDSC and polygenicity was defined according to the number of independently associated loci via clumping). When focusing on traits with substantial heritability (SNP-h² > 0.05, n=141), multi-ancestry snpboost models proved to be less complex than eur snpboost models (average variant count 3,539 versus 5,064).

Testing across superpopulations revealed that multi-ancestry models outperformed eur models for non-European groups. Notably, the East Asian (EAS) population had an average R² improvement of 0.0278, the African (AFR) group an average increase of 0.0185,

and the South Asian (SAS) population a modest rise of 0.0067, while the European (EUR) group experienced a minimal decrease in R² of -0.00485.

Although these improvements in prediction performance are slight, they are relevant given the UK Biobank's predominance of European samples (95%). Thus including a small proportion of non-European ancestries already consistently enhanced the PRS models' generalizability. Furthermore, the larger sparsity of multi-ancestry models suggests that training with multi-ancestry individuals could decrease the confounding effects of population-specific linkage disequilibrium, thereby aiding in the selection of the most informative independent genetic variants.

Special Sessions

SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data

2024-02-29 15:00 - 16:20, AM S2

Organizers

Tim Friede (gemeinsame Präsidiumskommission „Methodenaspkte in der Arbeit des IQWiG und IQTIG“ der GMDS und der IBS-DR), Ralf Bender (IQWiG) and Jona Cederbaum (IQTIG)

2024-02-29 15:00 - 15:20, AM S2

SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data

Performing subgroup analyses in HTA applications

Grouven, Ulrich; Skipka, Guido

IQWiG, Germany

The performance of subgroup analyses is an important component in the preparation of HTA reports by the Institute for Quality and Efficiency in Health Care (IQWiG). The legal requirements explicitly oblige the Institute to conduct subgroup analyses with regard to age, sex, and disease severity [1].

The historical development of the methodological procedure for conducting subgroup analyses is outlined and the current methodological procedure described in IQWiG's methods paper is presented [2]. The starting point is to consider possible heterogeneity between subgroups and to perform an interaction test based on Cochran's Q statistic [3]. An alternative is to perform an F-test in the context of a meta-regression [4].

The concrete stepwise procedure for performing subgroup analyses and for deriving benefit statements is described in detail and illustrated with concrete examples from IQWiG's benefit assessment. Possible problems and limitations are discussed.

References

1. Deutscher Bundestag. Gesetzentwurf der Fraktionen SPD, CDU/CSU und BÜNDNIS 90/DIE GRÜNEN: Entwurf eines Gesetzes zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG) [online]. 2003 [Accessed: 04.02.2021]. URL: <http://dipbt.bundestag.de/doc/btd/15/015/1501525.pdf>.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.1 [online]. 2022 [Zugriff: 27.01.2022].
URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf
3. Borenstein M. Introduction to Meta-Analysis. Chichester, U.K.: John Wiley & Sons; 2009.
4. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J et al (Ed). Cochrane handbook for systematic reviews of interventions. Hoboken: Wiley-Blackwell; 2019.

2024-02-29 15:20 - 15:40, AM S2

SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data

Using meta-regression for investigating subgroups in a meta-analysis

Röver, Christian¹; Kramer, Malte²; Friede, Tim¹

¹University Medical Center Göttingen

²German Rheumatism Research Center, Berlin

Meta-analyses commonly include reports of additional subgroup-analyses, which may be useful as a sensitivity analysis, or to investigate whether a subset alone may provide sufficient (or consistent) evidence. Technically, the default procedure often is to perform separate analyses, overall and for data subsets.

Given that meta-analysis methods often perform poorly when only few studies are involved, matters get worse when study numbers are reduced to subsets. As performance issues commonly relate to the estimation of between-study heterogeneity (while the main focus is on overall effects), a promising approach may be to reduce model complexity by considering a single, common heterogeneity parameter. On the technical side, this means approaching subgroup analyses as a meta-regression problem (Röver and Friede, 2023). This methodological approach has been advocated previously (Dias et al., 2013), but is still rarely implemented in practice (Donegan et al., 2015).

We investigate the alternative approaches from both frequentist and Bayesian viewpoints and demonstrate the potential performance gains, as well as sensitivity to potential assumption violations using simulations as well as a larger number of published meta-analyses from the Cochrane Database of Systematic Reviews.

References

Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: Heterogeneity—subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making*, 33(5): 618-640, 2013.
<https://doi.org/10.1177/0272989X13485157>

Donegan S, Williams L, Dias S, Tudur-Smith C, Welton N. Exploring treatment by covariate interactions using subgroup analysis and meta-regression in Cochrane reviews: A review of recent practice. *PLoS ONE*, 10(6): e0128804, 2015.
<https://doi.org/10.1371/journal.pone.0128804>

Röver C, Friede T. Using the bayesmeta R package for Bayesian random-effects meta-

regression. Computer Methods and Programs in Biomedicine, 229:107303, 2023.
<https://doi.org/10.1016/j.cmpb.2022.107303>

2024-02-29 15:40 - 16:00, AM S2

SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data

Statistical analysis of aggregate results of health care providers

Rauh, Johannes

IQTIG, Berlin, Germany

In order to assure quality of health care systems, it is important to understand how quality differs between different health care providers. Treatment quality is usually quantified using quality indicators. Most indicators are based on rates of adverse events (such as delayed treatment, complications or readmissions) that occur during specific treatments. For many quality indicators, it is necessary to take into account patient characteristics using some form of risk adjustment, e.g. indirect standardization.

Quality indicators are often instantiated to identify performance outliers. They are also used to study how quality of care correlates with other characteristics of health care providers. Examples of such characteristics include their caseload or the availability of specialized equipment or personnel. The study of such associations naturally leads to regression problems. When risk adjustment is necessary, these regression problems are ideally modelled on the patient level: Looking at the patient level makes it possible to cleanly separate the effect of patient and provider characteristics on the outcome. On the other hand, due to data protection rules, patient specific data are often not freely available. In contrast, aggregated results at the provider level are more commonly available publically. This makes it worthwhile to also investigate statistically sound methods that can cope with such aggregated information.

We highlight some of the pitfalls when drawing conclusions from aggregated data. We also discuss examples from the literature and make suggestions for best practices.

2024-02-29 16:00 - 16:20, AM S2

SES-05: Special: IQWiG/IQTIG: Statistical analyses based on aggregated data

Volume-outcome analyses based on aggregated data

Gutzeit, Maurilio; Rauh, Johannes; Cederbaum, Jona

IQTIG, Germany

How does the quality of care of a health care provider depend on the number of treated patients? The study of this question is called a volume-outcome analysis. Volume-outcome analyses are ideally performed using case specific data. This makes it possible to cleanly separate the effect of patient covariates and provider covariates (like the caseload) on treatment outcomes. However, due to data protection rules, case specific data is often not freely available. On the other hand, aggregated data about health care providers, such as caseloads, complication rates and some information about the risk profile of the provider's cases, may be available more freely. We discuss what should be considered when using aggregated data for volume-outcome analyses. Moreover, we compare volume-outcome analyses using aggregated data with our proposed approach using case specific data.

SES-09: Special: Statistics in Toxicology

2024-03-01 14:40 - 16:00, AM S2

Organizer

Bernd-Wolfgang Igl (Boehringer-Ingelheim)

2024-03-01 14:40 - 15:00, AM S2

SES-09: Special: Statistics in Toxicology

How to benefit from statistics in toxicology

Rahnenführer, Jörg; Kappenberg, Franziska

TU Dortmund University, Germany

The role of statistics for analyzing and understanding toxicological data must be rated extremely highly. Bretz and Greenhouse (2023) argue that in biopharmaceutical research, where toxicology is often the first step, “the practice of statistics is built on the foundation of good statistical thinking”. They describe the four general steps of problem-solving: A) Understanding and representing the problem, B) determining the data strategy, including an inventory of available data and possibly the collection of additional data, C) developing and executing a solution strategy, including exploratory analysis and model building, and D) evaluating and communicating the research results. They show how these steps align well with the way “clinical biostatisticians typically engage in collaborative clinical research”.

In this talk, we discuss these points based on own research projects, carried out in the Research Training Group RTG 2624 “Biostatistical methods for high-dimensional data in toxicology”. We analyze real world data from toxicological experiments, both classical low-dimensional data, where a manageable number of measured variables are available, and high-dimensional data from measurements of gene expression or protein quantity. Rapid developments in biotechnology and computer technology have enabled the generation of such complex data, driving a paradigm shift in molecular biology, from hypothesis-based research to hypothesis-generating research.

One example presented is a new toxicogenomics database that was built using model selection for time-expression curves. We explain the rationale for the development of such a database. We also give an overview of typical necessary steps in the analysis of the corresponding data, including preprocessing and normalization steps, as well as sta-

tistical analyses tailored to the research question. Moreover, we briefly discuss popular complex analysis methods, such as penalization techniques in high-dimensional regression and classification, and enrichment tests for identifying important biological processes and molecular functions based on lists of genes or proteins identified as relevant in experiments.

2024-03-01 15:00 - 15:20, AM S2

SES-09: Special: Statistics in Toxicology

Virtual Control Groups in Toxicity Studies

Vaas, Lea A.I.¹; Gurjanov, Alexander²; Ulbrich, Hannes-Friedrich¹; Kreuchwig, Annika²; Steger-Hartmann, Thomas²

¹Bayer AG, Research & Development, Pharmaceuticals, Research & Pre-Clinical Statistics

²Bayer AG, Research & Development, Pharmaceuticals, Investigational Toxicology

In systemic toxicity studies replacement of concurrent control animals by so-called Virtual Control Groups (VCGs) may reduce the use of animals thus contributing to the 3R's principle of animal experimentation: Reduce, Refine, Replace.

Although VCGs are an established concept in clinical trials, the idea of replacing living beings with virtual data from historical data sets has so far not been introduced into the design of regulatory animal studies¹. However, major steps facilitating review of methodology for derivation of ViCoGs from historical control data and performance testing in statistical analysis, are the collection, curation and sharing of suitable sets of historical control data from preclinical toxicity studies.

This talk will summarize accomplished and ongoing efforts for cross-industry provision of data resources, standardization and curation activities and line out both general ideas and specific methodology for derivation of ViCoGs.

A discussion of advantages and pitfalls along this journey will include real-world examples, case studies² and potential solutions. Further, ideas for transferring these insights into regulations and guidelines, especially possibilities for reaching out for regulatory advice to gain acceptance of this concept as early as possible¹ will be discussed.

References

- [1] Steger-Hartmann, T., Kreuchwig, A., Vaas, L., Wichard, J., Bringezu, F., Amberg, A., Muster, W., Pognan, F. and Barber, C. (2020) Introducing the concept of virtual control groups into preclinical toxicology testing, ALTEX - Alternatives to animal experimentation, 37(3), pp. 343–349. doi: 10.14573/altex.2001311.
- [2] Gurjanov, A., Steger-Hartmann, T., Kreuchwig, A., and Vaas, L.A.I.(2023) Hurdles and Signposts on the Road to Virtual Control Groups -A Case study illustrating the Influence of Anesthesia Protocols on Electrolyte Levels in Rats, Front. Pharmacol. Sec.

2024-03-01 15:20 - 15:40, AM S2

SES-09: Special: Statistics in Toxicology

Robust and simple experimental designs for fitting dose response curves in toxicology

Holland-Letz, Tim

DKFZ, Germany

Experimental designs in dose response experiments often use simple setups where the dose levels are increased by a fixed factor on the log scale. More efficient or even formally optimal experimental designs exist for this context, but these are often unpopular among applied scientists, mainly for two reasons: i) The designs usually depend on the unknown true value of some of the parameters and ii) they often propose using only a small number of distinct dose levels, casting further doubt on their robustness to parameter misspecifications. On the other hand, more generally optimal designs such as quasi-Bayesian designs are often quite complicated, and still require specifying an assumed a-priori distribution of parameters.

In this talk we propose a single graphical representation of design robustness which shows the performance of any given experimental design under a wide range of possible parameters. Using this representation, we propose three different possible designs, which are both simple and still provide reasonable efficiency under many parameter constellations, without needing anything but the most coarse prior knowledge about these parameters.

2024-03-01 15:40 - 16:00, AM S2
SES-09: Special: Statistics in Toxicology

How to benefit from high-dimensional expression data in toxicology

Kappenberg, Franziska; Schorning, Kirsten; Rahnenführer, Jörg

TU Dortmund University, Germany

High-throughput methods for gene expression data allow the simultaneous measurement of tens of thousands of genes. However, due to cost, time and ethical constraints, the sample size in toxicological experiments is typically very small.

In this talk, several research projects on the toxicological analysis of high-dimensional concentration-response data, where the response is given by gene expression measurements, are presented. These projects focus on simultaneous statistical design, parametric modelling for high-dimensional gene expression data, and even benefitting from the high-dimensionality by estimating common parameters or by sharing information across genes. All projects are part of the research of several members of the Research Training Group 2624 "Biostatistical Methods for High-Dimensional Data in Toxicology".

While simultaneous comparisons against the negative control are well-established in toxicological analyses, fitting a parametric curve to the concentration-response relationships has been extensively investigated. The MCP-Mod procedure, combining a multiple comparison testing procedure and a modelling step, originally developed for Phase II clinical studies, has been shown to produce favorable results also when simultaneously applied to high-dimensional gene expression datasets.

The first step in any toxicological experiment is to plan the experimental design. It is well-known that corresponding statistical approaches often clearly improve analysis results, however they are rarely used in toxicological studies. Especially in high-dimensional gene expression data, one additional challenge is the requirement that a single 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 high-dimensional gene expression dataset, estimation of many parameters is required. In order to reduce the number of parameters, some common parameters are assumed within a group of genes. The use of a newly developed information criterion allows the selection of an appropriate model within the framework of common parameters.

A relaxation of the assumption of common parameters is given by overall information

Special Sessions

sharing. A newly developed empirical Bayes method shrinks the individual estimates of the parameter corresponding to the half-maximal effect (the ED₅₀) of a fitted curve towards the population mean, which is shown to result in an improvement of the estimation of this parameter. This idea is extended to a fully Bayesian hierarchical model, simultaneously inducing shrinkage on the ED₅₀ parameter and the slope parameter of a sigmoidal model.

Statistics in Practice 1 and 2

2024-03-01, 09:00 - 10:20 and 10:40 - 12:00, AM 2

Diagnostic accuracy studies: basic and advanced statistical methods

Speakers

- Prof. Dr. Antonia Zapf (Head of the Institute of Medical Biometry and Epidemiology (IMBE) at the University Medical Center Hamburg-Eppendorf)
- Alexander Fierenz, Denise Köster, and Philipp Weber (PhD students in the field of diagnostic studies at the IMBE)
- Dr. Maria Stark (Staburo GmbH München)

In the Statistics in Practice sessions, we will provide an overview of diagnostic accuracy studies. After a general introduction, the first part will focus on the planning of such a study, including study design and sample size (re-)estimation. The second part will focus on the analysis and reporting of the results of diagnostic accuracy studies and will also provide an outlook on randomized test-treatment studies. The content will be illustrated using example studies from the field of imaging diagnostics and questions and discussion are very welcome.

IBS-DR: Nachwuchspreise

Bernd-Streitberg Preise

2024-02-29 15:00 - 16:20, AM 1

SES-32: IBS-DR: Nachwuchspreise

Infusing structural assumptions into dimension reduction for single-cell RNA sequencing data to identify small gene sets

Brunn, Niklas

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Germany, Deutschland

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 informed assumptions about the underlying structure or the experimental design. For instance, dimensions that help to distinguish between cell groups intuitively should be characterized 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 characterize temporal changes of cell states. To this end, we propose the boosting autoencoder approach, which synergizes unsupervised deep learning models for dimension reduction and boosting to formalize constraints. Specifically, the approach selects distinct small sets of genes by maximizing a score function to explain cell state-related patterns in separate latent dimensions. We showcase the functionality of our approach through applications on simulated data, accounting for different structural assumptions. Additionally, we explore the diversity of neural cell identities and temporal patterns of embryonic development in different applications to gene expression data. Our approach presents a complementary dimension reduction strategy to identify cell stage-related genes without needing pre-defined cluster memberships.

2024-02-29 15:00 - 16:20, AM 1

SES-32: IBS-DR: Nachwuchspreise

Evaluation of Index-based Response-adaptive Randomization Procedures in Clinical Trials

Drescher, Sonja^{1,2,3}; Kunz, Cornelia Ursula²; Walther, Andrea³

¹Department of Medical Statistics, University Medical Center Göttingen

²Boehringer Ingelheim Pharma GmbH & Co KG, Germany

³Department of Mathematics, Humboldt-Universität zu Berlin

Response adaptive randomization (RAR) has been an active area of research since it was first proposed by Thompson in 1933 [1]. However, compared to the theoretical interest and the amount of methodological work, the use of RAR in clinical trials has been disproportionately low [2]. There are a number of common concerns about RAR that, despite the methodological work addressing these concerns, continue to impede the use of RAR in practice. The aim of the thesis was to assess the validity of common criticisms of RAR for a particular class of RAR procedures, the so-called index-based procedures, which result from the solution of a multi-armed bandit problem.

The discussion of these criticisms is based on simulations of multi-arm clinical trials using RAR. The index based-procedures are compared to each other and also to the permuted block design, as an example for a fixed randomization procedure, and to the Thall and Wathen procedure [3], which is a RAR procedure that is not based on indices.

One key finding is that the index-based procedures result in an undesirably high probability of a sample size imbalance in favor of an inferior treatment, when the response rate of the best treatment is close to the response rates of the inferior treatments. Furthermore, the effect of a delay between the randomization of a patient and the availability of their outcome on the performance metrics of RAR procedures was investigated. It was shown that although the performance metrics of RAR procedures, including patient benefit metrics and power, tend to be negatively affected by delayed responses, the index-based procedures still yield a notable improvement in patient benefit compared to permuted block randomization as long as the delay is not too long relative to the total trial duration.

The thesis was written during an internship at Boehringer Ingelheim and as part of the mathematics master's program at Humboldt-Universität zu Berlin

References

Special Sessions

- [1] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 5, no. 3-4, pp. 285-294, 1933.
- [2] D. S. Robertson, K. M. Lee, B. C. López-Kolkovska and S. S. Villar, "Response-adaptive randomization in clinical trials: from myths to practical considerations," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 38, no. 2, pp. 185-208, 2023.
- [3] P. F. Thall and J. K. Wathen, "Practical Bayesian adaptive randomisation in clinical trials," *European Journal of Cancer*, vol. 43, no. 5, pp. 859-866, 2007.

Gustav-Lienert Preis (1. Preis)

2024-02-29 15:00 - 16:20, AM 1

SES-32: IBS-DR: Nachwuchspreise

A connection between survival multistate models and causal inference for external treatment interruptions

Erdmann, Alexandra^{1,2}; Loos, Anja³; Beyersmann, Jan¹

¹Institute of Statistics, Ulm University, Ulm, Germany

²Development Biologicals, Boehringer Ingelheim Pharma GmbH, Biberach, Germany

³Global Biostatistics, Merck KGaA, Darmstadt, Germany

Recently, we investigated the impact of treatment discontinuations due to a clinical hold on the treatment effect using a multistate model framework [1]. A clinical hold order by the Food and Drug Administration (FDA) to the sponsor of a clinical trial is a measure to delay a proposed or to suspend an ongoing clinical investigation. The phase III clinical trial START [2] with primary endpoint overall survival served as the motivating data example to explore implications and potential statistical approaches for a trial continuing after a clinical hold is lifted. We proposed a multistate model incorporating the clinical hold as well as disease progression as intermediate events to investigate the impact of the clinical hold on the treatment effect.

In our presentation, we create the link to causal inference. We suggest a censoring approach to answer the question what treatment effect would have been observed if there had been no treatment interruption due to a clinical hold. Specifically, we explore the conditions under which independent, but not necessarily random, censoring leads to a causal interpretation of the partial empirical transition matrix. Furthermore, we also propose a causal filtering approach.

Using a realistic simulation study informed by the START data, we showed that our censoring and filtering approaches perform favorably compared to a naïve method ignoring the external impact. We pointed out that our censoring approach coincides with the causal truncated factorization formula and has a causal interpretation regarding the intention of the initial treatment.

We will discuss the assumptions that have to be fulfilled for the ‘causal’ censoring or filtering to address treatment interruptions in general settings with an external time-dependent covariate inducing a time-varying treatment.

References

Special Sessions

- [1] Nießl, Alexandra, Beyersmann, Jan and Loos, Anja. "Multistate modeling of clinical hold in randomized clinical trials." *Pharmaceutical Statistics* 19.3 (2020): 262-275.
- [2] Butts, Charles and others, Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial", *Lancet Oncology* 15.1 (2014): 59-68.
- [3] Erdmann, Alexandra, Loos, Anja and Beyersmann, Jan. "A connection between survival multistate models and causal inference for external treatment interruptions." *Statistical Methods in Medical Research* 32.2 (2022): 267-286.

Gustav-Lienert Preis (2. Preis)

2024-02-29 15:00 - 16:20, AM 1

SES-32: IBS-DR: Nachwuchspreise

Functional Additive Models on Manifolds of Planar Shapes and Forms

Stöcker, Almond

EPFL, Schweiz

In many imaging data problems, the coordinate system of recorded objects is arbitrary or explicitly not of interest. Statistical shape analysis addresses this by identifying the ultimate object of analysis as the „shape“ of an observation, i.e., its equivalence class modulo translation, rotation and re-scaling, or as its „form“ (or „size-and-shape“) modulo translation and rotation. The shape/form space of this equivalence class is endowed with a Riemannian manifold geometry, which needs to be considered in the analysis.

We introduce a flexible additive regression framework for modeling the shape or form of planar (potentially irregularly sampled) curves and/or landmark configurations in dependence on scalar covariates. Models are fit by a novel component-wise Riemannian L2-Boosting algorithm, which yields desirable means of regularization for high-dimensional scenarios and allows estimation of a large number of parameter-intense model terms with inherent model selection.

We utilize the framework A) to analyze configurations of 2D landmarks and outline segments describing the shape of astragali (ankle bones) of wild and domesticated sheep, taking also other "demographic" variables into account, and B) to analyze the form of (irregularly sampled) cell outlines generated from a cellular Potts model in dependence on different metric biophysical model parameter effects (including smooth interactions).

Graphic illustration usually plays an essential role in practical interpretation of smooth (non-linear) additive model effects but becomes a challenging task when the response presents an (equivalence classes of) planar curves or landmark configurations. Therefore, we suggest a novel visualization for multidimensional functional regression models. Analogous to principal component analysis often used for visualization of functional data, a suitable tensor-product factorization identifies an optimal intrinsic coordinate system for a covariate effect. After factorization, main effect directions can be illustrated on the level of curves, while the effect into the respective direction is visualized by standard effect plots for scalar additive models.

This allows us to not only extend additive models with their versatile applicability to

Special Sessions

(functional) shape/form responses, but also to visualize estimated effects in an intuitively accessible way.

Feier zum 70. Biometrischen Kolloquium (in German only)

2024-01-03, 13:00 - 14:20, AM 1

13:00 - 13:20

Edgar Brunner

Teil 1: Irgendwann hats angefangen; Teil2: Fitnesstraining: Oberwolfach und unsere Trainer

13:20 - 13:40

Gerta Rücker, Iris Pigeot, Lothar Kreienbrock

Die wilden Achtziger und die goldenen Neunziger

13:40 - 14:00

Susanne-Dahms Preisträger*in

Präsidentenschaften: Wirken nach innen und außen

14:00 - 14:20

AG Nachwuchs

AG Nachwuchs: Rückblicke, Einblicke, Ausblicke

Young Statisticians

2024-02-28, 16:20 - 17:40, AM 2

Organizer

AG Nachwuchs

2024-02-28 16:20 - 16:40, AM 2

SES-30: Young Statisticians

Sample size calculation and recalculation for non-inferiority trials in the ‘gold standard’ design using the studentized permutation test

Schulz, Max¹; Mütze, Tobias²; Friede, Tim¹

¹Department of Medical Statistics, University Medical Center Göttingen, Georg-August-University, Göttingen, Germany

²Novartis Pharma AG, Schweiz

A critical aspect of designing a clinical trial involves the determination of the sample size. Using the power function and the assumed treatment effect, one typically derives the required sample size as the smallest number for which the target power is obtained. To validate the assumptions made in the planning stage, trial designs may allow an adjustment of the initially planned sample size while the trial is in progress. For three-arm non-inferiority trials including an experimental treatment as well as an active control and a placebo control, Mütze et. al. [1] proposed a studentized permutation test for the retention-of-effect approach. This method does not require specific distributional assumptions on the data and therefore, proves useful for various applications and particularly small sample sizes. However, their study does not provide an investigation of the tests power and thus, a method to plan sample sizes. Tests that derive a rejection area based on permutations lack a closed-form power function and require simulations instead. This study aims to close that gap by simulating the tests power behaviour and subsequently deriving a procedure to determine sample sizes.

Firstly, this study investigates the power of the studentized permutation test by means of simulations. The results serve as a basis to propose an approximation formula for sample size planning. Secondly, this study examines possible procedures for sample size re-estimation based on nuisance parameter estimates from internal pilot study data. For this purpose, we compare the performance of several variance estimators, both unblinded and blinded. The simulation scenarios cover continuous data with both homogenous and

heterogeneous variance scenarios.

For continuous data, the simulations show that the power of the test behaves very similarly to its parametric equivalent [2], thus suggesting the use of its sample size formula. The studentized permutation test effectively achieves the desired power level when sample sizes are planned using the Hasler sample size formula. It offers a straightforward way to determine sample sizes that only requires specifying the expectation and variance parameters in advance. However, when variances are misspecified in the planning stage, the initially estimated sample sizes are inadequate. Sample size re-estimation procedures considerably improve the power level, but remain below the target level for smaller pilot study sizes. In such cases, inflating the re-estimated sample size improves the reliability of achieving the target power level.

References

- [1] Tobias Mütze, Frank Konietschke, Axel Munk, and Tim Friede. A studentized permutation test for three-arm trials in the 'gold standard' design. *Statistics in Medicine*, 36:883–898, 2017. doi: 10.1002/sim.7176.
- [2] Mario Hasler, Richardus Vonk, and Ludwig A. Hothorn. Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Statistics in Medicine*, 27(4):490–503, 2008. doi: 10.1002/sim.3052.

2024-02-28 16:40 - 17:00, AM 2

SES-30: Young Statisticians

A statistical deconvolution method with an application to secretomic and proteomic data

Anarat, Akin^{1,2}; Krutmann, Jean²; Schwender, Holger¹

¹Heinrich-Heine University Düsseldorf, Deutschland

²Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany

Aging processes such as intrinsic and extrinsic aging in the skin are influenced by genetic and environmental factors. In order to understand if these two types of aging develop independently or if they influence each other, the secretome or proteome of intrinsically and extrinsically aged skin fibroblasts from the same individuals can be compared. However, a challenge in analyzing extrinsic aging is that the entire human skin ages intrinsically. Therefore, it is only possible to measure the combination of intrinsic and extrinsic aging in areas of the skin that are exposed to external factors. To study the effects of intrinsic and extrinsic aging in these areas, it is thus necessary to separate the pure extrinsic signal from the combined intrinsic and extrinsic signal.

To achieve this goal, we propose a statistical deconvolution technique that allows for the extraction of signals from mixed data containing multiple sources. This nonparametric deconvolution method relies on the Fourier transform inversion theorem and the estimation of density functions using Poisson regression fit to histogram counts of proteomic data. Specifically, we apply this method to data obtained from intrinsically aged fibroblasts and combined intrinsically and extrinsically aged fibroblasts. By numerically integrating the Fourier transforms of predicted functions, we can estimate the pure extrinsic signal by considering the quotient of these transforms.

To assess the performance of the proposed nonparametric deconvolution method, we conduct a simulation study. The results demonstrate that this nonparametric deconvolution procedure is capable of extracting the signal from an unmeasurable source in a mixture of signals, provided that the signal from other components in the mixture can be measured. Additionally, we apply the proposed deconvolution method to secretomic and proteomic data from the Gerontosys study, which were collected from intrinsically aged fibroblasts as well as intrinsically and extrinsically aged fibroblasts belonging to individuals from different age groups.

2024-02-28 17:00 - 17:20, AM 2

SES-30: Young Statisticians

Evaluation of Time-To-Event-Endpoints in Oncology Biosimilar Trials – A Simulation Study

Bohlken, Jan-Georg; Wright, Marvin N.; Hemmelmann, Claudia

Sandoz, Germany

BACKGROUND: Biosimilars are close copies of established but costly biological agents with no clinically meaningful differences in safety and efficacy compared to the originator. Accurate assessment of equivalence is necessary to broaden access to effective treatments [1]. In Phase III oncology trials, the similarity of treatment effects is typically measured primarily based on response rates. However, this approach is criticized especially when comparing immunotherapies due to the delayed effects and unclear definition of a response [2].

OBJECTIVE: This work aims to evaluate the ability of comparative statistical procedures to reliably assess equivalence based on time-to-event endpoints like overall survival (OS) and progression-free survival (PFS) for immunotherapy agents. The effect of different sample sizes and follow-up times as well as the robustness of Cox regression-based hazard ratio estimation under the delayed effect is at focus.

METHODS: Based on two historic trials of an established immunotherapy reference product, equivalence margins were calculated and reconstructed data from the reference arms was modeled parametrically. Using this, equivalence trials with OS and PFS endpoints were simulated using piecewise exponential models to include a delayed effect and specify different degrees of biosimilarity, given by the range of the equivalence margin. As efficacy summary measures, the hazard ratio as well as differences and ratios of median survival time, restricted mean survival, and survival rates at specific time points are used to estimate the equivalence of the simulated study arms. Bootstrap confidence intervals were calculated in cases where the density function of the comparative measures cannot be approximated by the normal distribution.

RESULTS: The first results indicate that OS endpoints require too long follow-up times (> 3 years) to reasonably assess bioequivalence using median survival-based estimates for OS. But for utilizing PFS endpoints arguments might be found as the ratios of median survival times show interesting distributional patterns. Further analysis is required and will be presented at the conference.

References

Special Sessions

1. Isakov L, Jin B, Jacobs IA. Statistical Primer on Biosimilar Clinical Development. American Journal of Therapeutics. 2016;23(6):e1903-e10.
2. Uno H, Schrag D, Kim DH, Tang D, Tian L, Rugo HS, et al. Assessing Clinical Equivalence in Oncology Biosimilar Trials With Time-to-Event Outcomes. JNCI Cancer Spectr. 2019;3(4):pkz058.

2024-02-28 17:20 - 17:40, AM 2

SES-30: Young Statisticians

Real world vs. controlled diagnosis: The example of response in oncology

Miller, Carina¹; Beyersmann, Jan²; Herpers, Matthias³

¹Ulm University, Ulm, Germany

²Institute of Statistics, Ulm University, Ulm, Germany

³Global Biostatistics & Data Sciences, Boehringer Ingelheim Pharma, Ingelheim am Rhein, Germany

In the context of drug approval, clinical trials in oncology are associated with substantial time and financial costs since a high number of patients is required. To minimize those efforts data obtained in the real-world clinical practice is intended to support regulatory decision-making. This can accelerate medical product development, lead to reduced expenses, and minimize the number of patients needed in clinical trials.

Consequently, more evidence on the comparability between outcomes assessed in the real-world and those assessed in clinical trials needs to be obtained. This is not trivial as the patient cohorts as well as the assessment of outcomes differ in these two settings. In oncology trials, outcomes are measured in accordance with the strict RECIST guidelines, first by the local investigator and afterwards by the independent central imaging review. Due to the strict schedule of events, nearly all needed data regarding the analysis is available. Whereas in real-world clinical practice, the outcome of a patient is assessed more subjectively by the treating physician and many missings occur in the data. Real-world healthcare data providers derive EHR-generated real-world response variables which aim to imitate the response according to RECIST.

This thesis compares the outcome variables of the two different data sources to assess the use of real-world response as surrogate for RECIST. The research question examines associations between the oncological response to treatment assessed in the real-world and the study outcome defined by RECIST. To ensure comparability in the baseline characteristics, first, the patient cohorts are retrospectively aligned by applying the inclusion and exclusion criteria utilized in the clinical trials to the real-world patient cohort. To reduce further bias in the data that might come from the missing randomization, Propensity Score Matching is implemented, with 1:1 nearest neighbor matching with random order on the logit of the Propensity Score.

Estimating the group effects on the best overall response, an association between the real-world outcome and the study outcome assessed by the local investigator is observed when looking at the whole matched population. Thus, the best overall response assessed

Special Sessions

in the real-world can be considered a surrogate for the outcome in clinical trials for certain research questions. Given the promising results obtained from the analysis conducted in this thesis, its usage in clinical trials can be justified. However, given the limitations of this analyses, more analyses comparing real-world response with RECIST need to be conducted.

Panel Discussion (AG-Nachwuchs): Navigating the Academic Odyssey: Early Career Challenges

2024-02-29, 09:00 - 10:20, AM 2

Panelists:

- Jörg Rahnenführer (TU Dortmund University)
- Kathrin Möllenhoff (University of Cologne and Heinrich Heine University Düsseldorf)
- Dominik Thiele (TCC GmbH Hamburg)
- Björn-Hergen Laabs (University of Lübeck)

Workshops für Schulen: Künstliche Intelligenz - Spielerisch lernen

2024-03-01, 09:00 - 10:20 and 10:40 - 12:00, AM S4

70. Biometrisches Kolloquium
Navigating the Sea of Data: Biometrics Guides the Way
28. Februar – 1. März 2024
Universität zu Lübeck

Das Institut für Medizinische Biometrie und Statistik der Universität zu Lübeck organisiert vom 28.2.-1.3.2024 die wissenschaftliche Jahrestagung der Deutschen Region der Internationalen Biometrischen Gesellschaft.

In der Biometrie werden statistische Methoden entwickelt und genutzt, um zusammen mit klinischen Kooperationspartnern medizinische Fragestellungen zu beantworten. Einige dieser Methoden gehören zu den maschinellen Lernverfahren, einem Teilgebiet der Künstlichen Intelligenz.

Im Rahmen der Tagung wird für Schulklassen folgender Workshop angeboten:

Künstliche Intelligenz – Spielerisch lernen

Was genau ist Künstliche Intelligenz? Wie lernt überhaupt eine Maschine? Und was ist ein Algorithmus? In diesem Workshop entdeckt ihr spielerisch, wie Künstliche Intelligenz wichtige Aufgaben in der Biomedizin bewältigt. Ihr werdet selbst zu Datenpunkten im Raum und entwickelt clevere Algorithmen, um unterschiedliche Aufgaben zu lösen.

Datum: Freitag, 1.3.2024 (9:00 – 10:20 oder 10:40 – 12:00)

Ort: Audimax-Gebäude der Universität zu Lübeck

Leitung: Prof. Dr. Helena Zacharias, Medizinische Hochschule Hannover



Kontakt und Anmeldung:

PD Dr. Bärbel Kunze
(Leiterin des LoLa)
E-Mail: baerbel.kunze@uni-luebeck.de



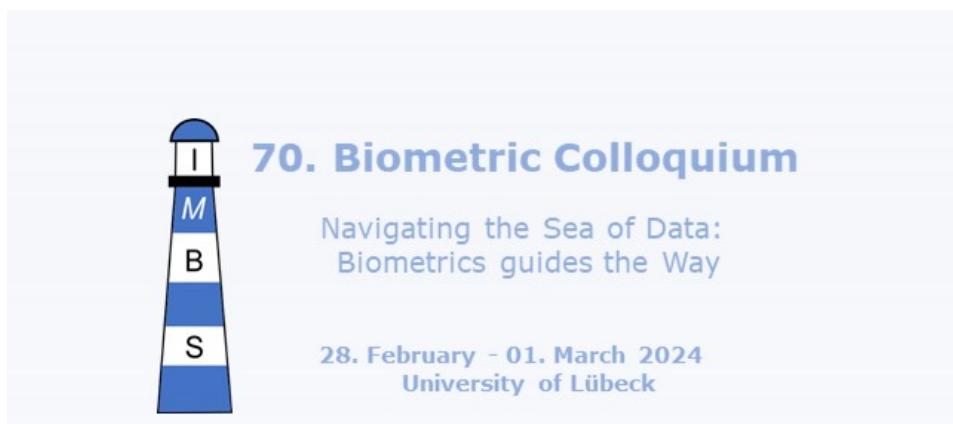
Deutsche Region der
Internationalen Biometrischen Gesellschaft
(IBS-DR)



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MEDIZINISCHE
BIOMETRIE UND STATISTIK

Vortrag für die Öffentlichkeit: Sich selbst mit dem Smartphone behandeln: Wie gut funktioniert das?

2024-03-01, 18:00 - 19:00, AM 1



The poster features a blue and white striped lighthouse icon on the left. To its right, the text "70. Biometric Colloquium" is written in a large, bold, blue font. Below it, the subtitle "Navigating the Sea of Data: Biometrics guides the Way" is displayed in a smaller blue font. At the bottom, the dates "28. February - 01. March 2024" and the location "University of Lübeck" are mentioned.



The central part of the poster is a white rectangular box with a blue border. Inside, there is a logo featuring a red heart with a white ECG line and a red blood drop. Next to the logo, the text "Öffentlichkeitsvortrag" is written in a blue font. Below the logo, the title of the lecture is given as "„Sich selbst mit dem Smartphone behandeln: Wie gut funktioniert das?“". Underneath the title, the speaker's information is provided: "Prof. Dr. Med. Stefan Sauerland", "Institut für Qualität und Wirtschaftlichkeit", and "im Gesundheitswesen". To the right of the text, there is an illustration of a smartphone displaying a red cross icon. At the bottom left of the white box, the date and time "01.03.2024 18:00 Uhr" and the location "Audimax AM1" are listed.

FREIER EINTRITT
Keine Anmeldung erforderlich

Abstracts of Contributed Talks

2024-02-28 14:40 - 15:00, AM S3

SES-01: Statistics in Epidemiology

Post-selection inference in case-control sampling designs: Results from the TORONTO Monte Carlo simulation study

Schöpe, Jakob; Wagenpfeil, Stefan

Institute for Medical Biometry, Epidemiology and Medical Informatics, Saarland University, Germany

Background: In the realm of observational studies, the prevalence of post-selection inference issues is quite substantial. Nevertheless, these issues tend to be overlooked and the results of the statistical modeling are often reported and interpreted without considering their inherent uncertainty. A recent proposal by Efron introduced a bootstrap approach to address the issue of incorporating uncertainty in statistical inference following model selection. Nonetheless, the performance of Efron's approach remains uncertain in the context of (matched) case-control sampling designs that involve a binary outcome variable and a mixture of confounding variables.

Objectives: The objectives of the TORONTO Monte Carlo simulation study were twofold: firstly, to evaluate the precision and accuracy of the bagged estimator employed in Efron's approach; and secondly, to compare and assess the coverage probability of the confidence interval estimators (specifically, smoothed (bias-corrected), percentile and standard) proposed by Efron within (matched) case-control sampling designs using a regularized procedure in the logit model and non-parametric bootstrapping.

Design: A comprehensive full-factorial Monte Carlo simulation study was conducted to evaluate the operating characteristics of Efron's approach in the context of (matched) case-control sampling designs. The data-generating process incorporated a total of sixteen random variables, each with a prespecified probability distribution. A Gaussian copula was employed to model the underlying prespecified stochastic dependence between the random variables, assuming an unstructured stochastic dependence structure. Pseudo-random samples were generated from a hypothetical population consisting of 100,000 units of observation in order to emulate a case-control sampling design for the prespecified scenarios. The matching for matched case-control sampling designs was ac-

complished through propensity score matching. Features of the Docker technology and R version 4.3.1 facilitated the execution on a virtual private server optimized for high-performance computing and the assurance of a completely reproducible environment.

Findings: The key findings of the TORONTO Monte Carlo simulation study will be summarized in this talk.

References

Efron B. Estimation and accuracy after model selection. *J Am Stat Assoc* 2014; 109: 991-1007.

2024-02-28 15:00 - 15:20, AM S3

SES-01: Statistics in Epidemiology

Maximum likelihood estimation for aggregated current status data: Simulation study using the illness-death model for chronic diseases with duration dependency

Mohammadi Saem, Maryam; Brinks, Ralph

Universität Witten-Herdecke, Germany

We use the illness-death model (IDM) for chronic conditions to derive a new analytical relation between the transition rates between the states of the IDM. The transition rates are the incidence rate (i) and the mortality rates of people without (m_0) and with disease (m_1). For the most generic case, the rates depend on age, calendar time and in case of m_1 also on the duration of the disease. It turns out that the prevalence-odds can be expressed as a kind of convolution product of the incidence rate and an exponentiated linear combination of i , m_0 and m_1 . The analytical expression can be used as the basis for a maximum likelihood estimation (MLE) and MLE's usual large sample asymptotics can be applied.

In a simulation study according to [1] where a cross-sectional trial about a chronic condition is mimicked, we estimate the duration dependency of the mortality rate m_1 based on aggregated current status data using the ML estimator. For this, the number of study participants and the number of diseased people in eleven age groups are considered. The ML estimator provides reasonable estimates for the parameters including their large sample confidence bounds.

Keywords: incidence, prevalence odd, mortality, simulation study.

References

1. Brinks, Ralph, et al. "Lexis diagram and illness-death model: simulating populations in chronic disease epidemiology." PLoS One (2014): e106043

2024-02-28 15:20 - 15:40, AM S3

SES-01: Statistics in Epidemiology

Recent methodological trends in Epidemiology: No need for data-driven variable selection?

Staerk, Christian¹; Byrd, Alliyah²; Mayr, Andreas³

¹RWTH Aachen University, Germany

²Duke University, Durham, USA

³University of Bonn, Germany

Variable selection in regression models is a particularly important issue in epidemiology, where one usually encounters observational studies. In contrast to randomized trials or experiments, confounding is often not controlled by the study design, but has to be accounted for by suitable statistical methods. For instance, when risk factors should be identified with unconfounded effect estimates, multivariable regression techniques can help to adjust for confounders.

We investigated the current practice of variable selection in four major epidemiological journals, including the American Journal of Epidemiology, Epidemiology (journal), the European Journal of Epidemiology, and the International Journal of Epidemiology. We found that 199 (73%) of the 272 included articles in our review used prior knowledge to select variables, among which 35 articles (13%) also employed causal graphs to summarize and justify the selection of variables based on subject-matter knowledge. In comparison with previous similar reviews, fewer articles applied data-driven variable selection (16% of articles in our review compared to 24% and 37% in previous reviews). Furthermore, for most articles the main aim of analysis was hypothesis-driven effect estimation in rather low-dimensional data situations (i.e., large sample size compared to the number of variables). Yet, we found that articles concerned with genetic epidemiology, which were excluded from previous reviews, usually incorporated some form of univariate selection of genetic variables, e.g., based on marginal p-values from genome-wide association studies.

We conclude that a priori variable selection based on subject-matter knowledge is increasingly predominant in epidemiological research with a focus on hypothesis-driven estimation in situations with a limited number of variables. However, there is a special need for data-driven variable selection in high-dimensional and large-scale data situations, which are typically encountered in genetic epidemiology. While further methodological research on scalable variable selection and post-selection inference methods is needed, a special focus should also be on the transfer of new methodological developments from statistics towards epidemiological research and practice.

Abstracts of Contributed Talks

This work is based on a recent article published in the American Journal of Epidemiology (Staerk et al., 2023).

References

Staerk C, Byrd A, Mayr A. Recent methodological trends in Epidemiology: No need for data-driven variable selection? Am J Epidemiol. 2023:kwad193. Available from: <https://doi.org/10.1093/aje/kwad193>

2024-02-28 15:40 - 16:00, AM S3

SES-01: Statistics in Epidemiology

The evaluation of clinical prediction models – methodological pitfalls illustrated with an application to peripartum depression

Mayr, Andreas¹; Costa, Raquel²; Martins, Rui³; de Sousa, Bruno⁴; Kneib, Thomas⁵

¹University of Bonn

²Universidade Lusófona

³University of Lisbon

⁴University of Coimbra

⁵University of Göttingen

The validation of a clinical prediction model is a very important but seemingly simple task from a methodological perspective. The model should either correctly assess the cross-sectional risk of a patient for a current illness (diagnostic model) or the risk to develop a clinically relevant condition in the future (prognostic model). However, in practice there are some larger methodological pitfalls when designing and performing validation trials that are easily overlooked.

While the methodological intuition might be to validate the prediction model on a population as close as possible to the development cohort, from a clinical perspective a fairer and more relevant validation might actually be performed on data that reflect the heterogeneous patient population the model will later be used for. The TRIPOD statement, in this context, clearly favors temporal validation and non-random splitting when performing validation on data from the same center to allow for heterogeneity across cohorts. External validation appears optimal, embracing different locations; however, defining what constitutes "external" remains ambiguous. Is it another research team, a different clinic, or even another country? Moreover, does any performance gap compared to internal validation then relate to shifting patient populations or the model's inability to generalize across different populations?

Selecting performance criteria poses another challenge. While classical discrimination and calibration measures are widely accepted for binary outcomes, the choice of the most suitable performance measure for continuous outcomes is more complex. For instance, the R^2 measure has gained attention in recent years for assessing prediction performance. Yet, surprisingly diverse definitions of R^2 on test data can yield vastly different results, especially in the presence of heterogeneous model performance.

We will illustrate and discuss these issues based on a recent study on peripartum depression that includes almost 12,000 pregnant and postpartum women from 11 European

Abstracts of Contributed Talks

countries.

2024-02-28 16:20 - 16:40, AM S3

SES-02: Clinical Epidemiology

First effects of HPV vaccination on cervical cancer incidence in Germany - an age-period-cohort analysis based on population-based cancer registry data

Grieger, Paula; Eisemann, Nora; Hammersen, Friederike; Rudolph, Christiane; Katalinic, Alexander; Waldmann, Annika

Institut für Sozialmedizin und Epidemiologie, Universität zu Lübeck, Deutschland

Introduction: Human papillomavirus (HPV) infection is the most common sexually transmitted disease. In Germany, HPV vaccination was implemented in 2007 for girls aged 12-17 years (modified to the range 9-14 years in 2014). The aim of this study is to assess whether there is an existing population-based effect of vaccination on HPV-associated cancer. We focus on the incidence of cervical cancer in women under the age of 30 who have been eligible for vaccination in the past.

Methods: We used data provided by the Centre for Cancer Registry Data (ZfKD) and included females with a maximum age of 120 years who were diagnosed with HPV-associated cervical cancer between 2004 and 2018 and who resided in Bavaria, Bremen, Hamburg, Lower Saxony, Saarland, Schleswig-Holstein, or the administrative district of Muenster. Age-period-cohort models were applied to investigate age, period, and cohort effects on cervical cancer incidence risk (invasive and *in situ*) using the US National Cancer Institute web tool. Data of each three years were aggregated (age groups: 9-11, 12-14, . . . ; diagnosis years: 2004-2006, 2007-2009, . . .).

Results: Overall, 96,330 cervical cancers contributed to the data base. For all age groups, the net drift or average annual percent change in the age-adjusted incidence rate of HPV-associated cervical cancer was 1.6% [95% CI: -0.09; 3.4]. The 1992 birth cohort, which was the first vaccine-eligible cohort, had a 23% lower incidence than the reference cohort born in 1989 [relative risk: 0.76, 95% CI: 0.68; 0.86]. The relative risk dropped down to 0.4 [95% CI: 0.06; 2.9] for the 2001 cohort and to 0.21 [95% CI: <0.01; 403] for the 2004 cohort. The overall trend was dominated by the trend for *in situ* cancers which accounted for 76% of all cancers. However, the incidence decrease was greater for invasive than for *in situ* cervical cancer.

Conclusions: The first positive effects of HPV vaccination are evident at the population level more than ten years after its introduction. In vaccine-eligible birth cohorts, the protective effect is reflected in a decline in cervical cancer incidence, although the proportion of 15-year-old girls with complete HPV vaccination was as low as 27% in 2010-2012.

2024-02-28 16:40 - 17:00, AM S3

SES-02: Clinical Epidemiology

The Skellam Distribution revisited - Estimating the unobserved incoming and outgoing ICU COVID-19 patients on a regional level in Germany

Rave, Martje; Kauermann, Göran

Ludwig Maximilian Universität, Germany

With the beginning of the COVID-19 pandemic, we became aware of the need for comprehensive data collection and its provision to scientists and experts for proper data analyses. In Germany, the Robert Koch Institute (RKI) has tried to keep up with this demand for data on COVID-19, but there were (and still are) relevant data missing that are needed to understand the whole picture of the pandemic. In this paper, we take a closer look at the severity of the course of COVID-19 in Germany, for which ideal information would be the number of incoming patients to ICU units. This information was (and still is) not available. Instead, the current occupancy of ICU units on the district level was reported daily. We demonstrate how this information can be used to predict the number of incoming as well as released COVID-19 patients using a stochastic version of the Expectation Maximization algorithm (SEM).

This, in turn, allows for estimating the influence of district-specific and age-specific infection rates as well as further covariates, including spatial effects, on the number of incoming patients. The paper demonstrates that even if relevant data are not recorded or provided officially, statistical modelling allows for reconstructing them. This also includes the quantification of uncertainty which naturally results from the application of the SEM algorithm.

2024-02-28 17:00 - 17:20, AM S3

SES-02: Clinical Epidemiology

When will I be cured from cancer? – An application of cure models to cancer registry data from Schleswig-Holstein

Baltus, Hannah; Schliemann, Antje; Schumann, Laura; Katalinic, Alexander; Eismann, Nora

Institute for Social Medicine and Epidemiology, University of Lübeck, Germany

Introduction: The question of whether and when one is cured from cancer is important for patients, health care providers, and others (e.g., insurance or credit companies). Yet survival probability is the most used measure to statistically describe the time survived after a cancer diagnosis. Cure from cancer can depend on many different factors such as gender, age, kind of tumour diagnosis, stage at diagnosis and treatment, but also on general health status and socio-demographics. Using population-based data from cancer registries, we want to answer the question “when will I be cured?”, also stratified for available prognostic variables.

Methods: Statistical “cure from cancer” is defined as having the same mortality probability as a sex- and age-matched individual from the general population.

We extracted data on all cancers diagnosed between 2004 and 2019 from the cancer registry of Schleswig-Holstein. We stratified our analyses by cancer entity, age group, sex, and stage at diagnosis. We compared the observed number of deaths in our population with the expected number of deaths in a group matched by sex and age using official cause-of-death statistics to determine attributable mortality. We examined the probability of cure as a function of time elapsed since tumour diagnosis.

Results: Higher stage at diagnosis and older age prolong time to cure and lower the number of individuals that reach cure. For prognostically favourable tumours, such as early-stage breast or prostate cancers, there is not much difference in time to cure. For some early-stage cancer cure is reached immediately after primary care, leading to same mortality as for the general population (e.g. breast cancer stage I). Prognostically unfavourable tumours, such as lung or pancreatic cancers, are mainly diagnosed at higher stages, and survival is low in the first years. In this group, the probability of being cured depends strongly on the time that has elapsed since diagnosis, but few patients reach the point of cure.

Discussion: We were able to show that established cure models can be applied to German cancer registry data and lead to realistic results. Next, we will analyse the data from all

Abstracts of Contributed Talks

over Germany and extend the time period considered.

To capture the impact of new therapies for advanced tumours that have been developed and approved in recent years (such as immunotherapies) time to cure should be monitored in the upcoming years, because the possible effect of these therapies might not be visible in the currently available data.

2024-02-29 09:00 - 09:20, AM S1

SES-03: Meta-Analysis 1

Overall and landmark survival estimates by stage for patients with non-small cell lung cancer treated with either surgery alone or surgery plus adjuvant systemic anticancer treatment – an analysis based on German cancer registry data

Baltus, Hannah; Labohm, Louisa; Katalinic, Alexander; Waldmann, Annika

Institute for Social Medicine and Epidemiology, University of Lübeck, Germany

Introduction: Data derived from population-based clinical cancer registration allows to describe survival as a function of treatment. We aim to describe overall and landmark survival estimates overall and stratified by stage, treatment and disease status six months after surgery.

Methods: We pooled data from four cancer registries (Baden-Württemberg, Hamburg, North Rhine-Westphalia, Schleswig-Holstein; covering 35% of the German population). We included patients with non-small cell lung cancer (ICD-10 C34; three main subtypes of NSCLC: squamous cell carcinoma, adenocarcinoma, large cell carcinoma; diagnosis: 2016-2019) and residence in the respective federal states at time of diagnosis. Treatment data (provided by the registries) was used to classify patients as resected (surgery alone; within 6 months after diagnosis), resected plus adjuvant systemic anticancer treatment (adj. SACT; within 84 days after surgery) or other/unknown treatment. Follow up data on recurrences or progression was also provided by the registries. Vital status was achieved by the registries via linkage with registration offices' data. We report Kaplan-Meier survival estimates for overall survival and landmark survival in a population re-indexed after surviving at least 6 months after diagnosis.

Results: Overall, 9,549 patients (median age: 68 years; females: 40%; UICC Stages: I (40%), II (24%), IIIA (16%), IIIB (4.4%), IIIC (0.4%), IV (15%)) with surgery (n=7422, 78%) or surgery plus adj. SACT (n=2127) and a follow-up of at least 1 day were included in our analysis. Overall 1-year survival was 77% (95% confidence interval: [0.76, 0.78]) (5-year estimate: 29% [0.24, 0.35]). Survival decreased with increasing stage. Stratification by treatment revealed that patients with surgery as compared to surgery plus SACT had lower survival estimates, regardless of UICC stage. The statistical significance of this difference is not stable over time. For the landmark population we considered disease status six months after surgery (early recurrence (n=946), free of disease (n=5335)). One-year landmark survival estimates were almost 90% in disease-free patients of stages I to IIIA, but remarkably lower in patients with early recurrences/progression. Stage IV disease-free patients (n=504) had a one-year landmark survival estimate of 66% [0.62, 0.71] (with recurrence/progression: 41% [0.34, 0.49], n=182).

Abstracts of Contributed Talks

Discussion: Albeit low survival in general, patients with early stage non-small cell lung cancer and those without early recurrences experience a comparably good survival. This information could be used to (better) inform patients about survival prospects.

2024-02-29 09:20 - 09:40, AM S1

SES-03: Meta-Analysis 1

A discrete time-to-event model for the meta-analysis of full ROC curves

Stoye, Ferdinand Valentin¹; Tschammler, Claudia¹; Kuss, Oliver²; Hoyer, Annika¹

¹Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Germany

²Institute for Biometry and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Germany

The development of new statistical models for the meta-analysis of diagnostic test accuracy (DTA) studies is still an ongoing field of research, especially with respect to summary receiver operating characteristic (ROC) curves. In the recently published updated version of the “Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy”, the authors point to the challenges of this specific kind of meta-analysis and propose two approaches (1). However, both of them come along with some disadvantages, such as longer computation times of Bayesian models or the requirement of a two-step approach where parameters are estimated for the individual studies, followed by summarizing the results. As an alternative, we propose a novel model by applying methods from time-to-event analysis. To this task we use the discrete proportional hazard approach to treat the different diagnostic thresholds, that provide means to estimate sensitivity and specificity and are reported by the single studies, as categorical variables in a generalized linear mixed model (GLMM), using both the logit- and the asymmetric cloglog-link (2). This leads to a model specification with threshold-specific discrete hazards, avoiding a linear dependency between thresholds, discrete hazard and sensitivity/specificity and thus increasing model flexibility. We compare the resulting models to approaches from the literature (3) in a simulation study. While the estimated area under the summary ROC curve is estimated comparably well in most approaches, the results depict substantial differences in the estimated sensitivity and specificity values as well as in the empirical coverage. We also show the practical applicability of the models to data from a meta-analysis for the screening of type 2 diabetes.

References

1. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. Cochrane handbook for systematic reviews of diagnostic test accuracy. 1st ed. Chichester (UK): John Wiley & Sons, Ltd; 2023. p. 249-325.
2. Tutz G, Schmid M. Modeling discrete time-to-event data. New York: Springer Series

Abstracts of Contributed Talks

in Statistics; 2016.

3. Zapf A, Albert C, Frömke C, Haase M, Hoyer A, Jones HE, Rücker G. Meta-analysis of diagnostic accuracy studies with multiple thresholds: comparison of different approaches. Biometrical journal. 2021;63:699-711.

2024-02-29 09:40 - 10:00, AM S1

SES-03: Meta-Analysis 1

Identifying the risk of sample overlap in meta-analysis of registry-based studies

Zhang, Zhentian; Mathes, Tim

Universitätsmedizin Göttingen, Germany

There is a risk of sample overlap, when observational studies using existing registries are included in a meta-analysis. More specifically, when combining the results of the studies in the meta-analysis, some observations could be counted multiple times because they are included in the samples used by multiple studies' reports, e.g., when different articles use the data from the same trial. This could increase the alpha error, which potentially leads to wrong conclusions.

The sample overlap could be resolved by excluding all overlapping observations, if identification information such as personal IDs is available. However, in already aggregated data such as journal articles and study reports, this information is usually not available. Here we will emphasize on overlap analysis using aggregated data, which could also be applied to IPD.

We start with the quantification of the degree of the overlap. Our approach transforms the task of finding overlapping observations between study samples into identifying overlaps in samples' key characteristics, e.g., the range of time and region of the observations. For the first step, we suggested and developed mathematical definitions of several central concepts of our approach, such as definitions of overlap structure, in order to keep the presentation of the approach concise. In our next step, we constructed estimators for the degree of overlap in each of the identified subsets. To do this, we first defined the key characteristics of a sample, whose mutual exclusiveness excludes the possibility of overlap among study samples. Then, we code the samples' key characteristics into a collection of binary vectors and use the product of a normalized multilinear function of the binary vectors (product of the normalized dot product in the case of pairs) to estimate the degree of overlap.

To complement our methods, we suggested ways to visualize the overlap structure among the studies within the meta-analysis. We used both fictional and real examples to show the value of our visualization.

We applied our methods to existing meta-analyses and were able to confirm the viability and necessity of our method in real-world scenarios.

Abstracts of Contributed Talks

Little research has been done to address this challenge so far. In our opinion, this will be a growing issue because of the increasing use of registries and other pre-existing data in evidence synthesis. Our approach could be a good starting point for further research on the sample overlap problem and could improve the quality of evidence synthesis in the future.

2024-02-29 10:00 - 10:20, AM S1

SES-03: Meta-Analysis 1

Addressing Challenges in Subgroup-Specific Treatment Effects and Aggregation Bias in Meta-Analysis

Panaro, Renato Valladares; Röver, Christian; Friede, Tim

University Medical Center Göttingen, Germany

Randomized controlled trials (RCTs) not only furnish overall effect estimates but also frequently present aggregated outcomes for subgroup-specific treatment effects. Subgroup-specific effects often grapple with imbalances in subgroup prevalence. Applying standard meta-analysis methods to scrutinize these subgroups independently may lead to inconsistencies in treatment-by-subgroup interaction estimates, especially when incorporating large single-subgroup trials (WHO, 2020; WHO, 2021). This issue is commonly acknowledged in the individual patient data (IPD) meta-analysis literature as aggregation/ecological bias (Hua et al., 2017).

Recent approaches in meta-analysis strive to address these challenges, such as deriving reference subgroup effects from interaction residuals (Godolphin et al., 2023) or conditioning the subgroup analysis on the prevalence (Sørensen et al., 2023). However, estimation-matching approaches have shown unsatisfactory behavior when including single-subgroup trials (prevalence of one), particularly in the presence of substantial between-trial heterogeneity in treatment-by-subgroup interactions. In response to these concerns, our study conducted a thorough examination of conflicting findings in unbalanced meta-analysis. The comparison in our study includes inverse-variance and alternative weighting schemes (Henmi and Copas, 2010), as well as frequentist and Bayesian approaches for treatment-by-subgroup interaction meta-analysis of single-subgroup trials. The investigation also includes the bivariate random-effects model by van Houwelingen et al. (2002) to account for potential correlation in between the random-effects.

References

The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. JAMA. 2020 Oct 6;324(13):1330-1341.

The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group. Association Between Administration of IL-6 Antagonists and Mortality Among Patients Hospitalized for COVID-19: A Meta-analysis. JAMA. 2021 Aug 10;326(6):499-518.

Abstracts of Contributed Talks

Hua, H., Burke, D. L., Crowther, M. J., Ensor, J., Tudur Smith, C., & Riley, R. D. (2017). One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. *Statistics in Medicine*, 36(5), 772-789.

Godolphin PJ, White IR, Tierney JF, Fisher DJ. Estimating interactions and subgroup-specific treatment effects in meta-analysis without aggregation bias: A within-trial framework. *Res Synth Methods*. 2023 Jan;14(1):68-78.

Sørensen, A. L., & Marschner, I. C. (2023). Linear mixed models for investigating effect modification in subgroup meta-analysis. *Statistical Methods in Medical Research*, 32(5), 994-1009.

Röver C, Bender R, Dias S, Schmid CH, Schmidli H, Sturtz S, Weber S, Friede T. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods*. 2021 Jul;12(4):448-474.

Henmi, Masayuki, and John B. Copas. "Confidence intervals for random effects meta-analysis and robustness to publication bias." *Statistics in Medicine* 29.29 (2010): 2969-2983.

Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002 Feb 28;21(4):589-624.

2024-02-29 13:20 - 13:40, AM S2

SES-04: Meta-Analysis 2

Trials and triangles - The geometric interpretation of multi-arm studies in network meta-analysis

Rücker, Gerta

Medical Center - University of Freiburg, Germany

To enter the treatment effect estimates of a multi-arm study in a network meta-analysis (NMA), it is sufficient to know the contrasts to a baseline treatment. However, this does not hold for the variances of the contrasts, as can be explained by a geometric analogy.

A three-arm study can be represented by an acute-angled triangle, its side lengths representing the standard errors of the pairwise contrasts, and its angles (more precisely, their cosines) representing the (non-negative) correlations between contrasts. A necessary condition for consistency of variances is the triangle inequality.

For the NMA, this analogy explains why we need at least three pieces of variance information for entering data of a three-arm study: Either the variances of all three contrasts, or two variances and a covariance (or correlation), or one variance and two covariances, or all three covariances.

A four-arm study is represented by a tetrahedron with its faces being acute-angled triangles, and we need information on all six variances for entering its data into a network meta-analysis. Analogous statements hold for multi-arm studies in general, that are represented by generalized polyhedrons.

References

Lu G, Ades AE. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009 Oct; 10(4):792–805. Available from:
<https://doi.org/10.1093/biostatistics/kxp032>

2024-02-29 13:40 - 14:00, AM S2

SES-04: Meta-Analysis 2

A comparison of different software tools to support systematic reviews – considering the size of the training dataset and the machine learning methods used.

Schröder, Christin; Brewer, Madisen

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Germany

Introduction: Systematic reviews (SRs) are essential for various fields to reach a refined, evidence-based conclusion and fill knowledge gaps. Countless primary research studies are published daily, resulting in increased workload for SRs. To cope with this growing workload, the development and use of automated software tools is becoming increasingly popular (1-6).

SR tools use different machine learning (ML) methods. Previous studies have not addressed the impact of ML methods, and there is insufficient information on the requirements of characteristics of a good training data set (TDS). The objective of this work is to compare the different software tools and their respective ML methods and to investigate the impact of the TDS characteristics on the ML algorithm results.

Method. We performed a general search for SR tools. This involved extracting SR tools named from previous review articles and searching for others via Pubmed and GoogleScholar. The SR tools found were examined based on various criteria, e.g. costs, multiple reviewer allowed or type of machine learning support. Tools, which were suitable to answer our study question, were used to validate selection performance. Three different training data sets were built from an existing review with different numbers of labeled articles. For all datasets in all tools, we determined the number of false positives, false negatives, true positives (relevant) and true negatives (irrelevant).

Results: A total of 36 tools were identified and studied. All of them support the SR process, but in different ways. For our research questions, Rayyan, EPPI-Reviewer, Nested Knowledge, and ASReview Lab were selected as the most appropriate tools. The results in terms of number of correctly classified articles, false positives, and percentage of work saved varied widely among the tools. One tool allowed the user to choose between different ML methods. Differences were found between the methods in terms of performance. Larger differences in performance were found between the datasets.

Discussion: It was difficult to compare our results with others, since these questions have received little attention so far or other software tools have been investigated. Our results in terms of labor time saved (45%) were within the range that other studies have found

so far (9-99%; 3, 7, 8). It was surprising that performance varied widely between TDSs and larger TDS did not lead to better performance. Further studies should investigate this in more detail.

References

- 1.Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environmental Evidence*. 2018;7(1):8.
- 2.Olofsson H, Brolund A, Hellberg C, Silverstein R, Stenström K, Österberg M, et al. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. *Res Synth Methods*. 2017;8(3):275-80.
- 3.Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Ab-strackr, a semi-automated online screening program for systematic reviewers. *Syst Rev*. 2015;4:80.
- 4.Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. *Syst Rev*. 2020;9(1):73.
- 5.van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*. 2021;3(2):125-33.
- 6.Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA, editors. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. International Health Informatics Symposium; 2012.
- 7.Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Systematic Reviews*. 2019;8(1):278.
- 8.Burgard T, Bittermann A. Reducing Literature Screening Workload With Machine Learning. *Zeitschrift für Psychologie*. 2023.

2024-02-29 14:00 - 14:20, AM S2

SES-04: Meta-Analysis 2

Path-Based Approach for Detecting and Assessing Inconsistency in Network Meta-Analysis: A Novel Method

Rajabzadehtahmasebi, Noosheen; Papakonstantinou, Theodoros; Nikolakopoulou, Adriani

Institut für Medizinische Biometrie und Statistik, uniklinik freiburg, Germany

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.

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 to 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[1], loop-specific[2], and the design-by-treatment interaction model[3], 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 in properly pinpointing the inconsistency in the network.

In the second part, we emphasize the necessity for a path-based approach[4]. 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.

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

References

1. Dias, S., Welton, N.J., Caldwell, D.M. and Ades, A.E. Checking consistency in mixed treatment comparison meta-analysis. *Statist. Med.* 2010; 29: 932-944.
<https://doi.org/10.1002/sim.3767>
2. Heiner C. Bucher, Gordon H. Guyatt, Lauren E. Griffith, Stephen D. Walter. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*. 1997; Volume 50, Issue 6, Pages 683-691, ISSN 0895-4356, [https://doi.org/10.1016/S0895-4356\(97\)00049-8](https://doi.org/10.1016/S0895-4356(97)00049-8)
3. Higgins, J.P.T., Jackson, D., Barrett, J.K., Lu, G., Ades, A.E. and White, I.R. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res. Syn. Meth.* 2012; 3: 98-110. <https://doi.org/10.1002/jrsm.1044>
4. Papakonstantinou T, Nikolakopoulou A, Rücker G et al. Estimating the contribution of studies in network meta-analysis: paths, flows and streams [version 3; peer review: 2 approved, 1 approved with reservations]. *F1000Research*. 2018; 7:610. <https://doi.org/10.12688/f1000research.14770.3>)

2024-02-29 14:20 - 14:40, AM S2

SES-04: Meta-Analysis 2

A re-analysis of about 60.000 sparse data meta-analyses suggests that using an adequate method for pooling matters.

Schulz, Maxi¹; Kramer, Malte²; Kuss, Oliver³; Mathes, Tim¹

¹Department of Medical Statistics, University Medical Center Göttingen, Germany

²Epidemiology Unit, German Rheumatism Research Centre Berlin, Germany

³Institute for Biometrics and Epidemiology, German Diabetes Center (DDZ), Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Germany

Meta-analyses involving a small number of trials or rare events pose a challenge for statistical analysis. Evidence suggests that conventional two-stage statistical methods can lead to distorted results in these sparse data situations. Though better-performing one-stage methods have become available in recent years for such meta-analyses, these methods appear not sufficiently implemented and two-stage methods are still often used. The actual impact of using two-stage methods in practice, however, remains unknown. This study aims to quantify the impact by re-analysing meta-analyses from the Cochrane Database of Systematic Reviews (CDSR) in two sparse data situations: when meta-analyses included trials with zero events in one or both arms, or when meta-analyses contained only a few trials.

For each scenario, we computed one-stage statistical methods, namely the generalized linear mixed model (GLMM), the beta-binomial model (BBM) and the Bayesian binomial-normal hierarchical model using a weakly informative prior (BNHM-WIP). We then compared their impact on the results to the conventionally used two-stage methods, namely the Peto-Odds-Ratio (PETO) and DerSimonian-Laird method (DL) in case of zero event trials and DL, the Paule-Mandel (PM) and restricted maximum likelihood (REML) method in the scenario of few trials.

While all methods showed similar estimates for the pooled treatment effect, the results showed large variability in the statistical precision (length of CI and statistical significance) between the methods. Specifically, two-stage methods in the zero event situation tended to estimate narrower CIs resulting in more significant meta-analyses than the one-stage methods. While differences between the two-stage and one-stage methods are less evident in the few trial situations, the one-stage methods proved less frequent statistically significant.

This confirms previous studies suggesting a high number of false-positive results in real meta-analyses in the case of sparse data situations. In addition, the differences in the

results provide evidence that method choice has a substantial impact on the outcome of meta-analyses and encourages the careful choice of an adequate method. Specifically in the situation of zero event trials, the BBM and BNHM-WIP appear to be promising candidates while using BBM, and additionally the PM and REML model for sensitivity analyses appears reasonable in the situation of few trials. Furthermore, Bayesian methods with carefully selected priors can be an alternative in the latter situation. Overall, the results advise against relying solely on the outcome of a single meta-analysis method in the case of sparse data situations.

2024-03-01 09:00 - 09:20, AM S2

SES-06: Agricultural and Biological Statistics 1

A random-regression model for analyzing the genetic variance of litter weight in mice

Reinsch, Norbert; Ding, Xi; Jähnel, Ricarda

Forschungsinstitut für Nutztierbiologie (FBN), Germany

In research with laboratory mice the trait litter weight has frequently served as an indicator for maternal ability. Measurements are taken by putting all members of a litter jointly on a scale. In order to avoid confounding with litter size the number of pups can be standardized to a predefined number. This is, however, not possible if litter weight is measured at birth and not e.g. at weaning. Models previously used for quantitative-genetic analyses of litter weight data do not sufficiently account for the variable numbers of animals in each litter, nor do they account for any possible genetic effect of the litter's father. As an alternative, we propose a random regression model for total litter weight with random linear regression slopes on litter size as genetic effects for the mother (dam) and for the father (sire) of each litter. Genetic (co-)variances for the dam and the sire effect were estimated for various mouse lines kept at the FBN, assuming constant residual variation over the distribution of litter sizes. The presentation subsumes details of model assumptions, interpretation of parameters and estimates of variance components, including genetic correlations with litter size. We briefly discuss aspects of possible applications in other multiparous species like pigs, rabbits, or minks.

2024-03-01 09:20 - 09:40, AM S2

SES-06: Agricultural and Biological Statistics 1

Optimising sampling efforts in hierarchical Lincoln-Petersen experiment: Towards precise population size estimation

Chin, Su Na^{1,2}; Overstall, Antony¹; Böhning, Dankmar¹

¹Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton.

²Faculty of Science and Natural Resources, Universiti Malaysia Sabah

Capture-recapture methods play a crucial role in population size estimation, yet the exploration of its sampling effort remains limited. This paper aims to investigate the concept of sampling effort within the framework of a Lincoln-Petersen type experiment. We explore scenarios where two capture occasions are partitioned into numerous sub-occasions, sum up to T. Our findings suggest that when individual capture probabilities are equal on both occasions, optimal allocation involves assigning $T/2$ sub-occasions to each occasion for even T; while for odd T, allocating $(T-1)/2 + 1$ sub-occasions to one occasion and distributing the rest to the other. In cases with varying capture probabilities across the two occasions, we established that the most favourable sub-occasion distribution can be determined by maximizing the product of detect probabilities in the two occasions. These findings are illustrated through the application of simulated wildlife population studies, demonstrating practical strategies to enhance the precision of population size estimations. This paper sheds light on the significance of taking into account the amount of sampling efforts required to obtain accurate estimates of population size and provides useful recommendations for improving the design of capture-recapture studies.

2024-03-01 09:40 - 10:00, AM S2

SES-06: Agricultural and Biological Statistics 1

Optimizing Genetic Gain and Haplotype Diversity in Genomic Selection through Mendelian Sampling-Based Similarity Matrices

Musa, Abdulraheem Arome; Reinsch, Norbert

Research Institute for Farm Animal Biology (FBN), Germany

In the pursuit of long-term sustainability in livestock breeding, optimizing both genetic gain and haplotype diversity is crucial. Mendelian Sampling Variability (MSV), influenced by the heterozygosity and linkage phases of parental haplotypes, serves as a key factor in maximizing genetic diversity and the probability of breeding elite offspring. Although recent genomic selection methods that are based on an index combining Genomic Expected Breeding Values (GEBV) with MSV offer better haplotype diversity than those relying solely on GEBV, they can also inadvertently reduce overall genetic diversity by selecting genetically similar parents. To address this issue, we introduce a novel haplotype similarity matrix, denoted as S , derived analytically using additive marker effects, phased genotypes, and recombination rates. The matrix element s_{ij} is defined as the absolute covariance between the additive genetic values of the paired gametes based on their segregating marker patterns on a chromosome for parents i and j . High similarity values suggest that parent pairs share many heterozygous markers in the same linkage phase, which have large additive effects on the trait. Unique to this matrix are individual MSVs incorporated into its diagonal elements, providing a nuanced view of MSV. Simulation studies substantiate the efficacy of our approach: compared to existing index-based methods, we observed substantial increases in genetic gain and the preservation of genetic diversity after fifty generations of selection. This approach could be particularly beneficial for breeding programs constrained by small population sizes or where intense selection is practiced. Future work aims to broaden the scope of methodological validation and investigate issues related to scalability and efficacy.

2024-03-01 10:00 - 10:20, AM S2

SES-06: Agricultural and Biological Statistics 1

Challenges in statistical consulting for Animal Science

Schnabel, Sabine K.

Wagenigen University and Research, Netherlands, The

Each research domain comes with its own data and analysis challenges. Here we are presenting a few of the common situations encountered when collaborating with researchers in Animal Science. We focus on experimental studies concerning farm animals that are housed in stationary buildings. Most data are collected on cattle, pigs and chickens. These animals are often housed in smaller or larger pens with multiple animals in one section. Units are often equipped with different installations in addition to the required standards. These are often part of the studies, for example different types of flooring, heating, number of animals per pen or extra equipment such as brushes.

As the aims of the analyses are very diverse, also the type of variables and aggregation levels can be very diverse. In addition to numeric variables we also often deal with percentages or proportions (e.g. proportions of active animals), scores (e.g. for health), binomial data etc. Data are recorded on animal level or aggregated e.g. on pen level. In addition observations are often longitudinal. In the examples we mainly used a (generalized) linear mixed model framework.

2024-03-01 10:40 - 11:00, AM S2

SES-07: Agricultural and Biological Statistics 2

Promises and limitations of applying structural equation modelling techniques from psychology in other disciplines exemplified in geoecology

Rieger, Alex¹; de Haan, Jan¹; Schneider, Anne-Kathrin²; Schibalski, Anett³; Egert, Frank¹; Schröder-Esselbach, Boris³

¹TU Braunschweig, Institut für Psychologie, Abteilung Psychologische Methodenlehre und Biopsychologie

²Thünen Institut für Biodiversität

³TU Braunschweig, Institut für Geoökologie, Abteilung Landschaftsökologie und Umwelt-systemanalyse

In many disciplines structural equation modelling (SEM) is used to represent the relationship of manifest variables and latent factors and the implied covariance structures. Manifest variables are then regarded as imperfect indicators of the supposedly underlying latent factor(s). These statistical methods (partly) originate in psychology, but have been picked up and used in many other disciplines such as economics, behavioral genetics, political science and ecology.

We aim to show how recent developments in SEM can be useful in other disciplines and to illustrate this by the example of (geo-)ecology. There, a relatively recent review paper by Fan et al. (1) called for the use of more advanced SEM techniques, outlining advantages of methods such as SEM for temporal dynamics, hierarchical SEM or bayesian SEM due to substantive ecological considerations.

Through an interdisciplinary cooperation with researchers from geocology, we will present promises for using SEM techniques. This is especially explored in the context of models of temporal dynamics, which are frequently relevant due to feedback loops being an important part of theories in both disciplines. Therefore, we will focus especially on potential applications of recently developed continuous-time dynamic SEMs from psychology (2) in ecology. We will contrast this with other more conventional models in ecology from e.g., dynamical systems theory, such as differential equations. This will illustrate advantages and disadvantages of the use of more recent SEM techniques and give insight into the possibilities of exporting and exchanging statistical methods among disciplines in the biological and life sciences. This perspective on SEM may furthermore offer insight into interpretations of latent variables in cases where they do not model psychological, but physical characteristics.

We will then illustrate the use of a Bayesian continuous-time dynamic SEM by modeling population dynamics of *Aporrectodea longa* over time to provide a further analysis approach to data collected by Schneider et al. (3).

References

- 1.Fan, Y., Chen, J., Shirkey, G. et al. Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecol Process* 5, 19 (2016).
<https://doi.org/10.1186/s13717-016-0063-3>
- 2.Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, 23(4), 774–799. <https://doi.org/10.1037/met0000168>
- 3.Schneider, A. K., Hohenbrink, T. L., Reck, A., Zangerle, A., Schroeder, B., Zehe, E., & van Schaik, L. (2018). Variability of earthworm-induced biopores and their hydrological effectiveness in space and time. *Pedobiologia*, 71, 8-19.

2024-03-01 11:00 - 11:20, AM S2

SES-07: Agricultural and Biological Statistics 2

Analysis of greenhouse gas emission and the effect on rainfall outcomes in nigeria

Dosumu, Ebun Adegbola

Bowling Green State University, United States of America

In this work, we investigated emission of Greenhouse gases (GHGs) in Nigeria from 1960 to 2014. Time series were extracted from the archives of the WorldBank on Global GHGs emission per country which contained records on Carbon-dioxide (CO₂) from 1960-2014; Methane (CH₄) and Nitrous-oxide (NO₂) from 1970-2012. The Mann-Kendall trend test and Sen's slope were employed in establishing the presence of significant upward trend in the GHGs emission figures. Results showed that over the years CO₂ emissions rose by an average of 9.32% annually with a peak in 2005, CH₄ and NO₂ rose by an average of 3.30% and 2.58% respectively each reaching their peaks in 2007.

Also, Rainfall records for 1971 to 2012 collected from the Nigerian Meteorological Agency (NIMET) was examined alongside GHGs emission quantities to ascertain the effect of the GHGs emissions on Climatic conditions such as Rainfall across climatic zones in Nigeria. Correlation results showed that significant relationship exist between the GHGs emission quantities and Annual rainfall in the Northern Guinea, Sudan and Sahel Savannah climatic zones. These results further corroborates findings from the Rainfall analysis where significant upward shifts of 15.8%, 23.6% and 18.4% were observed in average annual rainfall in the aforementioned climatic zones respectively since the 1990s – which are possible cause of incessant flooding experienced in the country since 2012 till date. Furthermore, using the double exponential smoothing method, predictions were made for the GHGs emissions up till 2050 which showed a steady increase in the emission figures which calls for immediate interventions in order to mitigate subsequent effects of climate change caused by these emissions.

2024-03-01 11:20 - 11:40, AM S2

SES-07: Agricultural and Biological Statistics 2

Antibiotic resistance in pigs - analysis of the VetAmUR data to investigate the influence of antibiotic use on the temporal development of resistance in pig farming

Mers, Fiona¹; Bonzelett, Clarissa²; Kreienbrock, Lothar²; Rehberg, Betty²; Ickstadt, Katja¹; Tug, Timur¹

¹Department of Statistics, TU Dortmund University, Dortmund, Germany

²Department of Biometry, Epidemiology and Information Processing, University of Veterinary Medicine, Hannover, Germany

The prevention and treatment of infections in food-producing animals is an important factor in both the economic success of an agricultural farm and the overall animal health. Depending on the biosecurity measures in place within those farms, infectious diseases can spread rapidly if treatment is not provided. For many years, antibiotics were administered at high levels and sometimes uncontrolled to diseased animals as well as healthy animals at the same time. This promoted the general development of antibiotic resistance in livestock.

In this master thesis, the influence of antibiotic substances on the development of resistances are analyzed, more specifically, whether previous antibiotic intake can be associated with later resistances. The goal is to determine whether the administration of an antibiotic substance is associated with the corresponding timely development of a resistance. This is intended to reduce the dispense of antibiotics and ultimately the development of resistances, or to be able to assess the future development of resistance.

The basis of this study is the data of the VetAmUR (Veterinary Antimicrobial Usage and Resistance) project, which provides information on antibiotic dispensing and resistance in food-producing animals in Germany. Observations collected for swine production, including piglets, sows, runner pigs and porkers, from 2017 to 2022 are used. The statistical methods employed for the analysis comprise of time series regression models, especially ARMA and dynamic models with the defined target variable relative resistance. Results of the adjusted models are compared with each other and, if necessary, differences or similarities are worked out. The analysis showed, among other things, that treatment with substances containing the active ingredient Tiamulin has a statistically significant impact on the development of resistance.

2024-03-01 13:00 - 13:20, AM S2

SES-08: Time Series and Longitudinal Data

On detecting change points in unlabelled multivariate time series

Balestra, Chiara¹; Li, Bin¹; Mayr, Andreas²; Müller, Emmanuel¹

¹TU Dortmund, Germany

²IMBIE, University Hospital of Bonn, Germany

In volatile multivariate time series, dealing with variations of the input dimensions and their correlation structure is a complex task. The different distributions before and after a change-point can significantly impact the performance of statistical models fitted on historical data. Consequently, timely re-estimating and updating of parameters is often required. However, detecting these change points in real-time represents a severe challenge, as data labelling is often expensive or delayed in streaming data. Classical concept drift detection methods struggle to detect changes in correlations of multivariate time series' input variables, while the delayed labels do not allow for fast and efficient detection of the decaying accuracy of model predictions.

Our new methodological approach for change point detection works on unlabelled data streams and involves tracking correlation changes in the input variables using the slidSHAP series, a proxy representation of the original time series based on Shapley values. By introducing slidSHAP, we propose a fully unsupervised change detection method tailored for multivariate time series with categorical value domains and without class labels. Our method enables the detection of correlation-based changes within the data stream and highlights that distributional changes, even happening only in a few univariate dimensions, might drastically change the correlation structure among all univariate dimensions. Therefore, slidSHAP is more sensitive than any prior change point detection method. Furthermore, the slidSHAP series directly helps to visualize correlation changes in unlabelled multivariate discrete time series. Technically, the approach is divided into three main steps. First, we derive the slidSHAP series, implicitly encoding the correlations among the time series' input variables as a function of time. Second, we detect the change points on the slidSHAP series through independent statistical tests, and third, we relocate the found points to the original time notion.

In contrast to the widespread application of Shapley values in interpretable machine learning, we use this foundational game-theoretic concept to extrapolate information on the correlation structure of data streams. We achieve higher sensitivity than other unsupervised concept drift detectors toward multiple types of changes in empirical evaluations of synthetic and real-world data. As an ongoing project, slidSHAP holds high expectations for practical applications. In healthcare research, for instance, it could be applied

to streaming patient records, allowing for the early detection of subtle changes in the correlation of time-dependent parameters, thereby accelerating the diagnosis of potential anomalies.

References

Balestra C, Huber F, Mayr A, Müller E. Unsupervised Features Ranking via Coalitional Game Theory for Categorical Data. In DaWaK, 2022.

Li B, Balestra C, Müller E. Enabling the visualization of distributional shift using Shapley values. In DistShift @ NeurIPS, 2022.

Balestra C, Li B, Müller E. slidSHAPs – sliding Shapley Values for correlation-based change detection in time series. In DSAA, 2023.

2024-03-01 13:20 - 13:40, AM S2
SES-08: Time Series and Longitudinal Data

Deep mixture of linear mixed models for complex longitudinal data

Kock, Lucas¹; Klein, Nadja¹; Nott, David J.²

¹Technische Universität Dortmund, Germany

²National University of Singapore, Singapore

Mixtures of linear mixed models are widely used for modeling longitudinal data for which observation times differ between subjects. In typical applications, temporal trends are described using a basis expansion, with basis coefficients treated as random effects varying by subject. A key advantage of these models is that they provide a natural mechanism for clustering, which can be helpful for interpretation in many applications. Current versions of mixtures of linear mixed models are not specifically designed for the case where there are many observations per subject and a complex temporal trend, which requires many basis functions to capture. In this case, the subject-specific basis coefficients are a high-dimensional random effects vector, for which the covariance matrix is hard to specify and estimate, especially if it varies between mixture components. To address this issue, we consider the use of recently-developed deep mixture of factor analyzers models as the prior for the random effects. The resulting deep mixture of linear mixed model (DMLMM) is well-suited to high-dimensional settings, and we describe an efficient variational inference approach to posterior computation. We demonstrate the adaptability of our DMLMM approach across a range of real-world applications, each presenting distinct challenges.

First, we consider within subject prediction for an unbalanced longitudinal study on CD4 percentages of HIV-infected men. In a diagnostic context, within subject forecasting is of particular interest and our method reconstructs meaningful trajectories by combining information from each individual and the entire dataset. The Gaussian mixture model representation of the DMLMM is helpful for interpretation. For example, the CDF and therefore the risk of the CD4 percentage of an individual falling below a threshold at a given time can be easily calculated.

Secondly, we consider real data on malaria transmission in Afghanistan. We are interested in forecasting future case counts based on observed data. A nonlinear ordinary differential equation model is used to describe the temporal population dynamics, resulting in a complex likelihood free model. DMLMM is a suitable choice for predictive likelihood free inference, where each sample is a high-dimensional time series and a large number of basis functions is needed to estimate a complex temporal trend. We also discuss how the DMLMM can be used for model recalibration using prior-data conflict checks.

Third, we showcase the DMLMM for missing data imputation for gene expression data and find that our model is well calibrated even for a high level of missing information.

2024-03-01 13:40 - 14:00, AM S2

SES-08: Time Series and Longitudinal Data

Type-I-error rate inflation in mixed models for repeated measures caused by ambiguous or incomplete model specifications

Häckl, Sebastian¹; Koch, Armin¹; Lasch, Florian²

¹Hannover Medical School, Institute of Biostatistics, Hannover, Germany

²European Medicines Agency, Amsterdam, Noord-Holland, The Netherlands

Pre-specification of the primary analysis model is a pre-requisite to control the family-wise type-I-error rate (T1E) at the intended level in confirmatory clinical trials. Consequently, both the ICH E9 guideline and the European Medicines Agency (EMA) guidelines for multiplicity and missing data require that the primary analysis model is defined unambiguously to avoid multiplicity issues or data-driven model selection. This requirement equally applies to more complex models, such as mixed models for longitudinal data (MMRM), where general guidance for appropriate model specification is lacking or contradictory. In an initial empirical evaluation we investigated compliance with EMA guidelines and showed that MMRMs are poorly specified in study protocols. The magnitude of a potential T1E rate inflation due to multiplicity issues, however, is still unknown. In our recent research we aimed to quantify the magnitude of the T1E rate inflation that is caused by imprecise or ambiguous model specification depending on the type and number of unspecified model items as well as different trial characteristics (sample size, allocation ratio). We simulated a randomized, double-blind, parallel group, phase III clinical trial under the assumption that there is no treatment effect at any time point. The simulated data was analysed using different clusters, each including several MMRMs that are compatible with the imprecise pre-specification of the MMRM. T1E rates for each cluster were estimated. A significant T1E rate inflation could be shown for ambiguous model specifications with a maximum T1E rate of 7.6% [7.1%; 8.1%]. The results show that the magnitude of the T1E rate inflation depends on the type and increases with a higher number of unspecified model items. T1E rates are also higher for small and unbalanced trials. The results clearly emphasize the need for a thorough and precise specification of the chosen model for the primary analysis, despite the model's proven ability for a good performance. In fact, our simulations may only represent the lower bound and rather underestimate the true T1E rate inflation since we only applied a simple data generating mechanism that used a simple variance-covariance structure for repeated measures and created only complete data sets in the absence of missing data. In conclusion, we have shown that imprecise or ambiguous MMRM specifications may lead to a substantial inflation of the T1E rate and can damage the ability to generate confirmatory evidence in pivotal clinical trials.

References

1. Häckl S, Koch A, Lasch F. Type-I-error rate inflation in mixed models for repeated measures caused by ambiguous or incomplete model specifications. *Pharmaceutical Statistics*. 2023;1-16. doi:10.1002/pst.2328
2. Häckl S, Koch A, Lasch F. Empirical evaluation of the implementation of the EMA guideline on missing data in confirmatory clinical trials: specification of mixed models for longitudinal data in study protocols. *Pharm Stat.* 2019;18(6):636-644. doi:10.1002/pst.1964
3. International Conference on Harmonisation (ICH). Statistical Principles for Clinical Trials (ICH E9). 1998; (February).
4. European Medicines Agency (EMA). Points to Consider on Multiplicity Issues in Clinical Trials Discussion in the Efficacy Working Party. 2002.
5. European Medicines Agency (EMA). Guideline on Missing Data in Confirmatory Clinical Trials. 2011:1–12.

2024-03-01 14:00 - 14:20, AM S2

SES-08: Time Series and Longitudinal Data

Function-on-Scalar Regression (FoSR) with Wavelet Basis Functions for the analysis of Periodic Time-Series

Neumann, Konrad

Charité-Universitätsmedizin Berlin, Germany

The analysis of periodic time-series is pivotal in research involving data from wearables, such as smartwatches or accelerometer devices. Function-on-scalar regression (FoSR) stands out as a widely used method for analysing such data ([1] and [2]). FoSR represents a family of multivariate regression models that elucidate the relationship between covariates and time-series as a response.

The analysis in function-on-scalar regression (FoSR) invariably commences with the specification of a set of basis functions within the space of square-integrable functions (L^2). Numerous series of basis functions are currently employed. In this presentation, we compare the utilization of sine-cosine basis functions with that of wavelet basis functions [3]. The second focal point of the talk will centre around the multiple testing problem and the construction of corrected confidence bands. We will illustrate these methods using an example from [4].

References

- [1] Goldsmith J., Liu X., Jacobson JS. & Rundle A. New Insights into Activity Patterns in Children, Found Using Functional Data Analyses. *Med. Sci. Sports Exerc.* 48, 1723–1729 (2016).
- [2] Xiao L. et al. Quantifying the lifetime circadian rhythm of physical activity: A covariate-dependent functional approach. *Biostatistics* 16, 352–367 (2015).
- [3] Percival DB., Walden AT. *Wavelet Methods for Time Series Analysis*. Cambridge University Press (2000).
- [4] Rackoll T., Neumann K., Passmann S., Grittner U., Külzow N., Ladenbauer J., Floel A. Applying time series analyses on continuous accelerometry data- A clinical example in older adults with and without cognitive impairment. *PloS One* 16(5) (2021).

2024-02-28 14:40 - 15:00, AM 2

SES-10: Estimands

Treatment effect measures in clinical trials with time-to-event outcomes: it is time to apply estimand thinking

Mütze, Tobias¹; Lanius, Vivian²

¹Novartis Pharma AG, Basel, Switzerland

²Bayer AG, Wuppertal, Germany

The ICH E9(R1) addendum on estimands and sensitivity analysis in clinical trials calls for clarity and precision when describing the clinical question of interest. It defines the estimand as a population-level summary of “what the outcomes would be in the same patients under different treatment conditions being compared”. Thus, while not explicitly using the term “causal”, both the framework and language used in ICH E9(R1) are aligned with causal reasoning.

In randomized clinical trials with a time to event endpoint, the hazard ratio is still the most common effect measure. Post-randomization (i.e., intercurrent) events are often addressed through censoring without explicitly discussing or stating the underlying clinical question of interest. Alternative summary measures, especially on a probability scale or time scale, are rarely considered in clinical trials despite being seemingly easier to interpret and potentially more meaningful to patients and practitioners.

In this talk we will present the status of ongoing discussions among a working group of statisticians from different pharmaceutical companies on estimands for clinical trials with time-to-event data. In detail, we will discuss what key clinically meaningful questions of interest are when measuring the effect of an intervention through a time-to-event endpoint. We will reflect on the interpretation of various summary measures, the role of causality when defining an estimand in a clinical trial, and on how the choice of the estimand affects the design of a trial with a time-to-event endpoint. We will also elaborate on the practicalities of summarizing the effect of treatment through a single number in a time to event setting and discuss separating testing and estimation.

2024-02-28 15:00 - 15:20, AM 2

SES-10: Estimands

Causal Inference with Continuous Interventions

Schomaker, Michael¹; McIleron, Helen²; Denti, Paolo²; Diaz, Ivan³

¹Ludwig-Maximilians Universität München, Germany

²Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, South Africa

³Division of Biostatistics, Department of Population Health, New York University Grossman School of Medicine
New York, United States of America

Currently, there are limited options to estimate the effect of variables that are continuous and measured at multiple time points on outcomes, i.e. through the dose-response curve. However, these situations may be of relevance: in pharmacology, one may be interested in how outcomes of people living with -and treated for- HIV, such as viral failure, would vary for time-varying interventions such as different drug concentration trajectories. A challenge for doing causal inference with continuous interventions is that the positivity assumption is typically violated. To address positivity violations, we develop projection functions, which reweigh and redefine the estimand of interest based on functions of the conditional support for the respective interventions. With these functions, we obtain the desired dose-response curve in areas of enough support, and otherwise a meaningful estimand that does not require the positivity assumption. We develop g-computation type plug-in estimators for this case. Those are contrasted with using g-computation estimators in a naïve manner, i.e. applying them to continuous interventions without addressing positivity violations. The ideas are illustrated with longitudinal data from HIV+ children treated with an efavirenz-based regimen. Simulations show in which situations a naïve g-computation approach is appropriate, and in which it leads to bias and how the proposed weighted estimation approach recovers the alternative estimand of interest.

2024-02-28 15:20 - 15:40, AM 2

SES-10: Estimands

The estimand framework for diagnostic accuracy studies

Fierenz, Alexander; Rackow, Britta; Badpa, Mahnaz; Zapf, Antonia

University Medical Centre Hamburg-Eppendorf, Germany

A diagnostic test provides a statement about a target condition of an individual. This target condition is evaluated based on diverse clinical information such as symptoms, laboratory values or physical examination.[1] Diagnostic accuracy studies assess the precision of a diagnostic test. In those studies, the diagnostic test is compared to the true state, defined by the reference test. Based on the result of the reference test, patients can be assigned to the two target conditions.

A diagnostic study is performed to estimate test accuracy in daily practice. If the study design deviates from the practical use, the estimated test accuracy may be biased. Therefore, components of the study objective must be defined a priori to avoid discrepancies between the study and daily practice. For example, the study population has to be selected based on the later target population. Moreover, various interfering events could occur that can lead either to non-existent test results or influenced test decisions. It should be determined how to handle these events.

The trial objective must be formulated during the planning of a diagnostic study.[2] This objective is then translated into the clinical question of interest. For treatment studies, the estimand framework consists of different attributes to define the estimand, which must be aligned with the stated clinical question of interest.[3] We will present an estimand framework for diagnostic studies, including the attributes target population, index test, target condition, accuracy measurement and strategies for interfering events.

To illustrate this framework, we will present an application example, evaluating a computed tomography (CT) scan to detect lung carcinoma. We will define the estimand for this study and discuss several potential interfering events and strategies to handle them, such as premature termination of the CT scan because of coughing.

References

1. Knottnerus JA, van Weel C, Muris JWM. Evaluation of diagnostic procedures. *BMJ* 2002; 324(7335):477–80.
2. Lawrence R, Degtyarev E, Griffiths P, et al. What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes

Abstracts of Contributed Talks

in clinical trials? J Patient Rep Outcomes 2020; 4: 68.

3. ICH. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials; 2020 [cited 2022 Mar 21]. Available from: URL: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.

2024-02-28 15:40 - 16:00, AM 2

SES-10: Estimands

Estimande in early phase studies, with an example in atopic dermatitis

Klein, Stefan; Friedrichs, Frauke; Kunz, Michael

Bayer AG, Germany

The estimand strategy will be presented for a still running phase2a study in atopic dermatitis, comparing active drug vs placebo in 72 planned patients, with EASI-75 response as primary endpoint. We will compare our estimands against the proposals given in Bissonnette et al. (1) for estimands in atopic dermatitis (AD), as well as against the estimand strategy in published phase III studies in AD.

Given this particular example, we will discuss some aspects of estimand concept in PoC studies. Especially we will discuss - scientific objectives to be explored in early phase studies, - the impact of different intercurrent event strategies in PoC studies, - the usage of principal stratum strategy and a useful choice of analysis population in PoC studies, - the availability of information on the type of intercurrent events that might occur, - methods to keep the assignment of patients to data sets and the assignment of datapoint sets unbiased, and - how usage of the hypothetical strategy might change the choice of endpoints in PoC studies.

References

1. Bissonnette R, Eichenfield LF, Simpson E, Thaci D, Kabashima K, Thyssen JP, et al. Estimands for atopic dermatitis clinical trials: Expert opinion on the importance of intercurrent events. *J Eur Acad Dermatol Venereol*. 2023;37:976–983.

2024-02-28 16:20 - 16:40, AM S1

SES-11: Missing Data

Recoverability of Causal Effects in a Longitudinal Study under Presence of Missing Data

Holovchak, Anastasiia¹; Schomaker, Michael²

¹ETH Zürich, Switzerland

²LMU München, Germany

Missing data in multiple variables is a common issue. We investigate the applicability of the framework of graphical models for handling missing data to a complex longitudinal pharmacological study of HIV-positive children treated with an efavirenz-based regimen as part of the CHAPAS-3 trial. Specifically, we examine whether the causal effects of interest, defined through static interventions on multiple continuous variables, can be recovered (estimated consistently) from the available data only. So far, there exists no general algorithm for deciding on recoverability, and decisions have to be made on a case-by-case basis. We emphasize sensitivity of recoverability to even the smallest changes in the graph structure, and present recoverability results for three plausible missingness DAGs in the CHAPAS-3 study (directed acyclic graphs), informed by clinical knowledge. Further, we propose the concept of "closed missingness mechanisms" and show that under these mechanisms an available case analysis is admissible for consistent estimation for any type of statistical and causal query, even if the underlying missingness mechanism is of MNAR type. Simulations demonstrate how estimation results vary depending on the modelled missingness DAG. Our analyses are possibly the first to show the applicability of missingness DAGs to complex longitudinal real-world data, while highlighting the sensitivity with respect to the assumed causal model.

2024-02-28 16:40 - 17:00, AM S1

SES-11: Missing Data

Comparing propensity score methods combined with multiple imputation for controlling confounding: a case study on mantle cell lymphoma treatment regimens

Gutmair, Katja¹; Cunningham, Nicholas²; Silkenstedt, Elisabeth³; Kluin-Nelemans, Hanneke⁴; Dreyling, Martin³; Villa, Diego²; Hoster, Eva¹

¹Institute for Medical Information Processing, Biometry, and Epidemiology - IBE, LMU Munich, Munich, Germany

²BC Centre for Lymphoid Cancer and University of British Columbia, Vancouver, BC, Canada

³Department of Medicine III, University Hospital LMU Munich, Munich, Germany

⁴Department of Hematology, University Medical Center Groningen, University of Groningen, Groningen, Netherland

Randomized trials are considered the gold standard for minimizing confounding in the estimation of treatment effects. However, data of randomized treatment groups are not always available, e.g. when comparing pooled cohorts from different studies, needing alternative approaches for confounding control. Besides the option to control for confounding with multiple regression models that incorporate confounding variables, propensity score (PS) methods have been developed to balance baseline characteristics between treatment groups. The objective of this case study was to compare confounding adjustment using multiple regression models alone and in combination with optimal matching and inverse probability of treatment weighting (IPTW), with and without multiple imputation (MI) due to a high percentage of missing values in some confounding variables. In particular, the focus was on the applicability of the methods and differences in its results.

To compare these statistical methods for confounding control, we used two pooled cohorts from two different settings (clinical trial vs. routine care) receiving a particular treatment for mantle-cell lymphoma (MCL). The objective was to compare the clinical outcome of these two treatment regimens. Multiple Cox regression models were fitted adjusted for the following relevant prognostic factors: MIPI score (a clinical prognostic score), Ki67 (a cell proliferation marker) and cytology alone and in combination. Additionally, PS optimal matching with a 1 : 1 ratio and IPTW were applied. The PS was calculated using a multiple logistic regression including sex and the MIPI score. The PS methods were additionally combined with MI using multivariate imputation of chained equation (MICE) because of a high percentage of missing values in Ki67 and cytology. After the imputation of the missing values in Ki67 and cytology, these variables could be added to the logistic regression model for PS calculation.

Abstracts of Contributed Talks

Balance statistics for relevant prognostic factors included in the logistic regression model for PS calculation were slightly better after IPTW than after optimal matching, especially for Ki67 and cytology when combined with multiple imputation. However, balance for individual MIPI variables remained insufficient in both PS methods, in contrast to the aim of their application. All analysis methods consistently demonstrated that there was no significant difference between the clinical outcome of both treatment regimens with minimal differences in the hazard ratio and its 95% confidence interval between the different analysis methods for confounding control. In summary, this case study demonstrates that PS methods may not always serve as a suitable replacement for randomization.

2024-02-28 17:00 - 17:20, AM S1

SES-11: Missing Data

Various approaches to deal with missing data when estimating causal effects with targeted maximum likelihood estimation

Wiederkehr, Christoph Dominik

Ludwig-Maximilians-Universität, Germany

Directed acyclic graphs (DAGs), and specifically missing data DAGs (m-DAGs), provide an alternative framework, offering a way to specify assumptions beyond the MAR (Missing At Random) - MNAR (Missing Not At Random) dichotomy. The full potential of the m-DAG approach is realized when the emphasis is shifted away from classifying data as MAR or MNAR , and instead, it is directed toward defining detailed mechanisms. A considerable portion of these mechanisms not only enable unbiased estimation, but by definition, also facilitate the consistent estimation of recoverable parameters based solely on the available data, provided that an appropriate method is utilized [1].

Nonetheless, DAGs come with their limitations, including the inability to depict interactions.

Moreover, conditions required for the recoverability of a target parameter are similar to those for the identification of causal effects, namely consistency and well-defined interventions, positivity, and conditional independence conditions. Positivity violation, both in the sense of causal effect estimation and concerning data support, also plays a crucial role.

Building upon a previous simulation study that evaluated several methods for handling missing data in the context of estimating the average treatment effect (ATE) using targeted maximum likelihood estimation (TMLE) with data-adaptive approaches [2], this paper reproduces and extends the analysis and further investigates cases under positivity violation and more complex data generating processes (DGP). Furthermore the study implements and assesses a variety of TMLE methods in R. Results align with previous findings that no single method consistently outperforms across all m-DAGs and different scenarios. However multiple imputation (MI) with Amelia performed best regarding root mean square error (RMSE), while also achieving low bias on the reproduced simulation.

Non-multiple imputation methods like complete-case and extended TMLE exhibit lower relative bias for certain m-DAGs across various DGPs but underestimate model standard errors (ModSEs) especially in complex scenarios, providing poor coverage.

Abstracts of Contributed Talks

Parametric MI including interaction terms performed superior across all m-Dags and DGPs without interactions in terms of bias and also standard errors (SEs). For the complex DGPs all methods provided biased estimations, whereas MI with classification and regression trees performed sufficient coverage.

TMLE with the tmle-package [3] provided lowest bias, empSE, and RMSE across all scenarios and disabling cross validation demonstrates additionally the fastest runtime.

The study reveals the implications of increased complexity on the reliability of imputation and estimation methods, guiding future estimation, and sensitivity analysis procedures, and aiding the treatment of missing data in point-exposure studies.

References

- [1] American Journal of Epidemiology, Volume 187, Issue 12, December 2018, Pages 2705–2715, <https://doi.org/10.1093/aje/kwy173>
- [2] Dashti, S. G., Lee, K. J., Simpson, J. A., White, I. R., Carlin, J. B., and Moreno-Betancur, M. (2021). Handling missing data when estimating causal effects with targeted maximum likelihood estimation. arXiv, 1
- [3] Gruber, S. and van der Laan, M. (2012). tmle: An R package for targeted maximum likelihood estimation. Journal of Statistical Software, 51(13):1–35. doi:10.18637/jss.v051.i13.

2024-02-28 17:20 - 17:40, AM S1

SES-11: Missing Data

Adaptive predictor-set linear model: an imputation-free method for linear regression prediction on datasets with missing values

Planterose Jiménez, Benjamin¹; Kayser, Manfred¹; Vidaki, Athina¹; Caliebe, Amke^{2,3}

¹Department of Genetic Identification, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands

²Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany

³University Medical Centre Schleswig-Holstein, Kiel, Germany

Linear regression (LR) is vastly used in data analysis on continuous outcomes in biomedical research and clinical practice. Despite its popularity, LR is incompatible with missing data, a common occurrence in health sciences. Complete-case analysis or imputation can typically supplement LR in the context of inference with missing data. However, both workarounds are inadequate under the prediction paradigm, since they either fail to predict on incomplete records or ignore missingness-induced reduction in prediction accuracy. Here, we derive adaptive predictor-set linear model (aps-lm), capable of making predictions for incomplete data without the need for imputation. We employ a predictor selection operation, the Moore-Penrose pseudoinverse and the reduced QR-decomposition for the development of aps-lm. aps-lm are fitted on a reference dataset and yield a set of privacy-preserving parameters which can be shared to the public domain and applied on external datasets with missing entries without requiring imputation. Moreover, aps-lm provide a platform for robust prediction error modelling that take into account the pattern of missing values in the incomplete data even under extreme missingness. Via a simulation study, we benchmark our novel method against popular imputation strategies. As a result, adaptive predictor-set linear models display greater prediction accuracy and reduced bias compared to other tested alternatives under a wide range of sample sizes, goodness-of-fit, missing value types and covariance structures. Finally, as a proof-of-principle, we explore aps-lm applications in the context of epigenetic aging clocks, linear models that can assess a person's biological age from epigenetic data with promising clinical applications.

2024-02-29 09:00 - 09:20, AM S3

SES-12: High Dimensional Molecular Data

Robust statistical detection of interaction effects in high-throughput sequencing data

Stadler, Mara Stefanie^{1,2}; Müller, Christian L.^{1,2,3}

¹Helmholtz Munich, Germany

²Ludwig Maximilians University Munich, Germany

³Flatiron Institute, New York, USA

The advent of large-scale data (e.g., from biotechnology) has made the development of suitable statistical techniques a cornerstone of modern interdisciplinary research. These data often contain many features but limited sample size, and are accompanied by experimental noise. A common research question in data-driven observational studies is to determine how features impact a readout of interest. Typically, only a subset of features is relevant, and they may interact in a concerted fashion. Thus, a major concern is to identify these relevant effects from a large number of possible combinations of features. To address this, we propose a robust statistical workflow to recover interactions in the data-scarce regime. Our multi-stage approach uses a lasso model for hierarchical interactions combined with stability-based model selection in a replicate consistent workflow. We demonstrate its superior performance compared to state-of-the-art techniques using synthetic data and show its wide applicability in a number of different biological applications including histone modification-protein interactions.

2024-02-29 09:20 - 09:40, AM S3

SES-12: High Dimensional Molecular Data

Using gene-set tests on expression data of mRNA targets to predict miRNAs involved during West Nile virus infections

Boege, Franz Leonard; Ruff, Sergej; Selle, Michael; Hemandhar Kumar, Shamini; Jung, Klaus

University of Veterinary Medicine Hannover, Germany

Gene-set tests are regularly used to study the enrichment of pathways, GO-terms or other gene sets in a set of differentially expressed genes. mRNA target sets of miRNAs have also been defined as gene sets and analysed with gene-set tests [1, 2]. mRNA expression levels are affected by the deregulation of miRNAs, with each miRNA having its own target set of mRNAs. Here, we used gene-set tests to identify miRNAs involved in West Nile virus (WNV) infections. In contrast to the availability of public mRNA expression data, public miRNA expression data is scarcely available. For our study, we hypothesize that the relationship between miRNA and mRNA can be extorted by computational approaches via set-based tests on mRNA expression profiles to indirectly predict changes in miRNA expression profiles during viral infections.

Based on this hypothesis, we developed a computational algorithm for predicting miRNAs involved in WNV infections. Using five parallel miRNA and mRNA datasets we initially conducted a differential analysis on the miRNA data to identify potentially differentially expressed miRNAs between infected and non-infected individuals. Subsequently, mRNA data underwent gene set enrichment analysis, with mRNAs grouped based on miRNA target sets from several public databases.

For evaluation, we tested multiple different parameters of the algorithm. We utilized different target set databases (mirdb, pita, miranda, tarbase, mirtarbase, targetscan, mirecords) and related prediction cutoffs (which reflect the probability of an mRNA-miRNA interaction) and common statistical enrichment tests (romer, roast, global test, Fisher).

Our results show better prediction capabilities using the set information from targetscan and mirdb and prediction cutoffs >80. By analysing 8 independent mRNA datasets with the gene-set approach in the absence of parallel miRNA data, we identified a number of miRNAs related to WNV infections that occurred in several of these 8 data sets.

References

Abstracts of Contributed Talks

- [1] Artmann, S., Jung, K., Bleckmann, A., & Beissbarth, T. (2012). Detection of simultaneous group effects in microRNA expression and related target gene sets. PLoS One, 7(6), e38365.
- [2] van Iterson, M., Bervoets, S., de Meijer, E. J., Buermans, H. P., 't Hoen, P. A., Menezes, R. X., & Boer, J. M. (2013). Integrated analysis of microRNA and mRNA expression: adding biological significance to microRNA target predictions. Nucleic acids research, 41(15), e146-e146.

2024-02-29 09:40 - 10:00, AM S3

SES-12: High Dimensional Molecular Data

Development of metabolomic risk scores for Alzheimer's Disease

Tug, Timur^{1,2}; Liang, Donghai^{2,3}; Tan, Youran³; Gearing, Marla^{4,5}; Levey, Allan I.⁵; Lah, James J.⁵; Wingo, Aliza P.^{6,7}; Wingo, Thomas S.^{5,8}; Ickstadt, Katja¹; Hüls, Anke^{2,3}

¹Department of Statistics, TU Dortmund University, Dortmund, Germany

²Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

³Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

⁴Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA

⁵Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA

⁶Division of Mental Health, Atlanta VA Medical Center, Decatur, GA, USA

⁷Department of Psychiatry, Emory University School of Medicine, Atlanta, GA, USA

⁸Department of Human Genetics, Emory University, Atlanta, Georgia, USA

Background: Alzheimer's disease (AD) is a neurodegenerative disorder that affects memory, thinking and behavior. There is currently no cure for AD. To develop and optimize disease modifying therapies, it is essential to obtain knowledge about the specific biological pathways and molecules involved in the development of AD. Metabolomics is an emerging omics technology capable of characterizing tens of thousands of exogenous and endogenous metabolites and perturbations of biological pathways. Here, we developed metabolomics risk scores (MetRS), defined as weighted sums of selected metabolic features, to predict AD neuropathology and provide insights on the strength of association between an individual's metabolomic profile and AD.

Methods: Untargeted metabolomics profiling was conducted on frontal cortex samples from 162 donors of the Emory Goizueta AD Research Center brain bank. AD neuropathology was assessed using the CERAD (Consortium to Establish a Registry for AD) score, Braak stage and ABC score. We calculated MetRS for AD based on different regularization methods to select metabolic features and estimate weights for the relative importance of each feature: 1.) elastic net (linearly combines the L1 and L2 penalties of the lasso and ridge methods), 2.) pruning and thresholding (choose representant of each cluster (hierarchical clustering) and then removing variants with a p value over chosen level of significance), 3.) random forest (machine learning algorithm, which combines the output of multiple decision trees to reach a single result). The data were split into training and test data, with the training data being used for the development of the MetRS

Abstracts of Contributed Talks

and the test data for its evaluation based on McFadden R2 and strength of association (odds ratio, p-value from an ordinal logistic regression) with AD neuropathology.

Results: The donors' mean age at death was 76.7 years (SD: 10.2 years) and 73.8% had pathology-confirmed AD (classified as intermediate or high ABC grading). The number of metabolic features selected for the MetRS varied depending on the approach (elastic net: 58-62, pruning and thresholding: 823-906, random forest: 1238-1281). The pruning and thresholding approach identified significant associations between MetRS and all three neuropathology outcomes with $R^2=0.16-0.23$, also the random forest approach for all outcomes ($R^2=0.13-0.29$) and elastic net approach only for ABC score ($R^2=0.10$).

Conclusion/Outlook: These results provide a valuable first step in understanding and applying risk score methodology to metabolomics data. This approach might consequently be used to predict certain health outcomes including AD.

2024-02-29 10:00 - 10:20, AM S3

SES-12: High Dimensional Molecular Data

Evaluating deep learning models for cell detection and multi-class cell classification: a comparative analysis of metrics and solutions

Pfrang, David¹; Ghete, Denisia-Tabita²; Pontones, Martina²; Metzler, Markus²; Kock, Farina¹; Höfener, Henning¹; Westphal, Max¹

¹Fraunhofer Institute for Digital Medicine (MEVIS)

²University Hospital Erlangen (UKER)

Compared to binary classification tasks, multi-class problems are less commonly encountered in medical machine learning. At the same time, they pose additional methodological difficulties, particularly in cases with a large number of classes. In this work, we address the challenge of comparing and evaluating deep learning models for cell detection and subsequent multi-class cell type classification on image data of bone marrow aspirate smears from pediatric patients. The dataset consists of data from 117 patients, with approximately 50,000 cells classified into 50 cell classes by human annotators. Approximately 20% of the data is reserved for the test set. The data was obtained in a retrospective multi-center study and is currently being used for model development. We are in the process of planning the evaluation study using the previously unused test set. Our primary focus is to determine the metrics for evaluation and to identify the most important aspects of the statistical analysis.

Our methodology involves conducting literature searches, consulting with clinical experts, and employing analytical thinking. Through this process, we identified several important questions that impact the study design and statistical analysis. These questions include: (a) the choice of observational unit (cell vs. patient), (b) suitable metrics (cost-agnostic vs. cost-sensitive), (c) defining reasonable minimal acceptance criteria based on between-reader agreement, and (d) handling detected cells that are not suitable or not relevant for classification.

We have explored various methodological solutions with their respective advantages and drawbacks which will be outlined in this work. Based on a comparative analysis, we were able to find the most adequate solutions for our specific use case. These solutions have been specified in a statistical analysis plan for the forthcoming evaluation study.

2024-02-29 13:20 - 13:40, AM S3

SES-13: Application with Molecular Data

Companion diagnostics in oncology clinical trials: lessons from the practice

Weispfenning, Anke; Descamps, Tine

Bayer AG

Oncology clinical trials are looking increasingly towards inclusion of companion diagnostics (CDx) to tailor drug development towards patients personalized medicine. This abstract will discuss the practical aspects of developing CDx from the statistical point of view within a clinical trial through an anonymized case study.

A first in human study has been conducted within NSCLC patients. Based on pre-clinical research patients with a specific mutational profile are predicted to be more susceptible to treatment than patients without these mutations.

The development of a CDx method includes completing a bridging study in which the mutations of the eligible patient cohort are reassessed with the chosen diagnostic tool, this to confirm both the concordance and clinical validity of the investigational CDx.

Practical aspects complicating these studies include the nature of the original local method to diagnose the patients and the time of sampling. Patients enrolled in the FiH study have their mutational profile assessed at time of cancer typing at local sites using different biopsy (tissue vs liquid biopsy) and diagnostic methods ((rt)PCR, NGS, ...). As a result, original samples are taken between 5 years to 1 month before enrolment in the study and performance of the diagnostic method varies widely. Secondly a high number of missing values is to be expected, this due to both patient consent and CDx sample isolation or sequencing failures.

The sample size required for the bridging study must reflect these complications. The calculation adjusts for the expected discordance between samples taken for the local assessment and those taken for the CDx assessment. Positive samples are selected from patients in the eligible cohort, while negative samples will be obtained commercially. To estimate the sample size assumptions are made for the local sensitivity and specificity. These are both set at 95%. Additional discordance is expected between the local assessment and CDx assessment due to the difference in sampling time and biopsy nature, the estimated additional disagreement is set at 5%. The power has been set to 80% with alpha = 0.05. The sample size is calculated for an aimed CDx PPA and NPA of 90%. This brings the estimated sample size to 141 each.

The impact of missing data must be estimated using a sensitivity analysis, where multiple imputation method is used to impute missing values for clinical trial samples.

2024-02-29 13:40 - 14:00, AM S3

SES-13: Application with Molecular Data

Genome-wide association studies on hedonic eating behaviour

Schliemann, Antje¹; König, Inke R.²; von Holt, Björn-Hergen²

¹Institut für Sozialmedizin und Epidemiologie, Germany

²Institut für Medizinische Biometrie und Statistik, Germany

Background: The brain's homeostatic process, which is normally responsible for maintaining a balance between food intake and energy expenditure, interact with hedonic, reward-driven processes. The latter can disrupt this balance and contribute to the development of obesity. An unexplored area in this context is the genetic predisposition to hedonic eating behaviour. The aim of our study is to identify genetic variants/chromosomal regions associated with this behaviour and thereby contribute to a better understanding of the genetic basis of eating behaviour.

Methods: The study included 447 participants from Luebeck (aged 26 to 51 years; 61.5% female), recruited as part of the SFB134 project (2014-2018), who showed differences in hedonic eating behaviour. We performed a genome-wide association study (GWAS) using the Illumina Infinium Global Screening Array (GSA) with 654,027 typed markers. After rigorous quality control and genotype imputation, 6,972,439 single nucleotide polymorphisms (SNPs) were tested for association with four factors derived from a previous factor analysis and associated with hedonic eating behaviour, adjusting for age, sex and population structure. Significant chromosomal regions were identified with a p-value of $< 5 \times 10^{-8}$, and regions with a smaller p-value of $< 1 \times 10^{-5}$ were also examined for potential significance in a larger sample.

Results: Significant associations with hedonic eating behaviour were found for two SNPs. The SNP rs78366984 achieved a p-value of 1.999×10^{-8} with an effect size of -0.463 factor units. This SNP is located between the genes ELP3 and PNOC. Similarly, SNP rs117126802 was identified with a p-value of 1.462×10^{-9} and an effect size of 1.744 factor units, located within the intron of the SLC26A4 gene. Other potential SNPs with a p-value $< 1 \times 10^{-5}$ were found: rs269965, rs61217966, rs328087, rs150078435, rs17344929, rs11767048, rs144292969 and rs7970517. These SNPs are located within or close to the genes SLC14A2, FAM207A, BNIP3L, PNMA2, CFAP53, ERBB4, CADPS2, LINC02760, LINC02759 and LOC643339. The effect sizes in these regions range from -0.348 to 1.229 factor units.

Conclusions: The results suggest the existence of loci associated with hedonic eating behaviour. Further analysis is crucial as only a limited number of these regions have been

investigated in genome-wide association studies. Investigation of genetic associations with hedonic eating behaviour may contribute to a better understanding of the causes of obesity and potentially offer new avenues for prevention and intervention.

2024-02-29 14:00 - 14:20, AM S3

SES-13: Application with Molecular Data

Partial interaction analysis in case-only defined clusters for high-dimensional biomarkers

Böhringer, Stefan; Maarseveen, Tjardo D.; Knevel, Rachel

Leiden University Medical Center, Netherlands, The

In case-control comparisons, cases might form sub-groups either through clinical definitions or the use of clustering on additional clinical or biomarker information. Information on cluster-membership might be lacking for controls either by definition or lack of measurements. We here investigate whether a given biomarker has different effect in the different case groups.

We first show that this problem cannot be analyzed by a standard interaction-model that uses cluster as main effect due to complete separation issues. Instead, we propose to use multivariate outcome models with outcomes representing case-group membership. We extend the MANOVA model to include shared effects of predictors on multiple outcomes which is nested within the full MANOVA model, allowing for model comparisons. The binary case is handled by an iterative weighted least squares approach. We proof consistency of our estimation procedure. We extend this approach to the high-dimensional setting by using penalized regression and perform inference using plausibility (1, 2). Penalized regressions use the vectorized formulation of MANOVA and can be performed with standard software.

In simulations, we investigate finite sample properties in the low and high-dimensional settings. We compare to a naive bootstrap procedure that stratifies the sample by case-subgroup.

In a data application, we analyze a rheumatoid arthritis (RA) data set (3) for which cases are clustered based on clinical criteria. These case groups are compared to a control group drawn from the general population. Biomarkers are SNP genotypes for which previous association with RA has been shown. We demonstrate the presence of effect heterogeneity of SNPs by comparing a shared effect model with the effect-per-subgroup model.

References

1. Martin, R. (2015). Plausibility functions and exact frequentist inference. *Journal of the American Statistical Association*, 110(512), 1552-1561.
2. Böhringer, S., & Lohmann, D. (2022). Exact model comparisons in the plausibility

- framework. *Journal of Statistical Planning and Inference*, 217, 224-240.
3. Maarseveen, T. D., Maurits, M. P., Bohringer, S., Steinz, N., Bergstra, S. A., Boxma-de Klerk, B. B., et al. & Knevel, R. (2023). Location of joint involvement differentiates Rheumatoid arthritis into different clinical subsets. *medRxiv*, 2023-09

2024-02-29 14:20 - 14:40, AM S3

SES-13: Application with Molecular Data

Is the early microbiome linked to childhood obesity? – A network perspective

Peschel, Stefanie^{1,2}; Depner, Martin³; von Mutius, Erika^{3,4,5,6}; Boulesteix, Anne-Laure^{2,7}; Müller, Christian L.^{1,2,8,9}

¹Department of Statistics, LMU München, Munich, Germany

²Munich Center for Machine Learning, Munich, Germany

³Institute of Asthma and Allergy Prevention, Helmholtz Zentrum München, Neuherberg, Germany

⁴Department of Pediatric Allergology, Dr Von Hauner Children's Hospital, LMU München, Munich, Germany

⁵Comprehensive Pneumology Center Munich (CPC-M), Munich, Germany

⁶German Center for Lung Research (DZL), Munich, Germany

⁷Institute for Medical Information Processing, Biometry and Epidemiology, LMU München, Munich, Germany

⁸Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany

⁹Center for Computational Mathematics, Flatiron Institute, New York, USA

Childhood obesity, a prevalent global health problem, is intricately linked to multiple factors including genetics, diet, and lifestyle. Over the past decade, the gut microbiome has emerged as another important factor contributing to obesity pathogenesis (1). Microbial association networks derived from high-throughput sequencing data offer a promising avenue for exploring the intricate dynamics of microbial communities in the context of obesity. Comparing association networks of the early microbiome between normal weight children and those who become obese at school age may shed new light on the factors contributing to the development of childhood obesity.

Due to the specific characteristics of microbiome data, such as compositionality and high dimensionality, not only the construction of microbial association networks, but also their comparison is challenging. In a previous project (2), we proposed a comprehensive analysis pipeline for the construction, analysis, and comparison of microbial association networks. Since network properties do usually not follow classical statistical distributions, permutation testing is an essential part of this workflow to compare networks between two groups. Due to the complexity of the methods and the high dimensionality of the data, performing a large number of permutations is often not feasible, resulting in low statistical power. A popular heuristic solution to this problem is to approximate extreme p-values by fitting a Generalized Pareto Distribution (GPD) to the tail of the distribution of the permutation test statistics (3). In another unpublished project, we proposed

an improvement to this method that uses a constrained fit of the GPD parameters to strictly avoid zero p-values. In the current study, we applied this method to data from the PASTURE study (4), which includes 16s rRNA sequencing data of the gut microbiome collected at 2 and 12 months of age. The aim is to detect differences in the microbial association networks of the early gut microbiome between children who become obese at school age and those who remain normal weight.

Following our original pipeline, where p-values are computed based on the empirical distribution of permutation test statistics, did not show significant differences after adjusting for multiple testing. The application of our newly proposed method based on the GPD approximation, however, revealed differences in network properties as well as differentially associated taxa. This allowed us to draw so-called differential networks, where two taxa are connected only if their association is significantly different between the two groups.

References

1. Castaner, O., Goday, A., Park, Y. M., Lee, S. H., Magkos, F., Shiow, S. A. T. E., & Schröder, H. (2018). The gut microbiome profile in obesity: A systematic review. *International Journal of Endocrinology*, 2018. <https://doi.org/10.1155/2018/4095789>
2. Peschel, S., Müller, C. L., von Mutius, E., Boulesteix, A. L., & Depner, M. (2021). NetCoMi: network construction and comparison for microbiome data in R. *Briefings in Bioinformatics*, 22(4), 1–18. <https://doi.org/10.1093/BIB/BBAA290>
3. Knijnenburg, T. A., Wessels, L. F. A., Reinders, M. J. T., & Shmulevich, I. (2009). Fewer permutations, more accurate P-values. *Bioinformatics*, 25(12), 161–168. <https://doi.org/10.1093/bioinformatics/btp211>
4. von Mutius, E., & Schmid, S. (2006). The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy*, 61(4), 407–413. <https://doi.org/10.1111/J.1398-9995.2006.01009.X>

2024-02-29 15:00 - 15:20, AM S3

SES-14: Genetic Epidemiology

Clinical utility of polygenic scores: A critical 2023 appraisal

Koch, Sebastian¹; Schmidtke, Jörg^{2,3}; Krawczak, Michael¹; Caliebe, Amke¹

¹Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität zu Kiel, Universitätsklinikum Schleswig-Holstein Campus Kiel, Kiel, Germany

²Amedes MVZ Wagnerstibbe, Hannover, Germany

³Institut für Humangenetik, Medizinische Hochschule Hannover, Hannover, Germany

Polygenic scores (PGSs) are numerical values that summarize the joint effects of multiple genetic variants on the risk of a common complex disease. They were first developed in 2009 in the context of schizophrenia and bipolar disorder, but have since been described for various common complex diseases. PGSs have been proposed as useful tools for disease prognosis, prevention and treatment in precision medicine. However, the clinical utility of PGSs in assessing disease risk or making therapeutic decisions is likely restricted as PGSs only account for the genetic component of a trait and disregard the impact of environmental and lifestyle factors.

Our survey examined the current state of PGSs for various diseases such as breast cancer, diabetes, prostate cancer, coronary artery disease, and Parkinson's disease, with a particular emphasis on how clinical scores could be enhanced by combining them with PGSs. Furthermore, we collected several examples of how PGSs can be and are used for clinical purposes, such as screening recommendations, drug stratification, and risk communication.

We observed that the diagnostic and prognostic performance of PGSs alone is for most diseases low, as expected from their limited heritability. A notable exception is type 1 diabetes, whose PGS achieved a remarkably high performance. We attributed this difference to the specific genetic architecture of type 1 diabetes, which is largely driven by the HLA locus. Moreover, combining a PGS with a clinical score led to moderate improvement of the power of either score at best. Combining PGSs with clinical risk scores improved the predictive performance by about 10% on average, depending on the disease, clinical score and age.

Despite the large number of PGSs reported in scientific literature, there are few external validations and prospective studies conducted on their clinical utility.

In conclusion, the benefit to individual patients or the health care system in general is still difficult to judge. PGSs may be suitable for optimizing conventional screening procedures

and the administration of medication, but there is a lack of prospective studies on this matter.

2024-02-29 15:20 - 15:40, AM S3

SES-14: Genetic Epidemiology

Tools for predicting the effects of genetic variants: a systematic review and practical guide

Riccio, Cristian^{1,2}; Jansen, Max Louis^{1,2}; Ziegler, Andreas^{1,2,3,4,5}

¹Cardio-CARE, Medizincampus Davos, Switzerland

²Swiss Institute of Bioinformatics, Lausanne, Switzerland

³Center for Population Health Innovation, University Heart and Vascular Center Hamburg, Germany

⁴University Medical Center Hamburg-Eppendorf, Germany

⁵School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Large-scale association analyses using whole-genome sequence data have become feasible, but understanding the functional impacts of these associations remains challenging. Many tools are available to predict the functional impacts of genetic variants. However, it is unclear which tool should be used in practice. In this work, we provide a practical guide to assist in selecting appropriate tools for variant annotation.

We conducted a MEDLINE search up to 7th June 2023 and included tools that are applicable to a broad range of phenotypes, can be used locally, and have been recently updated. Tools were categorized based on the types of variants they accept and the functional impacts they predict. Sequence Ontology terms were used for standardization.

In total, we identified 94 databases and software packages, encompassing 33 variation types and 130 functional impacts. Combining just the three tools Ensembl's Variant Effect Predictor, FAVOR, and WGSA allows researchers to make 90 of those functional predictions. 71 tools predict functional impacts that are not predicted by any other tool. 59 tools predict pathogenicity and can be used within the ACMG/AMP guidelines in a clinical context. We have developed a website that allows researchers to select tools based on desired variants and effects.

Despite the diverse functionalities of databases, some variation types, such as fusions, are not accepted by any tool. Future research incorporating tissue specificity, transcript variations, and trait specificity is needed to create tools for making more fine-grained and precise predictions.

URL: www.cardio-care.ch/research/projects/findable

2024-02-29 15:40 - 16:00, AM S3

SES-14: Genetic Epidemiology

Detecting interactions in High Dimensional Data using Cross Leverage Scores

Teschke, Sven^{1,2}; Ickstadt, Katja¹; Munteanu, Alexander¹; Schikowski, Tamara²

¹TU Dortmund, Deutschland

²IUF Düsseldorf

We are developing a variable selection method for regression models for Big Data in the context of Genetics. The method is intended for investigating the influence of SNPs and their interactions on health outcomes, which is a $p \gg n$ problem.

Motivated by Parry et al. (2021), we apply the so called cross leverage scores to directly detect important interactions while maintaining interpretability. The big advantage of this method is that it is not necessary to consider each possible interaction between variables individually, which would be very time consuming even for a moderately large amount of data. In a simulation study, we show that these cross leverage scores are directly linked to the importance of a variable in the sense of an interaction effect.

Furthermore, we are developing methods for calculating cross leverage scores in very large datasets, which are common in the context of genetics. We divide the data set into subsets of variables (batches) and calculate the scores successively. Thus, we avoid complex and time-consuming computations of high-dimensional matrices. We compare these methods to existing approximation methods for calculating cross leverage scores with sketching (e.g. Drineas et al. (2012)) and evaluate these in simulation studies. We also applied this method to a real data set, the SALIA cohort study (Study on the Influence of Air Pollution on Lung, Inflammation and Aging) (Schikowski et al. (2005)). This study investigates the influence of air pollution on lung function, inflammatory responses and aging processes in 517 elderly women from the Ruhr area. In addition to data on influences of various environmental factors, data on over 7 million SNPs are also available for these women. We are exploring the influence of both SNP interactions and SNP environment interactions on various health outcomes.

References

- Drineas P., Magdon-Ismail M., Mahoney M.W., Woodruff D.P. (2012). Fast approximation of matrix coherence and statistical leverage. *J Machine Learning Research* 13, 3475-3506, doi: 10.5555/2503308.2503352.
Parry, K., Geppert, L., Munteanu, A., Ickstadt, K. (2021). Cross-Leverage Scores for Selecting Subsets of Explanatory Variables. arXiv e-prints, abs/2109.08399,

Abstracts of Contributed Talks

<https://arxiv.org/abs/2109.08399>.

Schikowski, T., Sugiri, D., Ranft, U. et al. Long-term air pollution exposure and living close to busy roads are associated with COPD in women. *Respir Res* 6, 152 (2005).
<https://doi.org/10.1186/1465-9921-6-152>

2024-02-29 16:00 - 16:20, AM S3

SES-14: Genetic Epidemiology

Estimating sparse graphical models in high dimensions

Foraita, Ronja; Mose, Kristof; Hanke, Moritz; Didelez, Vanessa

Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany

Discovering structures in high dimensions is especially relevant in genetics to investigate regulatory patterns. Numerous procedures have been proposed and the focus is often on graphical, i.e. conditional independence structures.

The challenges in high dimensions are typically addressed by assuming sparsity which requires regularization and thus the suitable choice of one or more tuning parameters.

In this talk we investigate the properties of various versions of the Stability Approach to Regularization Selection (StARS, Liu et al, 2010; G-StARS, Müller et al, 2016) when estimating sparse graph structures based on the graphical Lasso (gLasso, Friedman et al, 2008). StARS aims to select the tuning parameter which gives the most stable results under subsampling. While the StARS algorithm can be shown to have desirable theoretical properties under certain assumptions, these may be implausible in real-world settings. We therefore investigate the properties in an extensive simulation study, where we are especially interested in the ability to find specific graphical structures such as hubs, chains or “graphlets”, as well as structures typical of Erdös Reny and Scale free random graphs.

Our simulation study demonstrated that the selected graphs by StARS depend much more than one might expect on the sparsity of the true graphs. Moreover, our results suggest that even though it seems plausible that the tuning parameter should yield more stable graphs around the truth, it still mainly drives the neighbourhood size of nodes in the selected graphs. Therefore, many practically relevant structures cannot reliably be identified with this approach and a single tuning parameter. For illustration, we apply gLasso with the different variants of StARS to real genetic data from The Cancer Genome Atlas (TCGA) Research Network (<https://www.cancer.gov/tcga>).

References

- J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008
- H. Liu, K. Roeder, and L. Wasserman, “Stability approach to regularization selection (stars) for high dimensional graphical models,” *Advances in Neural Information Process-*

Abstracts of Contributed Talks

ing Systems, vol. 24, no. 2, pp. 1432–1440, 2010.

C. L. Müller, R. Bonneau, and Z. Kurtz, “Generalized stability approach for regularized graphical models,” arXiv preprint arXiv:1605.07072, 2016.

2024-03-01 09:00 - 09:20, AM S1

SES-15: Time-to-Event 1

On historically controlled survival trials

Feld, Jannik; Danzer, Moritz; Faldum, Andreas; Schmidt, Rene

Institut für Biometrie und klinische Forschung, Germany

The preferred method for analyzing the results of single-arm survival trials is the one-sample log-rank test. This test compares the survival outcomes of patients with a pre-determined reference survival curve, typically representing the expected outcome under the standard of care. However, conventional one-sample log-rank tests assume that the reference curve is known, disregarding the fact that it is often estimated from historical data, making it susceptible to sampling errors. Failing to consider the variability in the reference curve can lead to an increased type I error rate, as demonstrated in a prior study. In this paper, we introduce an improved one-sample log-rank test that accounts the sampling error associated with the reference curve, even when the full historical survival time data is not available. Our new test enables a valid historical comparison of patient survival times when only a historical survival curve is accessible, making it suitable for situations where the two-sample log-rank test cannot be used as the method of choice due to the unavailability of complete historical survival time data.

References

- [1] Feld, J.* , Danzer, M. F.* , Faldum, A., Schmidt, R. (2022). *authors contributed equally Reference curve sampling variability in one-sample log-rank tests. PLOS ONE doi: 10.1371/journal.pone.0271094

2024-03-01 09:20 - 09:40, AM S1

SES-15: Time-to-Event 1

Conditional survival of younger patients with mantle cell lymphoma: novel insights into disease course and dynamic prediction by baseline and time-dependent prognostic factors

Jiang, Linmiao¹; Dreyling, Martin²; Hermine, Olivier³; Schumacher, Martin⁴; Hoster, Eva¹

¹Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, LMU Munich, Munich, Germany

²Department of Internal Medicine III, LMU University Hospital Munich, Munich, Germany

³Department Hematology, Hôpital Necker, Assistance Publique Hôpitaux de Paris, University Paris Descartes, Paris, France

⁴Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Background: Conditional survival (CS) is the future survival probability after a patient has already survived certain time. Estimating CS extends standard survival analyses by providing dynamic predictions over the treatment and disease course and more relevant information for patients who are still event-free after a few years. Our study aims to investigate conditional overall survival in younger patients with mantle cell lymphoma (MCL) with a focus on the dynamic outcome evolution and impact of baseline and time-dependent prognostic factors.

Methods: We used a mature dataset from a randomized phase III trial for untreated advanced-stage younger MCL patients with median follow-up of 10.6 years. The trial compared two different combinations of immuno-chemotherapies, R-CHOP arm and R-DHAP arm. All randomized patients with confirmed MCL diagnosis were included. Landmark Kaplan-Meier method was used to estimate 3- and 5-year CS probability for prediction times $s = 0, 1, 2, \dots, 8$ years. Landmark Cox regression models were applied for subgroups with different baseline Mantle Cell Lymphoma International Prognostic Index (MIPI) groups and dynamic treatment failure (TF) status by prediction times.

Results: Among 497 randomized patients, 473 MCL patients (237 in R-CHOP and 236 in R-DHAP arm) were included in the analysis. Kaplan-Meier estimates of both 3- and 5-year CS probabilities increased with prediction time in R-CHOP arm, but remained almost constant in R-DHAP arm. The 5-y CS differences between low-risk and intermediate/high-risk MIPI groups decreased with prediction time in R-CHOP arm until no differences could be observed after 3 years, while the differences stayed constant at all prediction times in R-DHAP arm. The 3-/5-year CS between patients stratified by

TF showed clear differences at all prediction times in both arms. Patients with TF within the first 3 years in R-DHAP arm had particularly short 3-/5-year CS. The combination of MIPI and TF revealed that patients with TF had inferior CS, independently of MIPI risk groups.

Conclusion: Different CS trends in two arms indicated differential disease courses and dynamic effects of baseline prognostic factors depending on the initial treatment groups, which were not visible from traditional survival analyses. TF, as a time-varying prognostic factor, showed different effects and dynamic evolutions in two arms, suggesting a need for updated risk analysis over time. Clinical interpretations of the observed CS differences under different treatment need to be further investigated. CS analyses could also provide additional insights when applied to other diseases in haematology, oncology and beyond.

2024-03-01 09:40 - 10:00, AM S1

SES-15: Time-to-Event 1

The challenge of time-to-event analysis for multiple events: Which method of analysis can we trust?

Schmeller, Sandra¹; Erdmann, Alexandra¹; Beyersmann, Jan¹; Ozga, Ann-Kathrin²

¹Institute of Statistics, Ulm University, Germany

²Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Germany

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 any other events. 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. 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. Pros and cons of the different methods will be introduced and explained on data from the Interdisciplinary Network for Heart Failure (1). This multi-centre (9 centers) randomized controlled trial investigated the efficacy of a nurse-coordinated disease management program (HNC) in heart failure compared to usual care for patients that were first hospitalized for systolic heart failure. A total of 1022 patients (513 in usual care, 509 in HNC group) with 663 deaths, 3016 re-hospitalizations (with a maximum of 27 events per patient) and a median follow up of 2596 days were observed.

Special attention is given on the prerequisites of the models, e.g. the Markov property, and their (causal) interpretation. A test for the Markov property is investigated within a simulation study and intensity versus rate modeling will be discussed. We explore the relation between summary measures like the mean number of hospitalizations which can be estimated also in non Markov models and the harder interpretable effects on the intensity or rate scale (2). The aim is to give an overview of existing methods, present the assumptions, with a special focus on the Markov assumption, and elaborate the difference in interpretation.

References

- (1) Christiane E Angermann et al. Mode of action and effects of standardized collaborative disease management on mortality and morbidity in patients with systolic heart failure: the interdisciplinary network for heart failure (inh) study. *Circulation: Heart Failure*, 5(1):25–35, 2012.
- (2) Alexandra Nießl et al. Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*, 25: 110-124, 2023.

2024-03-01 10:00 - 10:20, AM S1

SES-15: Time-to-Event 1

Unleashing the power of adjusted survival curves: Introducing the adjustedCurves R-package

Denz, Robin; Timmesfeld, Nina

Ruhr-University Bochum, Germany

Kaplan-Meier curves stratified by treatment allocation are the most popular way to depict causal effects in studies with right-censored time-to-event endpoints. If the treatment is randomly assigned and the sample size of the study is adequate, this method produces unbiased estimates of the population-averaged counterfactual survival curves. However, in the presence of confounding, this is no longer the case. Instead, specific methods that allow adjustment for confounding must be used. Although many such methods have been proposed in the literature, they are underused in practice. We suggest that one of the main reasons for this is the lack of user-friendly software implementations of these methods in widely used languages such as R and SAS. We present the adjustedCurves R-package, which aims to fill this gap by uniting the sometimes vastly different methods under one consistent framework. It includes 14 different methods to estimate confounder-adjusted survival curves, including different stratification-based estimators, inverse probability of treatment weighting, propensity score matching, g-computation, augmented inverse probability of treatment weighting, empirical likelihood estimation, an instrumental variable-based method, targeted maximum likelihood estimation, and some of their pseudo-values-based counterparts. Estimation of standard errors and confidence intervals is naturally supported for all methods via built-in non-parametric bootstrapping and approximate formulas (where available). The resulting estimates can then be plotted using the package's highly customisable plotting functions. In addition, they can be used directly to compare treatment groups by calculating differences in survival probabilities at a given point in time, differences in restricted mean survival times or by contrasting adjusted survival time quantiles. The use of these quantities for treatment comparisons has been recommended in the literature, because they are absolute measures of effect and are easy to interpret, unlike the hazard-ratio of a Cox proportional hazards model. We provide a brief overview of the included methods and demonstrate the use of the package with publicly available data from an observational study of 2982 breast cancer patients.

References

Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust survival curves for confounders. Stat Med. 2023; 42(10): 1461-1479

Denz R. adjustedCurves: confounder-adjusted survival curves and cumulative incidence functions. R package version 0.10.1; 2023.
<https://cran.r-project.org/package=adjustedCurves>.

Ozenne BMH, Scheike TH, Stærk L. On the estimation of average treatment effects with right-censored time to event outcome and competing risks. Biom J. 2020; 62(3): 751-763.

Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. Am J Epidemiol. 1996; 143(10): 1059-1068.

2024-03-01 10:40 - 11:00, AM S1

SES-16: Time-to-Event 2

Adaptive redesigning of combination testing procedures in survival analysis

Danzer, Moritz Fabian¹; Dormuth, Ina²

¹University of Münster, Germany

²TU Dortmund, Germany

In survival analysis, the assumption of proportional hazards is very common. In many practical scenarios, however, this assumption must be questioned. The possible deviations from this assumption are numerous. In such cases, the standard log-rank test loses its optimality and a differently weighted log-rank type test would be preferable. As the exact shape of the deviation is unknown, the best choice for the weight remains unclear, in particular during the planning stage of a trial.

Recently, testing procedures have been developed that combine the information from differently weighted log-rank type tests. Such procedures have a broader power function than log-rank type tests with a single weight as a large variety of deviations from the null hypothesis can be detected. Of course, this has the disadvantage that the procedure has less power than the test, whose weight would be optimal in the respective situation.

To narrow this gap between the optimal, but unknown test and the combined approach mentioned here, we propose an adaptive design. A start of the study with a broadly based combination approach can thus be combined with more refined procedures in later stages. For this, we want to use the information collected up to an interim analysis to redefine the testing procedure in terms of the weights of the log-rank tests. At the same time, other commonly used tools from adaptive designs (e.g. sample size recalculations) shall be applicable.

2024-03-01 11:00 - 11:20, AM S1

SES-16: Time-to-Event 2

A new parametric accelerated failure time model for semi-competing risk data

Dineva, Antoniya¹; Kuss, Oliver²; Hoyer, Annika¹

¹Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University

²Institute for Biometrics and Epidemiology, German Diabetes Center, Düsseldorf

Semi-competing risks refer to the general setting where interest focuses on the age at a non-terminal event (e.g., disease occurrence) when the investigated subjects are also under risk for experiencing a terminal event (e.g., death). That is, the terminal event might censor the non-terminal event, but remains observable if it occurs first (1).

The underlying setting can be described by an illness-death-model incorporating transitions between the three states “healthy”, “diseased” and “dead”. In recent publications, fitting a Cox frailty model for each transition is suggested (2,3,4). An appealing alternative, that has not gained a lot of attention yet is an accelerated failure time (AFT) model instead.

This work aims to fill this gap by fitting an AFT model for each of the three transitions. The major advantage of this approach is its intuitive and straightforward interpretation based on the survival instead of the hazard function, and facilitating communication of results. We propose a parametric model by assuming Weibull distributions for the ages in the different states, modelling 1) the age at disease onset, 2) the age at death for disease-free subjects, and 3) the age at death after disease onset. To adjust for intra-individual correlations, we add random effects to the linear predictor, yielding finally a trivariate random effects model.

Model parameters are estimated by the maximum likelihood principle. The likelihood function incorporates left truncation for both terminal and non-terminal events, reflecting the delayed entry into the study cohort. As in cohort studies the diagnosis usually is made at intermittent follow-up visits and the exact age at disease onset is only known to lie in the interval between the last two visits, we finally consider interval censoring for the disease occurrence.

The model is illustrated using data from the Paquid study (5), a large cohort study on mental and physical aging. The event of interest for our analysis is the occurrence of dementia in the study population. Our model leads to plausible results, indicating that people with dementia diagnosis die earlier compared with people without dementia.

Abstracts of Contributed Talks

References

- (1) Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001 Dec 1;88(4):907-19.
- (2) Gorfine M, Keret N, Ben Arie A, Zucker D, Hsu L. Marginalized frailty-based illness-death model: application to the UK-Biobank survival data. *Journal of the American Statistical Association*. 2021 Jul 3;116(535):1155-67.
- (3) Lee C, Gilsanz P, Haneuse S. Fitting a shared frailty illness-death model to left-truncated semi-competing risks data to examine the impact of education level on incident dementia. *BMC Medical Research Methodology*. 2021 Dec;21(1):1-3.
- (4) Xu J, Kalbfleisch JD, Tai B. Statistical analysis of illness-death processes and semi-competing risks data. *Biometrics*. 2010 Sep;66(3):716-25.
- (5) Letenneur L, Gilleron V, Commenges D, Helmer C, Orgogozo JM, Dartigues JF. Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project. *Journal of Neurology, Neurosurgery & Psychiatry*. 1999 Feb 1;66(2):177-83.

2024-03-01 11:20 - 11:40, AM S1

SES-16: Time-to-Event 2

Dynamic prediction of the risk of preeclampsia – Landmarking with continuous time-dependent covariates in left-truncated competing risks data

Stegherr, Regina¹; Aigner, Annette¹; Verloren, Stefan²

¹Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Biometry and Clinical Epidemiology, Charitéplatz 1, 10117 Berlin, Germany

²Department of Obstetrics, Charité – Universitätsmedizin Berlin, Campus Charité-Mitte, Charitéplatz 1, 10117 Berlin, Germany

Dynamic prediction based on time-to-event data with continuous time-dependent covariates has received much attention over the last years. Two approaches are frequently suggested: joint modelling and landmarking. However, the application to left-truncated competing risk data with time-dependent covariates is still rarely considered. At the same time, prediction models for preeclampsia have become increasingly complex in the last few years. The previously frequently used logistic regression is replaced by machine learning and Bayesian approaches. Even the importance of dealing with competing events was emphasized in a Bayesian approach [1]. However, complex time-to-event methods are rarely considered so far.

We use retrospectively collected data of about 2990 recorded pregnancies, of which 557 ended with a delivery with preeclampsia. Maternal characteristics and several irregularly and possibly repeatedly measured time-dependent biomarkers, such as the placental growth factor (PIGF) or mean arterial pressure (MAP), are used to predict the risk of preeclampsia. As there is no common study entry time, as pregnant women are only observed at a later stage of pregnancy, all pregnancies are left-truncated. The primary endpoint is delivery with preeclampsia; however, the competing event delivery without preeclampsia must also be considered.

We consider three possible approaches to develop a dynamic prediction model for the risk of preeclampsia. First, a two-stage approach predicting the unmeasured biomarker values using mixed effects models and incorporating them into Cox proportional hazards models for the cause-specific hazards, which are then used for prediction, is considered. Second, landmarking 2.0 [2] is extended to a left-truncated competing risk setting. Similar to the two-stage approach, in the first step, biomarkers are longitudinally modelled, and the biomarker value at the landmark time point is predicted based on Gaussian processes or revival. In the second step, time-dependent Cox proportional hazard models for each cause-specific hazard based on the predicted biomarker values are fitted. Lastly, dynamic

Abstracts of Contributed Talks

predictions for new patients are obtained. Third, predictions based on joint models are considered, which simultaneously model the longitudinal biomarker process and the time-to-event process. All three approaches are applied to the pregnant women data and in a simulation study, and we discuss which approach is the most appropriate.

References

- [1] Wright, D., Wright, A., & Nicolaides, K. H. (2020). The competing risk approach for prediction of preeclampsia. *American journal of obstetrics and gynecology*, 223(1), 12-23.
- [2] Putter, H., & van Houwelingen, H. C. (2022). Landmarking 2.0: Bridging the gap between joint models and landmarking. *Statistics in Medicine*, 41(11), 1901-1917.

2024-03-01 13:00 - 13:20, AM S1

SES-17: Time-to-Event: Estimation

Implication of the choice of time scales in survival analysis

**Vilsmeier, Judith¹; Büchele, Gisela²; Rehm, Martin²; Rothenbacher, Dietrich²;
Beyermann, Jan¹**

¹Institute of Statistics, Ulm University, Germany

²Institute of Epidemiology and Medical Biometry, Ulm University, Germany

The choice of time scale is an important and often discussed topic in time-to-event analysis. While the rule of thumb is to choose the natural time scale for the underlying problem, it is often not clear what this natural time scale is. This potentially leads to confusion and if researchers pick the "wrong" time scale for the underlying problem, this can lead to biased results. Two often discussed time scales are time since study entry and time since birth, i.e. age, but there are also other possible time scales. The time scale we focus on during this talk is the so-called calendar time. Its origin is an arbitrary day before the study entry of the first patients, and it is useful if calendar time holds important information in addition to just the time from study entry. However, we will illustrate situations where it can lead to an inflated estimation of the hazard ratio and overoptimistic p-values when using the Cox proportional hazard model. In this talk a real data example from the EvaCoM project is used to demonstrate this hazard ratio inflation and the reasons why it occurs. In this project, the impact of quality audit of healthcare providers was of interest, possible time scales being calendar time in which the audit acts or a patient's time-to-event. It serves as an example to raise awareness to why researchers should be careful when deciding which time scale to use, by highlighting a situation in which the choice of time scale might not be clear but impactful. Additionally, possible solutions for scenarios where more than one time scale seem to be important are proposed.

2024-03-01 13:20 - 13:40, AM S1

SES-17: Time-to-Event: Estimation

Hazards: key quantities for analysis, interpretation and understanding of time-to-event data

Beyersmann, Jan¹; Schmoor, Claudia²; Schumacher, Martin²

¹Ulm University, Germany

²Albert-Ludwigs-Universität Freiburg, Germany

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. 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. 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.

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.

2024-03-01 13:40 - 14:00, AM S1

SES-17: Time-to-Event: Estimation

Piecewise constant hazard estimation with the fused lasso

Rosenbaum, Manuel; Beyersmann, Jan; Vogt, Michael

Universität Ulm, Deutschland

In applied time-to-event analysis, assuming piecewise constant hazards is a flexible and useful parametric model. However, the change points and values of the piecewise constant hazard are usually unknown in practice and need to be estimated. In our project, we develop a fully data-driven procedure for piecewise constant hazard estimation. We work in the general framework of multiplicative intensity models, which nests a wide range of popular models including Cox's proportional hazards model. To construct our estimator, we set up a regression model for the increments of the Breslow estimator and then use fused lasso techniques to approximate the piecewise constant signal in this regression model. In the theoretical part of the project, we derive the convergence rate of our estimator as well as some results on how well the change points of the piecewise constant hazard are approximated by our method. We complement the theory by both simulation and a real data example, also illustrating that our results apply in rather general event histories such as multi-state models.

2024-03-01 14:00 - 14:20, AM S1

SES-17: Time-to-Event: Estimation

Estimation Within The Responder Stratified Exponential Survival Model

Kilian, Samuel; Kieser, Meinhard

Institute of Medical Biometry, Heidelberg University, Germany

The primary endpoint in oncology is usually overall survival, where differences between therapies may only be observable after many years. To avoid withholding of a promising therapy, preliminary approval based on a surrogate endpoint is possible. The approval can be confirmed later by assessing overall survival. When planning and analysing trials in this context, the correlation between surrogate endpoint and overall survival has to be taken into account. For the binary surrogate endpoint response, this relation can be modeled by means of the responder stratified exponential survival (RSES) model proposed by Xia et al. (1). The RSES model assumes that response is Bernoulli-distributed and survival conditional on response is exponentially distributed. This results in three model parameters: response probability, responder survival, and non-responder survival.

In this talk, we present methods for estimating the three model parameters of the RSES model. We derive estimators by the Maximum Likelihood method for different model parameterizations. We identify the parameterization with the best approximate normality by graphical evaluation. We investigate the correlation between the three estimators and find it to be very small. We construct approximate confidence intervals for the model parameters and find them to have very satisfying coverage probability. We outline how the exact distribution of the estimators can be derived. Furthermore, we outline how hypothesis tests for parameter differences and corresponding sample size calculation methods can be derived.

We conclude that the presented estimators have very good properties and can be used for planning and analysing studies in the context of a binary surrogate endpoint for survival. Also, the presented approach is applicable to other parametric survival models.

References

- (1) Xia, Y., Cui, L., & Yang, B. (2014). A note on breast cancer trials with pCR-based accelerated approval. *J Biopharm Stat.*, 24(5), 1102-14. doi:10.1080/10543406.2014.931410

2024-03-01 14:40 - 15:00, AM S1

SES-18: Time-to-Event: Machine Learning

A Large-Scale Neutral Comparison Study of Survival Models

Burk, Lukas^{1,2,3,4}; Bischl, Bernd^{2,4}; Bender, Andreas^{2,4}; Wright, Marvin^{1,3}; Lang, Michel^{5,6}; Sonabend, Raphael^{7,8}

¹Leibniz Institute for Prevention Research and Epidemiology - BIPS

²LMU Munich

³University of Bremen

⁴Munich Center for Machine Learning (MCML)

⁵TU Dortmund

⁶Research Center Trustworthy Data Science and Security

⁷OSPO Now

⁸Imperial College London

This work represents what we believe to be the first neutral and large-scale benchmark experiment for single-event, right-censored and low-dimensional survival data.

Benchmark experiments are crucial parts of methodological research to robustly compare new model classes in empirical experiments.

New methods are often benchmarked by the authors against a selection of existing methods, but these comparisons can not be considered neutral due to the authors' relationship to their own method.

Neutral benchmark experiments allow us to ensure that conclusions can be generalised to datasets that are representative of those included in the experiment.

In this context, we are able to generalise results to low-dimensional right-censored survival datasets, which are by far the most common datasets for time-to-event analysis available off-shelf (e.g., in R and Python libraries).

To draw these conclusions we benchmark 32 datasets, 17 models, and 7 measures, the largest benchmark experiment in survival analysis to-date.

While there are many benchmarking experiments available in the survival literature, they typically are either limited in scope (often focused on high-dimensional data) or are qualitative in nature where methods are reviewed as opposed to analytical quantitative comparisons.

Abstracts of Contributed Talks

So far, there has not been a comprehensive study to neutrally evaluate the performance of a sufficiently wide range of methods in low-dimensional settings.

We compare 17 models from classical statistical models (non-parametric estimators, proportional hazard models, etc.) to specialized machine learning methods such as boosting and random forests.

Model classes are compared on 32 low-dimensional survival datasets available publicly in software repositories - we have uploaded our full benchmarking suite to OpenML to ensure the experiment is completely reproducible.

Model evaluation is conducted using 7 different metrics, assessing calibration, discrimination, and overall predictive accuracy with scoring rules.

Machine learning models are tuned separately across these three metric classes to account for the effect of the tuning procedure on model performance.

2024-03-01 15:00 - 15:20, AM S1

SES-18: Time-to-Event: Machine Learning

Integrating Fine & Gray's Subdistribution Weights into Random Survival Forests for Competing Event Analysis

Behning, Charlotte¹; Bigerl, Alexander²; Wright, Marvin³; Berger, Moritz¹; Schmid, Matthias¹

¹Institute of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn

²DICE Group, Department of Computer Science, Paderborn University, Paderborn, Germany

³Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

Random Survival Forests (RSF) can be used to address medical research questions in survival analysis and time-to-event outcomes in other research areas. RSF can be advantageous over standard methods, especially in complex scenarios involving, for example, high dimensional data settings and if nonlinear relationships or unknown interactions between independent variables are present. A further level of complexity may occur in clinical contexts when not only the primary event of interest is observed, but when patients also may experience competing events other than the event of interest. Neglecting these competing events, e.g. by simply treating them as censoring, can lead to biased estimates of the cumulative incidence function (CIF). To address this challenge, Fine & Gray introduced the subdistribution hazard model that directly estimates the CIF within a framework defined on a subdistribution time scale.

In this work, we integrate concepts from Fine & Gray's subdistribution hazard modelling approach into RSF. For this, we utilize a property of random forests, namely the creation of multiple decision trees, each trained on a random sample of the data. In each node of the trees, the competing event times are imputed by potentially right-censored subdistribution times. We then apply the split rules designed for single-event RSF. Predictions from the individual trees are combined to obtain a final prediction. The performance of the proposed method is illustrated in a simulation study. In the simulation, the CIF is compared with an RSF that treats the competing risks as censoring events under various rates of the event of interest.

2024-03-01 15:20 - 15:40, AM S1

SES-18: Time-to-Event: Machine Learning

A random forest pseudo-value approach for modeling restricted mean survival times

Schenk, Alina¹; Basten, Vanessa^{1,2}; Schmid, Matthias¹

¹Department of Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn

²University of Applied Sciences Koblenz, Rhein-Ahr-Campus Remagen

Because of its simple interpretation, the restricted mean survival time RMST is often suggested as measure for treatment effects in time-to-event analysis in randomized controlled trials. The RMST is defined as the area under the survival function $S(t)$ up to $t > 0$ and can readily be interpreted as the life expectancy between $t = 0$ and a specific time horizon ($t = \tau$). Besides, the usage of RMST is advantageous in the sense of having minimal assumptions on the survival process (such as, e.g. proportional hazards of a treatment effect) [1].

In practice, the direct modeling of the RMST conditional on a set of covariates X is of particular interest when investigating covariate effects (e.g. treatment effects) on the expected life time. However, due to (right-censored) data leading to partly unobserved survival times, modeling of the RMST is not straightforward and requires special estimation and modeling techniques. Here, we use individual-specific leave-one-out jackknife pseudo-values providing, on average, a distribution-free estimate of the restricted mean survival time at τ for the entire sample [1]. These values can be treated in the same way as a continuous outcome in standard regression models for estimating the RMST [2].

Currently, the most popular strategy for modeling pseudo-values is the estimation of main covariate effects on the RMST via a generalized estimation equation (GEE) approach. While the GEE method yields consistent estimates in the case of correct model specification and random censoring, the commonly applied model is restricted to the inclusion of main covariate effects only. The inclusion of more flexible effect terms (e.g. interactions between the covariates) by pre-specification is often infeasible, as it would require prior knowledge on the, usually hidden, interaction structure in the data.

To extend standard pseudo-value models by higher-order interaction terms, we propose an alternative modeling approach with the aim of the estimation of individual-specific RMST, while at the same time selecting most relevant covariates in a data-driven way. Conceptually, our modeling approach comprises the estimation of pseudo-values for RMST used for the specification of a flexible random forest regression algorithm. We will present a simulation study and an application investigating the ability of our method to filter

out relevant covariates and to accurately estimate RMST as well as a comparison to established methods for RMST estimation and modeling.

References

- [1] P. Royston and M. K. B. Parmar. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30:2409–2421, 2011.
- [2] P. K. Andersen and M. Pohar Perme. Pseudo-observations in survival analysis. *Statistical Methods in Medical Research*, 19:71–99, 2010.

2024-03-01 15:40 - 16:00, AM S1

SES-18: Time-to-Event: Machine Learning

Oversimplifying machine learning on event time data – results of a literature and software review

Klein, Lukas; Grieser, Gunter; Jahn, Antje

Hochschule Darmstadt, University of Applied Sciences, Germany

The German organ transplant registry (TxReg) has recently become available [1], offering a unique opportunity to study the post-transplant survival of organ recipients in the German population and healthcare system. In recent years an increasing interest in applying machine-learning (ML) methods for predicting post-transplant survival was observed [2]. However, the issue of censoring in survival analysis prediction tasks seems to be often neglected in machine learning applications. Instead, the task is reduced to classification, a simplification that is only rarely seen in regression modeling. In the present work, we will investigate how ML is applied for event time endpoints in organ transplantation research, how censoring was handled and how both depend on the choice of R or Python as the software ecosystem.

First, we perform a literature review on how ML is applied for survival predictions from organ transplant registry data with respect to the choice of software, handling of censored observations, variable selection, reporting of prediction accuracy and uncertainty, and prediction communication. Second, we compare the R and Python libraries for ML for event time data with respect to their features and their computational speed when applied to large data. This will be based on an extensive simulation study in accordance with the ADEMP principles [3] and will be accompanied by an application to the German organ transplant registry data.

We show in our literature review that many studies using machine learning on organ transplantation event time data ignore censored observations, and this is seen more often when Python has been used. Although, commonly used Python implementations allow for IPC-weighting [4] to address censoring, this is only rarely applied.

When comparing ML algorithms for survival prediction, the Python libraries provide excellent possibilities for the integration of deep learning methods and machine learning pipelines for censored data. However, R provides better tools for model inspection, for assessing prediction performance for example by calibration curves and for scoring frameworks like TRIPOD [5].

Overall, our talk highlights that ML for survival prediction is very differently applied and

reported depending on whether R or Python is used. Whereas Python shows a strength in deep learning approaches, the risk of biased results is much higher as available ML libraries for event time data are not commonly used in Python.

References

1. Roland Nagel. Über das Transplantationsregister.
2. Neta Gotlieb, Amirhossein Azhie, Divya Sharma, Ashley Spann, Nan-Ji Suo, Jason Tran, Ani Orchanian-Cheff, Bo Wang, Anna Goldenberg, Michael Chassé, Heloise Cardinal, Joseph Paul Cohen, Andrea Lodi, Melanie Dieude, Mamatha Bhat. "The promise of machine learning applications in solid organ transplantation". *npj Digital Medicine* 2022; 5:89.
3. Tim P Morris, Ian R White, Michael J Crowther. "Using simulation studies to evaluate statistical methods.". *Statistics in medicine* 2019; 38:2074-2102.
4. David M Vock, Julian Wolfson, Sunayan Bandyopadhyay, Gediminas Adomavicius, Paul E Johnson, Gabriela Vazquez-Benitez, Patrick J O'Connor. "Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting.". *Journal of biomedical informatics* 2016; 61:119-31.
5. Maria C Haller, Constantin Aschauer, Christine Wallisch, Karen Leffondré, Maarten van Smeden, Rainer Oberbauer, Georg Heinze. "Prediction models for living organ transplantation are poorly developed reported and validated a systematic review". *Journal of Clinical Epidemiology* 2022; 145:126-135.

2024-02-28 14:40 - 15:00, AM S2

SES-19: Preclinical Drug Development and Toxicology 1

Interaction effects of UVA with UVB irradiation at the gene expression level in human skin cells

Talleb, Yassine¹; Rolfes, Katharina²; Dobner, Jochen²; Rossi, Andrea²; Haarmann-Stemmann, Thomas²; Krutmann, Jean²; Ickstadt, Katja¹

¹TU Dortmund, Germany

²IUF Düsseldorf, Germany

Ultraviolet (UV) B radiation (290-315 nm) and UVA (315 – 400 nm) are complete carcinogens and both are well known to contribute to the development of skin cancer in humans. Under physiological conditions, human skin is exposed to a mixture of UVB and UVA radiation from natural sunlight. Previous research, however, has primarily focused only on the effects of each type of radiation separately whereas the knowledge of the UVA-UVB-interaction under simultaneous exposure and its impact on human skin is quite unclear. In a previous photocarcinogenesis study we have found that simultaneous exposure of murine skin to UVB and non-carcinogenic doses of UVA increased UVB-induced photocarcinogenesis. These results indicate that detrimental effects caused by UVB and UVA radiation, if applied simultaneously, can enhance each other, even when the UVA dose per se does not cause significant skin damage. In the present study we would like to further analyze the interaction of UVB and UVA radiation, if applied simultaneously, by analyzing gene expression responses (by bulk RNAseq) in human skin cells. We are particularly interested in identifying the minimum UVA dose which is required to enhance UVB-induced skin damage.

The dose-response relationships involved will be examined more closely by using already existing RNAseq data sets of skin cells. Keratinocytes were irradiated either with only UVB or UVA, or with a combination (simultaneous irradiation) of both. The RNA of the cells was isolated at two different time points after irradiation (incubation time). Additional data was acquired for further incubation times and different physiologically relevant dosages and different ratios of UVA and UVB radiation. These settings were chosen to best support the statistical analysis.

When modelling univariate dose-response relationships separately for each incubation time, the information shared across the incubation times is not considered. However, it is critical to identify certain features of the relationship. We will consider data of all incubation times at once by modelling a two-dimensional surface with the help of tensor product B-splines. The identification of the minimum UVA dose, at which the combination irradiation shows a significantly different effect compared to UVB irradiation alone,

will ultimately also be investigated and described with the help of Gaussian processes and spline regression models.

2024-02-28 15:00 - 15:20, AM S2

SES-19: Preclinical Drug Development and Toxicology 1

Design of optimal concentrations for in vitro cytotoxicity experiments

Schürmeyer, Leonie¹; Schorning, Kirsten¹; Hengstler, Jan Georg²

¹TU Dortmund University, Germany

²IfADo Leibniz Research centre for working environment and human factors, Germany

Concentration dependent cytotoxicity tests are frequently used in toxicology. These tests usually serve to determine the EC50-value as the concentration of a substance that reduces vitality to 50% of solvent controls. An important challenge of in vitro cytotoxicity testing is the choice of adequate test concentrations, to ensure a precise estimation of the concentration-response relationship. A recent literature review of three major toxicological journals [1] showed that optimal design approaches are not considered at all for planning concentration-response experiments in toxicology, although plenty of research is available in this field ([2], [3], [4] among many others).

Therefore, the performance of three different design approaches (Bayesian, log-equidistant and intuition-based) in cytotoxicity concentration-response experiments is investigated in a user-friendly way. Here, two scenarios should be differentiated, on the one hand, where detailed previous knowledge on the substance is available and on the other hand, where this is not the case, for example because new substances are tested.

Investigating the scenario where detailed previous knowledge is available, first Bayesian D-optimal designs are determined based on three previously known EC50-values of valproic acid (VPA). Then the performance of the three different design approaches is analyzed based on an extensive cytotoxicity experiment with 50 concentrations of VPA (which include the concentrations of the different design approaches). A reference curve that serves as an approximation of the true concentration-response relationship of VPA is calculated based on the resulting data set. Then bootstrap analyses are performed in which distances of the curves obtained in the different bootstrap scenarios to the reference curve are determined: Here, both the Root mean squared error values and the distance between the EC50-values are considered. Our results actively demonstrate using Bayesian D-optimal designs leads to a higher model and EC50 precision than frequently used log-equidistant designs [1] or the intuition-based design.

For the second scenario, where there is no prior knowledge of the substance, data of 104 hepatotoxic substances are analyzed and a Bayesian D-optimal design is determined based on the distribution of the corresponding EC50-values. It is shown by an extensive simulation study that the resulting design outperforms the originally used designs for

almost all 104 substances.

The results and findings are presented in a user-friendly way to close the gap between methodological research and practice a bit further. In addition to this, we propose a recommendation, how to plan upcoming cytotoxicity experiments where more or less prior information is available.

References

- [1] Kappenberg, F., Duda, J. C., Schürmeyer, L., Gül, O., Brecklinghaus, T., Hengstler, J. G., Schorning, K. and Rahnenführer, J. (2023). Guidance for statistical design and analysis of toxicological dose–response experiments, based on a comprehensive literature review. *Archives of Toxicology*, 1-21.
- [2] Dette, H., Bretz, F., Pepelyshev, A., & Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, 103(483), 1225-1237.
- [3] Dette, H., Pepelyshev, A., Shpilev, P., and Wong, W. K. (2010). Optimal designs for discriminating between dose-response models in toxicology studies. *Bernoulli*, 1164-1176.
- [4] Holland-Letz, T., and Kopp-Schneider, A. (2015). Optimal experimental designs for dose–response studies with continuous endpoints. *Archives of toxicology*, 89, 2059-2068.

2024-02-28 15:20 - 15:40, AM S2

SES-19: Preclinical Drug Development and Toxicology 1

Probabilistic Approaches for Modeling Patient-Specific Effects of Antihypertensive Medication

Hunsdieck, Berit^{1,2}; Elci, Eren¹; Ickstadt, Katja²

¹Technische Universität Dortmund, Germany

²Bayer AG, Germany

The benefit of 24/7 blood pressure measurements during the titration phase of antihypertensive need is investigated. As more and more devices can measure blood pressure 24/7, the aim is to predict the drug effect during dose titration, i.e., to predict the saturation level of the drug after a few doses only. As no continuous blood pressure measurement data in relevant populations are currently available, an essential part of the work is to simulate realistic blood pressure data.

The primary focus of simulating real-world data is centered on the endeavor to create a detailed and realistic representation of hypertensive patients. In this pursuit, particular emphasis is placed on harnessing the valuable resource provided by the UK Biobank (see <https://www.ukbiobank.ac.uk/>).

To attain a comprehensive and realistic portrayal of hypertensive patients within the simulation, an extensive dataset from the UK Biobank is utilized. The process involves the integration of various facets of patient information like sex, age and individual diagnoses and their influence on the individual blood pressure. The individual daily rhythm is considered as well as the effect of medication with the help of PKPD modeling.

Several machine learning approaches are explored for the prediction. Parametric approaches in the form of a linear mixed effect model as well as a non-linear mixed effect (nlme) model and a non-parametric approach in the form of Gaussian processes are investigated. Model performance was evaluated and compared regarding the accuracy, bias, and precision. The focus is on two questions:

1. What time period is necessary for a solid prediction of the antihypertensive medication effect?
2. Which model performs best under which circumstances?

In scenarios where only a limited number of days of measurements are available, the Gaussian process is the preferred choice. However, when selecting the Gaussian process,

careful consideration should be given to whether a conservative estimation or a narrower prediction interval, which may not necessarily encompass the true values, is desired. When multiple days of data (more than 3) are available, the nlme model becomes the preferred choice.

References

Lasserson DS, Buclin T, Glasziou PHow quickly should we titrate antihypertensive medication? Systematic review modelling blood pressure response from trial data. Heart 2011;97:1771-1775

Stephen J. Phillips, W. Michael O'Fallon, and Jack P. Whisnant. A population-based model for predicting blood pressure. Mayo Clinic Proceedings, 63(7):700–706, 1988. doi: 10.1016/s0025-6196(12)65532-3

C. Van Rijn-Bikker, N. Snelder, O. Ackaert, R. M. Van Hest, B. A. Ploeger, G. A. Van Montfrans, R. P. Koopmans, and R. A. Mathot. Nonlinear mixed effects modeling of the diurnal blood pressure profile in a multiracial population. American Journal of Hypertension, 26(9):1103–1113, 2013. doi: 10.1093/ajh/hpt088

2024-02-28 16:20 - 16:40, AM S2

SES-20: Preclinical Drug Development and Toxicology 2

Shift from Frequentist to Bayesian Dose-Response modelling

Duda, Julia Christin¹; Wheeler, Matthew²

¹TU Dortmund University, Germany

²National Institute of Environmental Health Sciences, USA

Dose-response modelling is crucial in pharmaceutical development to understand the dose dependent effects of a compound. This includes both desirable therapeutic effects as well as unwanted toxic effects. For risk assessment, the European Food Safety Authority (EFSA) recently performed a paradigm shift from frequentist benchmark dose estimation to Bayesian approaches. The use of Bayesian approaches within the regulatory context poses various challenges. It also bears opportunities, as prior knowledge can be fueled into mathematical approaches and increase the reliability of the approach, especially in the case of few data.

The EFSA paradigm shift was stated in 2022 in the scientific committee's guideline. The previously recommended model averaging to calculate the benchmark dose (BMD) should now also be performed within a Bayesian framework. Due to the general applicability of the BMD approach for various, more specific test guidelines, the paradigm shift affects many study types simultaneously.

Model averaging is recommended to safeguard against model misspecification and consequential estimation errors. Among the considered models, the Hill model is well-known and always included as a model choice. Originally invented to describe the equilibrium relationship between oxygen tension and the saturation of hemoglobin, its sigmoidal shape is applicable in various dose-response estimation scenarios.

Within the context of new Bayesian methods becoming more prominent in pharmaceutical development, specifically risk assessment, we present a new Bayesian dose-response modeling approach. Based upon the work of Shin et al. (2020) we propose a non-parametric non-linear functional shrinkage approach that uses a priori beliefs in a parametric model while automatically accounting for possible deviations from the model. For our application, dose-response modelling, this is the space of Hill functions. While Shin et al. proposed this method for linear subspaces, we extended the approach for function spaces defined by non-linear functions, such as the Hill model. We compare our approach against other Bayesian approaches in a simulation study. We further apply the proposed method on a testosterone-level data example. At last, we discuss limitations, as well as possible extensions to increase suitability of the approach for risk assessment.

References

EFSA Scientific Committee (2022). Guidance on the use of the benchmark dose approach in risk assessment. *Efsa Journal*, 20(10), e07584.

Shin, M., Bhattacharya, A., & Johnson, V. E. (2020). Functional horseshoe priors for subspace shrinkage. *Journal of the American Statistical Association*, 115(532), 1784-1797.

2024-02-28 16:40 - 17:00, AM S2

SES-20: Preclinical Drug Development and Toxicology 2

Testing for similarity of multivariate mixed outcomes with application to efficacy-toxicity responses

Hagemann, Niklas^{1,2}; Marra, Giampiero³; Bretz, Frank⁴; Möllenhoff, Kathrin^{1,2}

¹Institute of Medical Statistics and Computational Biology, University of Cologne, Germany

²Mathematical Institute, Heinrich-Heine-University Düsseldorf, Germany

³Department of Statistical Science, University College London, London, United Kingdom

⁴Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

A common problem in clinical trials is to test whether an effect of an explanatory variable on the response, e.g. the effect of the dose of a compound on efficacy, is similar between two groups. In this context, similarity is defined as equivalence up to a pre-specified threshold specifying the accepted deviation between the groups. Such question is usually assessed by testing whether the (marginal) effects of the explanatory variable on the response are similar, based on, for example, confidence intervals for differences, or, to mention another example, the distance between two parametric models. These approaches typically assume a univariate continuous or binary outcome variable. An approach for associated bivariate binary response variables, based on the Gumbel model, has been recently introduced [1].

In this talk, we propose a flexible extension of such methodology that builds on a generalized joint regression framework with Gaussian copula. Compared to existing approaches, this allows for various scales of the outcome variables (e.g. continuous, binary, categorical, ordinal) including mixed outcomes as well as responses with more than two dimensions. We demonstrate the validity of our approach by means of a simulation study. An efficacy-toxicity case study demonstrates the practical relevance of the approach.

References

- [1] Möllenhoff, K., Dette, H., and Bretz, F. (2021). Testing for similarity of binary efficacy-toxicity responses. *Biostatistics*, 23(3), 949–966.

2024-02-28 17:00 - 17:20, AM S2

SES-20: Preclinical Drug Development and Toxicology 2

Prediction intervals for counted observations and their application in toxicological and medical quality control charts

Menssen, Max¹; Fneish, Firas²; Schaarschmidt, Frank¹

¹Leibniz University Hannover, Department of Biostatistics

²German Multiple Sclerosis Registry

In clinical trials as well as in toxicology, many endpoints of interest are counted events. If the endpoint is comprised of observations that are counted based on a given baseline unit, such as the number of relapses per patient which suffer from multiple sclerosis, it seems natural to model the observed events based on the Poisson distribution, whereas for a binary endpoint, such as the numbers of rats with a tumor vs. the number of rats without a tumor, a binomial model is considered as a common standard.

In quality-control it is of interest, to show that a process is stable over time, between patients or historical control groups. In many cases, this evaluation is done based on Sheward control charts. In such charts the number of events is plotted against time, patients or historical study ID, together with 95% or 99% control limits.

It has to be stressed, that the control limits used in Sheward control charts for counted observations (c , p , np) can be interpreted as heuristical prediction intervals which are explicitly based on the assumption that the observations are either Poisson (c chart) or binomial distributed (p or np chart). Nevertheless, this control limits have two major drawbacks: First, they do not account for the uncertainty of the estimates used for control limit calculation and hence should yield limits that do not reach the desired coverage probability. Second, they do not account for overdispersion, which is usually present in medical and toxicological data, since it is often the case, that the observed events per experimental unit (e.g. relapses per patient) are positively correlated, leading to higher variability between experimental units than possible under the simple Poisson or binomial model.

This work presents prediction intervals for overdispersed Poisson and binomial data, as well as their application in modified Sheward control charts on real-life-data. The proposed prediction intervals are publicly available within the framework of the R package „predint“.

2024-02-28 17:20 - 17:40, AM S2

SES-20: Preclinical Drug Development and Toxicology 2

Methods of model selection for models with common parameters

Gül, Onur; Schorning, Kirsten

TU Dortmund, Germany

The analysis of gene-expression data leads to a high-dimensional statistical problem where thousands of concentration-response data have to be analysed. For instance, the concentration-response data provided in the Valproic acid (VPA) data set (1) the information about the concentration-response relationship of more than 20.000 genes. Fitting each of these concentration-response data separately to a non-linear model leads to a complex model with many parameters and a corresponding high-dimensional estimator with high variance.

Assuming that some genes behave similarly and that the corresponding concentration-response data can be fitted by non-linear models with common parameters, can reduce the number of unknown parameters substantially. In particular, it might be reasonable that the concentrations at which 50% of the maximum effect is achieved (EC_{50}) are at least similar for some genes and therefore these parameters can be assumed to be the same across the considered non-linear models. This assumption causes a reduction of the variance of the lower-dimensional parameter estimator, but also a bias, as the assumed shared parameters are only similar, but not the same.

In this talk, we answer the question under which circumstances the less complex model with the additional assumption of common parameters should be used instead of the complex model where all genes are considered separately. More precisely, we derive asymptotic properties of the estimators in each of the models in order to calculate the asymptotic mean squared errors. Based on the asymptotics, we derive a model selection criterion which selects the model (with common parameters) leading to the smallest mean squared error.

We show in a simulation study that the derived model selection criterion performs well in comparison to other common selection criteria. Moreover, we apply the developed model selection criterion to the VPA data set in order to estimate the EC_{50} .

References

1. Krug, A. K., Kolde, R., Gaspar, J. A., Rempel, E., Balmer, N. V., Meganathan, K., Vojnits, K., Baquie, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R.

M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S., Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J. R., Hothorn, L., Bremer, S., van Thriel, C., Krause, K. H., Hengstler, J. G., Rahnenführer, J., Leist, M. und Sachinidis, A. (2013): Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach, Arch Toxicol, 87, S. 123–143.

2024-02-29 09:00 - 09:20, AM S2

SES-21: Clinical Trials

"Randomize the first patient" - old, but still most important concept

Großhennig, Anika¹; Koch, Armin¹; Beutel, Gernot²; Theodor, Framke¹

¹Institute of Biostatistics, Hannover Medical School, Germany

²Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, Germany

Already since the first half of the 20th century randomization, replication, and blocking are the three established principles in the design of experiments. They were advocated by Ronald Fisher who also made statistical analysis methods accessible to a broader scientific audience in various editions of his book Statistical Methods for Research Workers, e.g. (1). Although reminders for the need for randomization have been published repeatedly, e.g. (2, 3), there are still many non-randomized trials. Particularly in rare diseases, where the available sample size for a clinical trial is restricted, control groups are often omitted with this argument. To illustrate the importance of randomization we discuss a case study in the field of stem cell transplantation (4). Different designs for one and the same research questions and hypothetical and real results are reported and discussed. We argue that randomization techniques should be implemented routinely whenever applicable and even if the available number of patients is small. This discussion is especially relevant in the context of the current draft EMA Reflection Paper on single-arm trials (5).

References

1. Fisher, R. A. (1934). Statistical methods for research workers (5th ed.). Edinburgh; London: Oliver and Boyd.
2. Altman, D G. (1991). Randomisation. BMJ, 302(6791), 1481–1482.
<https://doi.org/10.1136/bmj.302.6791.1481>
3. Senn, S. (2013). Seven myths of randomisation in clinical trials. Statistics in Medicine, 32(9), 1439–1450. <https://doi.org/10.1002/sim.5713>
4. Hannover Medical School. (2018). Randomized, Multi-centre, Phase II Trial to compare the Event-Free Survival of Clofarabine / Ara-C (CI AraC) or of FLAMSA Treatment in Patients with High Risk AML or Advanced MDS scheduled for Allogeneic Stem Cell Transplantation. Retrieved from <https://portal.dimdi.de/data/ctr/O-4037145-2-0-940DAB-20180515091404.pdf>
5. EMA. Reflection paper on establishing efficacy based on single arm trials submitted as pivotal evidence in a marketing. Retrieved from
https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing_en.pdf

2024-02-29 09:20 - 09:40, AM S2

SES-21: Clinical Trials

A holistic approach to improve chronic kidney disease trials - unlocking the potential of hierarchical composite endpoints

Tasto, Christoph

Bayer AG, Germany

Clinical trials in chronic kidney disease (CKD) often utilize composite endpoints comprising clinical events such as onset of end-stage kidney disease (ESKD) and initiation of kidney function replacement therapy (KFRT), along with a sustained large (e.g., 50%) decrease in glomerular filtration rate (GFR). Such events typically occur late in the disease course, resulting in large and long trials in which most participants do not contribute clinical events. More recently, the rate of GFR decline over time (i.e., GFR slope) has been suggested as a more efficient endpoint, and the EMA published a Draft Qualification opinion for GFR slope as a Surrogate Endpoint in CKD trials. This endpoint is considered particularly useful in early CKD stages as well as patient populations with slower CKD progression.

We introduce the use of hierarchical composite endpoints (HCEs) in clinical trials of CKD progression, emphasizing the potential to combine clinical events with the continuous GFR slope, while ranking all components according to clinical importance. Post-hoc analyses of several large CKD trials illustrate the application of the newly developed kidney HCE including bootstrap-based efficiency comparisons with the established endpoints.

The prioritization of clinical outcomes and ability to combine clinical outcomes with GFR slope make the HCE an attractive alternative endpoint that holistically captures CKD progression.

References

Heerspink et al: Development and Validation of a New Hierarchical Composite End Point for Clinical Trials of Kidney Disease Progression. *Journal of the American Society of Nephrology*, 2023

Little et al: Validity and Utility of a Hierarchical Composite End Point for Clinical Trials of Kidney Disease Progression: A Review. *Journal of the American Society of Nephrology*, 2023

2024-02-29 09:40 - 10:00, AM S2

SES-21: Clinical Trials

Sample size calculation for cluster randomized trial with heterogeneous cluster size within cluster variances

Franco Castiblanco, Ana Carolina; Brannath, Werner

University of Bremen, Germany

The standard sample size formulae for cluster randomized trials assumed that the within cluster variability is homogeneous among clusters. In practice, however, the within cluster variability may not be constant and the standard formula may be biased. We propose a general sample size formula for cluster randomized trials when the within cluster variability is heterogeneous for both constant and variable cluster-wise sample sizes, and its simplification for two and three level trials. In addition, we propose estimators for the variance components based on conditional means. Furthermore, we conduct simulation study to investigate the behavior of the proposed sample size formula and variance components estimation and we compare it with the standard sample size formulae and the estimation of the variance components via multilevel linear models.

2024-02-29 10:00 - 10:20, AM S2

SES-21: Clinical Trials

Optimal standardization as an alternative to matching using propensity scores

Glimm, Ekkehard; Yau, Lillian

Novartis Pharma, Switzerland

To compare the effectiveness of different medical treatments in observational studies, or across different clinical studies, it is necessary to eliminate the influence of confounding factors if these are differently distributed in the treatment groups. A popular method for confounder adjustment is inverse probability weighting using propensity scores estimated from logistic regression as weights. While this method achieves "roughly matched" groups, determining when the matching is deemed "close enough" often sparks extensive debates.

In this talk, we propose a novel approach to the matching problem by reframing it as a constrained optimization problem. We explore the conditions under which a perfect match can be achieved, in the sense that the average value of confounders becomes identical in the treatment groups post-matching. We discuss the utilization of different objective functions, such as a function maximizing the effective sample size (ESS), to identify a specific set of weights that satisfy the given constraints. Depending on the chosen objective function, targeted optimizers like LPSOLVE or quadprog in R can be employed to efficiently determine these matching weights.

Our approach is closely related to the matching-adjusted indirect comparison approach by Signorovitch et al (2010). However, we go beyond their suggestion by not insisting on a specific functional form for the matching weights. In addition, the suggested approach can be applied to individual patient data from a treatment groups as well as in situations where in some groups only aggregated data is available.

In the talk, we will introduce the basic idea, apply the proposed approach to a dataset from an observational study, and compare the results with those obtained from propensity score matching.

References

Signorovitch JE, Wu EQ, Andrew P, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *PharmacoEconomics*. 2010;28(10):935-945.

Abstracts of Contributed Talks

Glimm, E. and Yau, L. (2022): Geometric approaches to assessing the numerical feasibility for conducting matching-adjusted indirect comparisons. Pharm Stat 21, 974-987.

2024-03-01 14:40 - 15:00, AM S3

SES-26: Teaching Statistics

The Power of Data: A Story by global Biostatistics and Data Sciences (gBDS)

Kunz, Cornelia Ursula

Boehringer Ingelheim Pharma GmbH & Co KG, Germany

In today's data-driven world, understanding the fundamentals of clinical trials and statistics is crucial not just for statisticians, data scientists and health care professionals but also for the general public. For a department whose jobs are computer based, it is often a challenge to visualize our daily tasks to people.

"The Power of Data" is a concept designed by the gBDS department of Boehringer Ingelheim for the Open Day. At the center of the concept is a story of aliens facing an outbreak of a disease for which they must develop a treatment.

To engage with as many people as possible considering different ages, backgrounds, languages etc., we designed a cartoon explaining in colorful pictures the different aspects of the core competencies in our department, namely statistics, programming, data management and agile coaching. A colored version of the cartoon was printed as a poster to attract visitors while a black and white version was available for children to colorize. We also sewed cuddly toys and crafted cardboard versions of the aliens to make the story more experienceable – especially for children. Based on a self-made flip flap book, visitors could create their own aliens by mixing a matching different heads, bodies, and feet. Each part of the flip flap book contained a partial code that put together served as the access code for an R Shiny App we also developed. Upon entry of the access code, data like age, planet, treatment group and outcome of the treatment were shown together with a picture of the alien. In addition, the app featured tabs for browsing through the data of all entered aliens as well as summary statistics

Overall, the concept was a huge success. A total of 487 aliens were entered into the app. Given that it was often the child of the family who created the alien, we estimated that around 1500 visitors engaged with us. The concept was also well received by colleagues.

2024-03-01 15:00 - 15:20, AM S3

SES-26: Teaching Statistics

The Alienator App: Unveiling Data Science in Clinical Trials to a Lay Audience

Andersen, Lars

Boehringer Ingelheim Pharma GmbH & Co KG, Germany

In the current changing environment of the pharmaceutical industry, understanding the fundamentals of clinical trials including statistics, programming, and data management is becoming increasingly important for making informed decisions about healthcare and medical treatments.

"The Alienator App" is an R-Shiny app developed to introduce these essential concepts to a lay audience. Using elements of gamification, the app presents clinical trial principles through an engaging story involving aliens seeking to develop a treatment for a disease. We created this tool for our open-door day to bring the complex topic of the clinical trial process from data collection to data analysis closer to a lay audience.

Our exploration of clinical trials begins with simulating aliens by randomly combining various body parts (heads, bodies, and feet) which would then serve as the study participant. A picture of the generated alien is shown including whether the alien is still diseased or cured. The aliens are then randomly assigned to either the treatment or placebo group. Another tab features a data base containing all aliens generated so far while a different tab showcases the R code behind the app. Various forms of visualizations of the aggregated information were also displayed. Summary statistics for age or planet of origin as well as disease status by group together with the results of a Chi-squared test were then shown in the last tab.

Each tab also featured a cartoon picture of an alien stating their expertise in the clinical trial as either data manager, programmer or statistician covering all skills of a clinical data scientist as well as a clinical data engineer. The app was accompanied by a cartoon story featuring the aliens from the app. A black and white version of the cartoon was also available for children to colorize.

The app was used on two different occasions and received very positive feedback – from parents and their children as well as colleagues from other departments. Overall, we conclude that color stories with interactive technologies can aid in bringing our profession to life for a general audience.

2024-03-01 15:20 - 15:40, AM S3

SES-26: Teaching Statistics

If four programs do the same thing, it's still not the same... (A model building catalog)

Sahlmann, Jörg

Universitätsklinikum Freiburg / IMBI, Deutschland

When the statistical software landscape was still manageable and was led by SAS and SPSS, there was no question of how to formulate and calculate regression models.

The SAS code provided was considered the gold standard.

Meanwhile, other programs such as R, Python and Julia claim to be at least equal in the field of statistics and to be able to replace the old top dogs.

If you have statistical models in front of you that have been provided as SAS code, it becomes clear that with more complicated models it is not that easy to map them 1 to 1 and reproduce them with the other programs.

Various models from simple linear regressions to generalized linear mixed models are presented in different computer languages and converted into one another.

Similarities and differences are discussed in order to explain any different results with the same data and the same model.

The various models and their implementation in the SAS, R, Python and Julia languages are provided as a growing model building catalog on a public GitHub repo to be presented there for teaching purposes. At the same time, this repo should provide a platform for other models and other languages (e.g. SPSS, Stata, Matlab, Octave).

2024-03-01 15:40 - 16:00, AM S3

SES-26: Teaching Statistics

Overview over 5 years of Academia meets Industry Workshop

**Lehn, Annette¹; Schulte-Göbel, Marlene¹; Jahn, Antje²; Stucke-Straub, Kathrin³;
Beyermann, Jan⁴; Lanius, Vivian⁵; Lang, Tina⁵; Scharpenberg, Martin⁶; Brannath,
Werner⁶; Kunz, Cornelia Ursula⁷**

¹Merck Healthcare KGaA

²Hochschule Darmstadt

³Technische Hochschule Ulm

⁴Universität Ulm

⁵Bayer

⁶Universität Bremen

⁷Boehringer Ingelheim Pharma GmbH & Co KG, Germany

The "Academia meets Industry (AMI)" workshop, an IBS-DR initiative, has been running successfully for the past five years. This annual event is organized on a voluntary basis by a dedicated pharmaceutical company, ideally jointly with academic institutions. Its primary objective has been to bring together various companies and universities to support the next generation of statisticians and data scientists. This platform has provided a dynamic environment for students and industrial partners to connect and foster relationships. By participating in the workshop, companies and universities showcase their commitment to the field's development and contribute to the growth of young professionals.

We will present the concept of and the experiences with the AMI workshop and will propose future plans for the upcoming event. Within the past 5 years, 20 companies and 8 universities joined the event, presenting potential topics for internships or theses or details about study programs – opening doors for BSc, MSc, and PhD students. The workshop has also shed light on the diverse career paths available in the field of Medical Statistics. Throughout the breakout sessions, 40-50 students actively engage with practitioners, gaining insights and networking opportunities.

One of the workshop's significant achievements is its focus on nurturing the next generation of statisticians and data scientists. By connecting academia and industry, it provides students a clear understanding of the practical applications of their academic knowledge and the skills required in the industry.

In conclusion, the "Academia meets Industry" workshop plays a pivotal role in shaping the future of Medical Statistics and successfully fosters the relationships between academia and industry.

2024-03-01 13:00 - 13:20, AM S3

SES-27: Nonparametric Methods

Inference for Random Effects in Nonparametric Repeated Measures Designs with Missing Data via Randomization

Amro, Lubna; Dobler, Dennis; Kuhn, Jörg-Tobias

TU Dortmund University, Germany

Relative effects are a useful measure for quantifying trends, e.g., in ordinal data. The methodology behind relative effects has been extended in many different directions. In this talk, we will focus on repeated measures designs with missing data. Here, relative (marginal) effects can be used to find a time or other effects on the outcomes. In a previous work, relevant theory had been developed and hypothesis tests were derived. In this presentation, we re-visit the testing problems by means of a randomization procedure which will give rise to asymptotically exact inference procedures. Also, we will investigate the role of data missing at random. Simulations demonstrate the small sample performance and an analysis of real data on children learning math illustrate several aspects of our method.

2024-03-01 13:20 - 13:40, AM S3

SES-27: Nonparametric Methods

NANCOVA: Nonparametric Analysis of Covariance for Rare Disease Research

Thiel, Konstantin Emil^{1,2}; Sattler, Paavo³; Bathke, Arne C.²; Zimmermann, Georg¹

¹Research Programme Biomedical Data Science, Paracelsus Medical University, Salzburg, Austria

²Faculty of Digital and Analytical Sciences, University of Salzburg, Austria

³Department of Statistics, TU Dortmund University, Germany

Developing treatments for rare diseases is challenging as the small sample sizes faced in randomized clinical trials (RCTs) strictly limit the precision of statistical tests for treatment effects. Consequently, it is crucial to improve precision by leveraging covariates, i.e., variables associated with the outcome measure of interest. The idea is to explain observed variance by covariates, and thus, increase power of tests for treatment effects. In recent guidelines, even the European Medicines Agency recommends such covariate adjustment in RCTs. Thereby, the validity of model assumptions must be checked to avoid false conclusions caused by inflated type-I errors. Again, this is challenging with rare diseases as it is infeasible to check distributional assumptions when sample sizes are small. Hence, the strict assumption of normally distributed data prevents the use of parametric ANCOVA (analysis of covariance), which is otherwise a standard tool for covariate adjustment in RCTs.

To solve this problem, we propose NANCOVA, a nonparametric tool based on relative effects. Therein, we provide a new bootstrap test that can be reliably used in small sample RCTs. The use of nonparametric methods even extends the scope of NANCOVA to scenarios with purely ordinal data such as patient questionnaires. The suitability of our proposed bootstrap test is examined both theoretically and empirically.

2024-03-01 13:40 - 14:00, AM S3

SES-27: Nonparametric Methods

Sample size planning for rank-based multiple contrast tests

Pöhlmann, Anna¹; Brunner, Edgar²; Konietzschke, Frank¹

¹Charité - Universitätsmedizin Berlin, Germany

²University Medical Center Göttingen, Germany

Any preclinical trial should start with meticulous sample size considerations. Errors in statistical planning have severe consequences on conclusions drawn from the data. Also, planning with the wrong number of subjects is ethically unacceptable. Many experiments consist of multiple samples, e.g., when different dose levels are investigated. We develop methods for sample size computations suitable for analyzing several samples without relying on any distributional assumption.

While rank methods have long been established as valuable tools for comparing two or more independent groups, the field has lacked comprehensive statistical planning methods to determine the necessary sample size(s) to detect specific alternatives with predefined power. In response to this need, we introduce new approaches for sample size planning specifically tailored for pseudo-rank-based multiple contrast tests.

We discuss the treatment effects in detail and offer various approaches to approximate variance parameters within the estimation scheme. Additionally, the study conducts a thorough comparison between pairwise and global rank methods. To assess the practicality and accuracy of the proposed sample size estimators, extensive simulation studies have been carried out, demonstrating the reliability of these methods under various conditions. To illustrate the real-world applicability of these techniques, a real data example is presented, showing the potential of the proposed methods in guiding preclinical trial design.

2024-03-01 14:00 - 14:20, AM S3

SES-27: Nonparametric Methods

A non-parametric proportional risk model to assess a treatment effect in an application to long-term carcinogenicity assays

Ameis, Lucia¹; Kuß, Oliver²; Hoyer, Annika³; Möllenhoff, Kathrin¹

¹Heinrich Heine University Düsseldorf, Düsseldorf, Germany

²German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Institute for Biometrics and Epidemiology, Düsseldorf, Germany

³Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany

Time-to-event analysis often relies on assumed parametric models or, if a semi-parametric approach was chosen, Cox's model that is inherently tied to the assumption of proportional hazards. This limits the quality of the results in case of any violation of these assumptions. Especially the assumption of proportional hazards was recently criticized for being rarely verified. In addition, most interpretations focus on the hazard ratio, that is often misinterpreted as the relative risk and comes with the restriction of being a conditional measure. Our approach introduces an alternative to the proportional hazard assumption and allows for a direct estimation of the relative risk as well as the absolute measure of the number needed to harm, therefore provides the possibility of an easy and holistic interpretation.

In this talk, we propose a new non-parametric estimator to assess the relative risk of two groups to experience an event under the assumption that the risk is constant over time, namely the proportional risk assumption. Precisely, we first estimate the respective cumulative distribution functions of both groups by means of the Kaplan-Meier estimator and second combine their ratio at different time points to estimate the mean relative risk. We then combine the result with one of the estimated cumulative distribution functions to assess the number needed to harm. This offers the possibility to interpret the treatment effect solely based on a Kaplan-Meier estimator, a well-established method without prior parametric assumptions, and offers a flexible alternative to Cox's proportional hazard model if the proportional hazard assumption is violated.

We demonstrate the validity of the approach by means of a simulation study and present an application to mortality data of mice from a study investigating the long-term carcinogenicity of piperonyl butoxide.

2024-02-28 14:40 - 15:00, AM S1

SES-28: Complex Models

Multiple contrast tests for possible overdispersed count data: small sample approximations

Pigorsch, Mareen¹; Hothorn, Ludwig A.²; Konietschke, Frank¹

¹Charité - Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology, Germany

²31867 Lauenau

Count data occur frequently in statistical practice [1]. Typically, researchers choose a Poisson or Negative Binomial distribution model for making statistical inferences. These data, however, often exhibit over/ under-dispersion or zero- inflation making these distributional assumptions doubtful and test results inaccurate. Statistical methods not relying on specific distributional assumptions therefore are a great alternative often posing more robust and accurate conclusions. In this talk, we discuss multiple contrast tests [2] allowing for the use of general contrasts (e.g., many-to-one or all-pairs comparisons) for the analysis of count data in multi-arm trials. The methods differ in effect and variance estimation as well as in the approximation of the joint distribution of the multiple test statistics. An extensive simulation study shows that a resampling version controls the type-I error rate fairly well in various situations as well as the methods' limitations with too liberal type-I error rates. Some of the standard methods show type-I error rates of over 30% in situations, where the resampling version controls the type-I error rate well. Real data applications demonstrate their applicability.

References

- [1] Eduard Szöcs and Ralf B Schäfer. Ecotoxicology is not normal: A comparison of statistical approaches for analysis of count and proportion data in ecotoxicology. Environmental Science and Pollution Research, 22:13990–13999, 2015.
- [2] Frank Konietschke. Ludwig A. Hothorn. "Rank-based multiple test procedures and simultaneous confidence intervals." Electron. J. Statist. 6 738 – 759, 2012.

2024-02-28 15:00 - 15:20, AM S1

SES-28: Complex Models

Personalized patient-specific mechanistic mathematical models of acute myeloid leukaemia disease dynamics

Görlich, Dennis

University of Münster, Germany

Mathematical models and statistical data analysis are traditionally focused on elucidating the average response of observable factors within a population. However, in the context of clinical predictions, a more personalized viewpoint often proves essential for making informed treatment decisions tailored to individual patients.

In this discussion, we present our method for fitting mechanistic mathematical models of acute myeloid leukemia to patient data. Specifically, we employed the ordinary differential equation (ODE) model outlined in Banck & Görlich (2019) for this purpose.

Our strategy involves aligning outcome measures commonly found in clinical datasets for acute myeloid leukemia with the predictive capabilities of our model. While several clinical outcomes, such as blast load, complete remission, and survival, are considered, only blast cell count (leukemic burden) is currently well-suited for integration into the model fitting process (Görlich, 2021).

The process of determining personalized model parameters entails solving an optimization problem based on a time series of blast cell counts. By constraining the search space to a subset of model parameters, the level of personalization can be tailored. A preliminary grid search indicated that an optimal solution to the optimization problem typically exists, although the fitness landscape can be noisy. In instances of high noise, gradient-based solvers may perform inadequately, necessitating the selection of alternative algorithms. For our analysis, we opted for the use of the differential evolution algorithm outlined in Mullen et al. (2011).

To validate our approach, we applied it to data from 126 acute myeloid leukemia patients in a proof-of-concept analysis. A subsequent cluster analysis identified 5 subgroups characterized by distinct model parameter combinations. These clusters demonstrated correlations with clinical parameters and outcomes, notably including an association with overall response rate (therapy success) and overall patient survival.

In summary, our modeling approach facilitates the integration of combination chemotherapy of cytarabine and anthracyclines into acute myeloid leukemia dynamics. The proof-

of-concept analysis yielded promising results, establishing validity by correlating clinical data. However, it also raised questions about model identification that warrant exploration in future research. Ultimately, the utilization of validated personalized models in clinical care holds the potential to become a valuable tool for supporting informed clinical decision-making in the long term.

References

Görlich D. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, Volume 3: BIOINFORMATICS 2021, Ronny Lorenz, Ana Fred, Hugo Gamboa (Eds.) ISBN: 978-989-758-490-9, published by SCITEPRESS.

Banck JC & Görlich D. In-silico comparison of two induction regimens (7 + 3 vs 7 + 3 plus additional bone marrow evaluation) in acute myeloid leukemia treatment. BMC Syst Biol. 2019;13(1).

Mullen KM, Ardia D, Gil DL, Windover D, Cline J. DEoptim: An R Package for Global Optimization by Differential Evolution. J. Stat. Soft.. 2011;40(6):1-26.

2024-02-28 15:20 - 15:40, AM S1

SES-28: Complex Models

Flexible Modeling of Biomarker Ratios with Correlated Components

Berger, Moritz¹; Klein, Nadja²; Schmid, Matthias¹

¹Department of Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn, Germany

²Chair of Uncertainty Quantification and Statistical Learning, Research Center for Trustworthy Data Science and Security (UA Ruhr) and Department of Statistics (TU Dortmund)

A frequent objective in observational studies is to model the ratio of two non-negative components. It's of major relevance in biomedical research, where biomarker ratios are used for early diagnoses of specific diseases. Examples include, among others, the neutrophil/lymphocyte ratio and platelet/lymphocyte ratio for neurological diseases, the LDL/HDL cholesterol ratio in cardiovascular research as well as the CD4/CD8 ratio of T helper cells in infectiology. In such studies, the focus typically is not only on the characterization of the marginal ratio distribution, but also on modeling this distribution as a function of covariates.

When setting up a regression model that relates the ratio outcome to covariates, two strategies are in common use. The first one assumes that the ratio outcome follows a normal distribution after transformation, the second one assumes gamma distributed components, thereby accounting for the positivity of the component values and the skewness of their distributions. We consider both approaches for modeling ratio outcomes and embed them into a distributional regression framework that allows us to relate all distributional parameters to a set of covariates. These are more flexible than classical regression approaches (which focus on the mean ratio only), as they enable to derive the conditional probability density function as well as quantities of interest (like the median or quantiles) as functions of covariates. We also derive valid tools for inference in finite samples. On the model side, we consider (i) Gaussian models with log-transformed outcome and Box-Cox-transformed outcome, and (ii) gamma distribution-based models, where the joint distribution of the two components is defined by either Kibble's bivariate gamma distribution ("extended GB2 model") or Frank's copula ("FCGAM model"). While the extended GB2 model is tailored to the case of positively correlated components, the FCGAM model allows for positive or negative correlation between the two components.

We evaluate and compare these four approaches in an analysis of data from dementia research, where cerebrospinal fluid biomarkers are used for early diagnoses of Alzheimer's

disease. In this application, measurements of the widely employed amyloid-beta 42 and amyloid-beta 40 proteins are positively correlated, whereas amyloid-beta 42 and total tau proteins exhibit a clearly negative correlation. Our results demonstrate the efficacy of the gamma distribution-based approaches.

2024-02-28 15:40 - 16:00, AM S1

SES-28: Complex Models

Comparative analysis of proportional odds models in a simulation study. A robust alternative to linear regression?

Klinger, Andreas; Heinze, Georg; Dunkler, Daniela; Gregorich, Mariella; Kammer, Michael; Kraemmer, Daniel

Medical University of Vienna, Center for Medical Data Science, Institute of Clinical Biometrics

Ordinal outcomes occur frequently in medicine. Three slightly different models for ordinal outcomes with a proportional odds (PO) assumption are the Cumulative Logit (CLPO), Adjacent Categories (ACPO) and Continuous Ratio (CRPO) model. Proportional odds models can be a robust alternative to a linear regression model if the assumption of Gaussian errors is in doubt.

In three simulation studies, we aimed at 1) evaluating if there are differences in performance between the three models, 2) comparing the CLPO model to linear regression when data was generated with CLPO, a Gaussian error model or a multinomial model with non-proportional odds and 3) comparing the CLPO model to negative binomial regression when the dependent variable resembles count data. We assumed three binary and three continuous predictors with moderate correlation. We evaluated the performance in prediction and the actual type I error rate and the power to detect a non-zero effect of a binary covariate. We also compared estimated parameters and predictions in selected data sets.

When data followed a CLPO, ACPO or CRPO model, there were only little differences in the performance when either of these models was used for analysis.

With Gaussian error data, both prediction and test performances of the CLPO model were nearly as good as those of the linear regression model. When the data was generated by the CLPO model, the linear regression model was much worse in both aspects.

When outcomes were generated with a negative-binomial distribution, the CLPO model's performance did not reach that of the data-generating model.

Although regression coefficients of PO and Gaussian models have quite different interpretation, they are similar in magnitude and have a high association across simulations. We conclude that proportional odds regression models are a viable alternative to linear regression, in particular in situations where the assumptions of the latter are in doubt. This

may be the case with bounded outcomes such as patient-reported outcome scores.

2024-03-01 09:00 - 09:20, AM S3

SES-33: Multiple Testing

Surviving the multiple testing problem: RMST-based tests in general factorial designs

Munko, Merle¹; Ditzhaus, Marc¹; Dobler, Dennis²; Genuneit, Jon³

¹Otto-von-Guericke University Magdeburg, Germany

²TU Dortmund University, Germany

³Leipzig University, Germany

Several methods in survival analysis are based on the proportional hazards assumption. However, this assumption is very restrictive and often not justifiable in practice. Therefore, effect estimands that do not rely on the proportional hazards assumption, such as the restricted mean survival time (RMST), are highly desirable in practical applications. The RMST is defined as the area under the survival curve up to a prespecified time point and, thus, summarizes the survival curve into a meaningful estimand. For two-sample comparisons based on the RMST, there is an inflation of the type-I error of the asymptotic test for small samples and, therefore, a two-sample permutation test has already been developed. The first goal is to further extend the permutation test for general factorial designs and general contrast hypotheses by considering a Wald-type test statistic and its asymptotic behavior. Additionally, a groupwise bootstrap approach is considered. In a second step, multiple tests for the RMST are developed to infer several null hypotheses simultaneously. Hereby, the asymptotically exact dependence structure between the local test statistics is incorporated to gain more power. The small sample performance of the proposed global and multiple testing procedures is analyzed in simulations and finally illustrated by analyzing a real data example.

2024-03-01 09:20 - 09:40, AM S3

SES-33: Multiple Testing

Informative simultaneous confidence intervals for graphical test procedures

Scharpenberg, Martin; Brannath, Werner

University of Bremen, Deutschland

Simultaneous confidence intervals (SCIs) that are compatible with a given closed test procedure are often non-informative. More precisely, for a one-sided null hypothesis, the bound of the SCI can stick to the border of the null hypothesis, irrespective of how far the point estimate deviates from the null hypothesis. This has been illustrated for the Bonferroni-Holm and fall-back procedures, for which alternative SCIs have been suggested, that are free of this deficiency. These informative SCIs are not fully compatible with the initial multiple test, but are close to it and hence provide similar power advantages. They provide a multiple hypothesis test with strong family-wise error rate control that can be used in replacement of the initial multiple test. In the talk we extend previous work for informative SCIs to graphical test procedures. The information gained from the newly suggested SCIs is always increasing with increasing evidence against a null hypothesis. The new SCIs provide a compromise between information gain and the goal to reject as many hypotheses as possible. They are defined via a family of dual graphs and the projection method. A simple iterative algorithm for the computation of the intervals will be presented. We investigate the new SCIs in simulation studies.

2024-03-01 09:40 - 10:00, AM S3

SES-33: Multiple Testing

Multiple contrast testing procedures for semiparametric MANCOVA

Baumeister, Marléne¹; Thiel, Konstantin Emil²; Matits, Lynn³; Kolassa, Iris-Tatjana³; Pauly, Markus¹; Zimmermann, Georg²

¹Department of Statistics, TU Dortmund University, Germany

²Research Programme Biomedical Data Science, Paracelsus Medical University Salzburg, Austria

³Institute of Psychology and Education, Clinical & Biologica IPsychology, Ulm University, Germany

In biological, medical and psychological research, there is a need for multiple inference methods. At the same time there is a need to adjust for covariates. In particular, the European Medicines Agency (EMA) recommends covariate adjustments in randomized clinical trials. However, especially for multivariate outcomes, there is a need for statistical inference methods that can deal with both issues. Powerful solutions for multiple testing are multiple contrast test procedures (MCTPs) as they redefine the rejection of the global hypothesis: it is simultaneously rejected if one of the individual comparisons is rejected. MCTPs are known to be effective for lots of models and estimates (e.g., Konietzschke et al. 2013 or Umlauft et al. 2019).

In this talk we present a new method for multiple testing in multivariate analysis of covariance settings (MANCOVA). Our method works in a semiparametric model without normality, homoscedasticity or non-singularity assumption and generalizes the global testing approach from Zimmermann et. al. (2020). We prove asymptotic validity of the new method and employ different bootstrap approaches to realize a good small sample performance as shown in extensive simulations. Our method is motivated and illustrated by a recent psychological hypnosis data set.

Konietzschke, F., Bösiger, S., Brunner, E., & Hothorn, L. A. (2013). Are multiple contrast tests superior to the ANOVA?. *The International Journal of Biostatistics*, 9(1), 63-73.

Umlauft, M., Placzek, M., Konietzschke, F., & Pauly, M. (2019). Wild bootstrapping rank-based procedures: Multiple testing in nonparametric factorial repeated measures designs. *Journal of Multivariate Analysis*, 171, 176-192.

Zimmermann, G., Pauly, M., & Bathke, A. C. (2020). Multivariate analysis of covariance with potentially singular covariance matrices and non-normal responses. *Journal of Multivariate Analysis*, 177, 104594.

References

1. Konietzschke F, Bösiger S, Brunner E, Hothorn LA. Are Multiple Contrast Tests Superior to the ANOVA? *The International Journal of Biostatistics*. 1. August 2013;9(1):63–73.
2. Umlauft M, Placzek M, Konietzschke F, Pauly M. Wild bootstrapping rank-based procedures: multiple testing in nonparametric factorial repeated measures designs. *Journal of Multivariate Analysis*. 1. Mai 2019;171:176–92.
3. Zimmermann G, Pauly M, Bathke AC. Multivariate analysis of covariance with potentially singular covariance matrices and non-normal responses. *Journal of Multivariate Analysis*. Mai 2020;177:104594.

2024-03-01 10:00 - 10:20, AM S3

SES-33: Multiple Testing

Asymptotic online familywise error rate control for dependent test statistics

Jankovic, Vincent; Fischer, Lasse; Brannath, Werner

University of Bremen, Germany

In online multiple testing, an a priori unknown number of hypotheses is tested sequentially, i.e. at each time point a test decision for the current hypothesis has to be made using only the data available so far. Although many powerful test procedures have been developed for online error control in recent years, most of them are designed solely for independent or at most locally dependent test statistics. In this work, we provide a new framework for deriving online multiple test procedures which ensure asymptotical (with respect to the sample size) control of the familywise error rate (FWER), regardless of the dependence structure between test statistics. In this context, we give a few concrete examples of such test procedures and discuss their properties. Furthermore, we conduct a simulation study to investigate the behaviour regarding type I error and power of these test procedures for finite sample size.

2024-03-01 10:40 - 11:00, AM S3

SES-34: Simulation Studies

Statistical Plemode Simulations - Potentials, Challenges and Recommendations

Schreck, Nicholas¹; Slyko, Alla²; Saadati, Maral³; Benner, Axel¹

¹DKFZ Heidelberg, Germany

²University of Waterloo, Canada

³Saadati Solutions, Germany

Statistical data simulation is essential in the development of statistical models and methods as well as in their performance evaluation. To capture complex data structures, in particular for high-dimensional data, a variety of simulation approaches have been introduced including parametric and the so-called plemode simulations. While there are concerns about the realism of parametrically simulated data, it is widely claimed that plmodes come very close to reality with some aspects of the "truth" known. However, there are no explicit guidelines or state-of-the-art on how to perform plemode data simulations. We first review existing literature and introduce the concept of statistical plemode simulation. We then discuss advantages and challenges of statistical plmodes and provide a step-wise procedure for their generation, including key steps to their implementation and reporting. We illustrate the concept of statistical plmodes as well as the proposed plemode generation procedure by means of a public real RNA dataset on breast carcinoma patients.

2024-03-01 11:00 - 11:20, AM S3

SES-34: Simulation Studies

When is Plasmode simulation superior to parametric simulation?

Stolte, Marieke; Rahnenführer, Jörg; Bommert, Andrea

Department of Statistics, TU Dortmund University, Germany

Simulation is an important tool for evaluating and comparing statistical methods. Consequently, designing fair and neutral simulation studies is of great interest to both researchers developing new methods and practitioners who are faced with choosing the most appropriate methods (1). Simulation generally refers to parametric simulation, i.e. computer experiments using artificial data composed of pseudo-random numbers. Plasmode simulation is a computer experiment that is often claimed to produce more realistic data by combining covariate data resampled from real-life data with a known user-selected outcome-generating model (OGM) to generate target variables.

Both parametric and Plasmode simulation have their advantages and disadvantages. Schreck et al. (2) discuss these and compare the two approaches on a theoretical basis. To our knowledge, there are no empirical comparisons of the performance of parametric and Plasmode simulation in practice so far. If the actual data-generating process (DGP) and the outcome-generating model (OGM) are known, parametric simulation is the best choice since in that case, it allows to easily generate large amounts of data from the true DGP and OGM. However, in reality, both the true DGP and the true OGM are usually unknown, so researchers must make assumptions: in Plasmode simulation studies on the OGM, in parametric simulations on both the DGP and the OGM. These assumptions may not fully reflect the truth. Here, we want to find out how assumptions that differ from true DGPs and true OGMs affect the performance of parametric simulations and Plasmode simulations and in which scenarios which simulation type is preferable.

We compare parametric simulation and Plasmode simulation to estimate the mean squared error (MSE) of the least squares estimator in linear regression. The results show that the preferred simulation method depends on many factors, including the number of covariates, and the way and extent to which assumptions for parametric simulation differ from the true unknown DGP. In addition, the resampling strategy used in Plasmode has an impact on the results. In particular, subsampling with a small resampling proportion can be recommended.

References

- (1) Boulesteix A-L, Lauer S, Eugster MJA. A Plea for Neutral Comparison Studies in

Computational Sciences. PLoS ONE 2013;8(4): e61562.

(2) Schreck, N, Slynko, A, Saadati, M, & Benner, A. Statistical Plasmode Simulations - Potentials, Challenges and Recommendations. arXiv preprint 2023; arXiv:2305.06028.

2024-03-01 11:20 - 11:40, AM S3

SES-34: Simulation Studies

On the role of benchmarking data sets and simulations in method comparison studies

Friedrich, Sarah¹; Friede, Tim²

¹University of Augsburg, Germany

²University Medical Center Göttingen, Germany

Method comparisons are essential to provide recommendations and guidance for applied researchers, who often have to choose from a plethora of available approaches. While many comparisons exist in the literature, these are often not neutral but favor a novel method. Apart from the choice of design and a proper reporting of the findings, there are different approaches concerning the underlying data for such method comparison studies. Most manuscripts on statistical methodology rely on simulation studies and provide a single real-world data set as an example to motivate and illustrate the methodology investigated. In the context of supervised learning, in contrast, methods are often evaluated using so-called benchmarking data sets, that is, real-world data that serve as gold standard in the community. Simulation studies, on the other hand, are much less common in this context. In this talk, we will investigate differences and similarities between these approaches, discuss their advantages and disadvantages, and ultimately develop new approaches picking the best of both worlds.

2024-03-01 11:40 - 12:00, AM S3

SES-34: Simulation Studies

Translating methodological simulation studies into practice: a reproducible application

Callahan, Patrick; Boulesteix, Anne-Laure

LMU München, Institut für Medizinische Informationsverarbeitung Biometrie und Epidemiologie

Simulations are useful for empirically assessing the properties of statistical methods. They are primarily used in methodological research to draw general conclusions on new or existing methods. Applied statisticians and data analysts may also use published simulation results or perform their own simulations with a specific application in mind such as choosing an appropriate statistical method or sample size. In our experience, however, the direct use of published simulation results for this latter purpose is hardly feasible because the settings considered in methodological publications rarely fit the characteristics of applied scenarios closely enough to be of utility. The translation of methodological simulation studies into practical settings is thus hindered. On the other end of this equation, however, applied practitioners may find it difficult to construct their own simulations or adapt methodological research to better their specific data and parameter settings due to time and coding constraints.

In an effort to improve the translation of simulation studies into practice, we propose this gap can be bridged through the development of open source prebuilt software programs. Such tools should abstract away the coding-heavy aspects of running simulation studies while still allowing sufficiently flexible (hyper?)parameter selection to meet applied statisticians' needs.

To demonstrate this in practice, we introduce an R package containing a Shiny web application capable of running simulations, enabling users to select a powerful testing method or to determine the necessary sample size to achieve a certain power in the context of two-group comparisons of ordinal (clinical trial) endpoints. Indeed, corresponding simulation results from the literature are limited to special simple settings, and are thus of poor utility for applied statisticians planning a two-arm trial with ordinal endpoint. Elaborating a simulation study from scratch in this early stage of the project is often also not an option, especially without solid experience of simulation studies. The app fills this gap. Its concept can be principally extended to any simulation study intended to support applied statisticians in the choice of methods and sample size in practice.

2024-03-01 13:00 - 13:20, AM 2

SES-35: Diagnostic Studies

Unblinded sample size re-estimation in diagnostic test accuracy studies

Köster, Denise¹; Hoyer, Annika²; Zapf, Antonia¹

¹Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf

²Biostatistics and Medical Biometry, Medical School EWL, Bielefeld University

Diagnostic test accuracy studies aim to evaluate the accuracy of diagnostic test measured by sensitivity and specificity (true positive and true negative rate) as co-primary endpoints. Within such studies, an experimental diagnostic test is compared to a reference standard, that defines the true disease status, and possibly in addition to another test, referred to as comparator test. This leads to a single-test or comparative design. In a single-test design, the experimental test is compared to the reference standard, proving a minimum sensitivity and specificity. If there is a comparator test, both tests and the reference standard should be applied in all study participants if it is ethically acceptable and feasible (1).

A sample size calculation should be performed a priori in diagnostic accuracy studies. Assumptions for the sample size calculation are often uncertain and from small preliminary studies leading to possible biased estimates (2). Adaptive designs can be used in diagnostic accuracy studies to correct these uncertain assumptions during the ongoing study. These designs are long-established in therapeutic studies (3) and allow pre-specified adaptations, e.g. the adjustment of the sample size based on an interim analysis. There are two ways to perform interim analyses: blinded and unblinded. In blinded interim analysis the prevalence can be estimated and, e.g. used in a sample size re-calculation. Whereas in unblinded interim analysis sensitivity and specificity are estimated and, e.g. used in a sample size re-calculation. In this case, the type one error needs to be adjusted accordingly. In this talk we propose methods for single-test and comparative diagnostic accuracy studies in an adaptive two-stage design with an unblinded sample size re-estimation illustrated with an example study. Furthermore, the results of our simulation study will be presented.

References

1. European Medicines Agency. CPMP/EWP/1119/98 - Guideline on clinical evaluation of diagnostic agents; 2009. Available from: URL: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-evaluation-diagnostic-agents_en.pdf [cited 2020 Jun 7].

2. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174(4):469–76. Available from: URL: <https://www.ncbi.nlm.nih.gov/pubmed/16477057>.
3. Wassmer G, Brannath W. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. Cham: Springer International Publishing; 2016.

2024-03-01 13:20 - 13:40, AM 2

SES-35: Diagnostic Studies

**Comparing methods to handle missing values in the index test in diagnostic studies
- a simulation study**

Stahlmann, Katharina¹; Kellerhuis, Bas²; Zapf, Antonia¹

¹University Medical Center Hamburg-Eppendorf, Hamburg, Germany

²University Medical Center Utrecht, Utrecht, The Netherlands

Introduction: Inappropriate handling of missing values can lead to biased results in diagnostic accuracy studies (1). While there are well-known methods to handle missing values in the reference standard (2), missing values in the index test are often ignored or replaced by single imputation (3,4). One reason for this may be low awareness of methods to handle missing values in the index test and the lack of systematic comparison of these methods (5). Therefore, this simulation study compares the performance of methods for estimating the area under the ROC curve (AUC) of a continuous index test with missing values in a diagnostic study.

Methods: We simulated data including a continuous index test with missing values, a binary reference standard, and three covariates (one binary, two continuous). We varied the following factors: the sample size, the true AUC, the prevalence of the target condition, the percentage of missing values, and the missingness mechanism (missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR)). The methods for estimating the AUC were chosen from a previous review of methods for handling missing values in diagnostic studies (5). They comprised complete case analysis, several multiple imputation (MI) approaches, single imputation with empirical likelihood confidence intervals, robust inverse probability weighting, and a kernel-based approach. To determine performance, we examined bias, root mean squared error (RMSE), coverage of confidence intervals, and statistical power.

Results: Preliminary results show that bias and RMSE increase with a higher proportion of missing values, and with MAR and MNAR compared to MCAR. Especially the single imputation with empirical likelihood approach and the complete case analysis tend to underestimate the AUC. Some MI methods and the kernel-based method have a lower coverage probability with increasing percentage of missing values, while the single imputation method has a substantially reduced coverage probability if the true AUC is high. The robust inverse probability methods perform quite well across all performance parameter. However, it takes the most time to run.

Discussion: The choice of the “best” method depends, among others, on the missingness

mechanism and the percentage of missing values. If MCAR is a reasonable assumption, a complete case analysis may lead to valid results, but with decreased statistical power. However, some MI approaches or robust inverse probability may be a better choice for MAR. Sensitivity analysis should be conducted to explore the robustness of the results under different missingness assumptions.

References

- 1.Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013; 66: 1093-1104. doi: 10.1016/j.jclinepi.2013.05.014 .
- 2.Chikere CMU, Wilson K, Graziadio S, et al. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One* 2019; 14: e0223832. doi: 10.1371/journal.pone.0223832 .
- 3.Shinkins B, Thompson M, Mallett S, et al. Diagnostic accuracy studies: how to report and analyse inconclusive test results. *BMJ* 2013; 346: f2778. doi: 10.1136/bmj.f2778.
- 4.Schuetz GM, Schlattmann P and Dewey M. Use of 3×2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *BMJ* 2012; 345: e6717. doi: 10.1136/bmj.e6717.
- 5.Stahlmann K, Reitsma JB, Zapf A. Missing values and inconclusive results in diagnostic studies - A scoping review of methods. *Stat Methods Med Res.* 2023;32(9):1842-1855. doi: 10.1177/09622802231192954.

2024-03-01 13:40 - 14:00, AM 2

SES-35: Diagnostic Studies

Covariate adjustment, factorial designs and clustered data in diagnostic accuracy studies

Weber, Philipp¹; Kramer, Katharina²; Zapf, Antonia¹

¹University Medical Center Hamburg-Eppendorf, Germany

²University of Augsburg, Germany

The accuracy of diagnostic tests is commonly evaluated by estimating the area under the receiver operating characteristic curve (AUC), as well as sensitivity and specificity at given diagnostic cut-offs. However, many diagnostic trials use factorial designs. For example, different combinations of readers and methods may be used to diagnose a patient. Furthermore, diagnostic studies may generate clustered data by repeated measurements over time or several lesions, for example different brain regions. Dependencies between a person's observations must be taken into account in the analysis in order to prevent variance deflation. Lange [1] developed a nonparametric mathematical framework to deal with both of these design aspects, and Lange and Brunner generalized the approach from the AUC to sensitivity and specificity [3].

Additionally, it may be of interest to adjust the estimation procedure of the above mentioned accuracy measures for covariates. For example, it may be the case that age, weight or height influence the diagnostic accuracy of a test. Zapf [2] proposed a nonparametric methodological approach to adjust the AUC for such covariates, while also allowing for factorial designs, but not yet for clustered data.

In this talk we present a new, unified method that enables covariate adjustment of the AUC, sensitivity and specificity in studies with factorial designs and clustered data. We will show the properties of the approach using simulated data and illustrate the approach with an example study.

References

1) Lange, K. (2011, March 4). Nichtparametrische analyse diagnostischer Gütemaße bei Clusterdaten. Retrieved February 27, 2023, from <http://dx.doi.org/10.53846/goediss-3538>

2) Zapf, A. (2009, October 23). Multivariates nichtparametrisches Behrens-Fisher-problem MIT Kovariablen. Retrieved February 27, 2023, from <http://dx.doi.org/10.53846/goediss-2488>

- 3)Lange, K., & Brunner, E. (2012). Sensitivity, specificity and ROC-curves in multiple reader diagnostic trials—a unified, nonparametric approach. *Statistical Methodology*, 9(4), 490–500. <https://doi.org/10.1016/j.stamet.2011.12.002>

2024-03-01 14:00 - 14:20, AM 2

SES-35: Diagnostic Studies

Power and sample size estimation for comparing diagnostic methods with imperfect reference standards

Paul, Roman Harald¹; Othman, Ahmed²; Altmann, Sebastian²; Schmidtmann, Irene¹

¹Institute of Medical Biostatistics, Epidemiology or Informatics, University Medical Center Mainz

²Department of Neuroradiology, University Medical Center Mainz

Introduction: New imaging protocols have to be evaluated before being used routinely. Ideally, this is done by comparing the new method to a perfect reference standard. However, obtaining such a reference standard is not always feasible. Consequently, it may be necessary to assess interchangeability of an existing well-established method, thus providing similar outcomes for individual patients. Multi-reader multi-case (MRMC) studies are commonly used to perform comparative assessments of imaging protocols as the interpretation of images also depends on the ability of the reader.

Obuchowski (1) introduced a measure of interchangeability and derived bootstrap confidence intervals, which can be used to test for equivalence. In a subsequent paper, Obuchowski et al. (2) provided tests for interchangeability. A general method for sample size for such studies is currently lacking.

Methods: We considered an MRMC for diagnostic methods with a binary outcome. We assumed that probability of correct decision with each method is related to patient difficulty and rater ability via a logistic regression model, thus allowing for heterogeneity and dependence.

Multiple MRMC simulations were executed with a diverse set of linear predictors, generating a range of underlying values for sensitivity and specificity for the two compared methods. To test for interchangeability, we estimated the probability of agreement using the existing method twice, each time assessed by different readers, and the probability of agreement using each method once. Estimation of these probabilities was conducted using a Generalized Estimating Equation (GEE) model. Subsequently, we derived confidence intervals for the differences in these probabilities using the delta method. Method interchangeability could be demonstrated, if the confidence interval limits did not surpass a predetermined margin, representing the maximum acceptable difference, as proposed by Obuchowski et al. (2).

Results: Our simulations indicate small to moderate effects of variable sensitivity and

specificity on power for given sample size. Lower prevalence is associated with higher power for given sample size, while lower underlying accuracy is associated with lower power for given sample size.

Discussion: The proposed method allows to estimate the power of a test for interchangeability, given the number of cases, the number of readers, and some plausible assumptions about the acceptable method difference, the distribution of case difficulty, reader ability, and their link to sensitivity and specificity. However, the method depends heavily on prevalence of the trait that is to be diagnosed.

References

1. Obuchowski NA. Can electronic medical images replace hard-copy film? Defining and testing the equivalence of diagnostic tests. *Stat Med*. 15. Oktober 2001;20(19):2845–63.
2. Obuchowski NA, Subhas N, Schoenhagen P. Testing for Interchangeability of Imaging Tests. *Acad Radiol*. November 2014;21(11):1483–9.
3. Almansour H, Herrmann J, Gassenmaier S, Afat S, Jacoby J, Koerzdoerfer G, Nickel D, Mahmoud M, Mariappan N, Othman AE. Deep Learning Reconstruction for Accelerated Spine MRI: Prospective Analysis of Interchangeability. *Radiology*. 1. November 2022;212922.
4. Altmann S, Grauhahn N, Brockstedt L, Kondova M, Schmidtmann I, Paul RH, Clifford B, Feiweier T, Hosseini Z, Uphaus, T, Groppa S, Brockmann MA, Othman AE. Ultra-fast brain MRI with deep learning reconstruction for suspected acute ischemic stroke. (under Review at Radiology)

2024-03-01 14:40 - 15:00, AM 2

SES-36: Data Sharing and Reproduciility

Federated Generalized Additive Models for Location, Scale and Shape in DataSHIELD

Swenne, Annika¹; Intemann, Timm¹; Pigeot, Iris^{1,2}

¹Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany

²University of Bremen, Faculty of Mathematics and Computer Science, Germany

Generalized additive models for location, scale and shape (GAMLSS) (1) are flexible non-parametric regression models with a wide range of applications. In particular, they are used for estimating reference percentile curves for clinical biomarkers in children and adolescents. To establish such percentile curves as an international reference, the underlying analysis data set should not only be large but it should also cover a diverse population of children and adolescents. This can only be achieved by pooling data from multiple studies and study centers worldwide. However, due to ethical and legal constraints, physically sharing and pooling sensitive individual-level data might not always be permitted. Thus, a privacy preserving approach has to be used instead.

DataSHIELD (2) is a software infrastructure for R that overcomes these constraints and allows to co-analyze sensitive individual-level data from multiple sources without physically sharing the data. The federated DataSHIELD approach yields identical results as compared to the analysis of physically pooled individual-level data for generalized linear models (3). However, GAMLSS implementation for the joint analysis of federated data is currently not available in DataSHIELD. To close this gap, we developed an algorithm to estimate a single joint GAMLSS model for federated data in DataSHIELD.

This newly developed federated GAMLSS algorithm is based on the original GAMLSS algorithm, consisting of Fisher scoring and backfitting cycles to maximize the penalized likelihood, and works for the case of physically separated data. We prove that the federated GAMLSS yields identical results compared to the original GAMLSS. Furthermore, we provide the first implementation of a GAMLSS model for the joint analysis of federated data in the DataSHIELD infrastructure. We illustrate this implementation by estimating age-dependent height reference curves for children using spline functions and different distributions (e.g. normal, Box-Cox Cole and Green and Box-Cox power exponential distribution) based on two physically separated data sets. Our implementation yields identical results as the original GAMLSS algorithm for the same but physically pooled data sets. Despite these promising results, there are some limitations in terms of computation time, model diagnostics and available spline functions.

References

1. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2005;54(3):507-54.
2. Wilson RC, Butters OW, Avraam D, Baker J, Tedds JA, Turner A, et al. DataSHIELD – new directions and dimensions. *Data Science Journal*. 2017;16(21):1-21.
3. Jones EM, Sheehan NA, Masca N, Wallace SE, Murtagh MJ, Burton PR. DataSHIELD – shared individual-level analysis without sharing the data: a biostatistical perspective. *Norsk Epidemiologi*. 2012;21(2):231-39.

2024-03-01 15:00 - 15:20, AM 2

SES-36: Data Sharing and Reproduciility

The CodeClub at the MPI of Psychiatry – better code and better reproducibility

Hagenberg, Jonas^{1,2,3}; Karlbauer, Vera N.^{1,4}; Dieckmann, Linda^{1,2}

¹Max Planck Institute of Psychiatry, Germany

²International Max Planck Research School for Translational Psychiatry (IMPRS-TP)

³Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg

⁴Graduate School of Systemic Neurosciences, LMU Munich

At the Max Planck Institute of Psychiatry (MPIP) people with different backgrounds and programming skills, from wet-lab scientists to bioinformaticians, work together. Also, many PhD students who have not received formal education in computer science or statistics perform primarily computational work. Therefore, we have founded the CodeClub at the MPIP. We organize and lead monthly meetings, where we have presentations about programming topics, discuss problems encountered in the last month and find partners for code review.

One focus of the CodeClub is open science and tools to improve reproducibility. In the past, we covered topics such as using `renv` to create reproducible R environments, documenting analyses in RMarkdown or how to obtain DOIs from Zenodo for publishing code. As a result of discussions in the CodeClub, we developed a standard analysis template for RStudio projects. The template is freely available at <https://github.com/jonas-hag/analysistemplates> and makes it easy to start an analysis with a standardized and sensible structure.

Based on our experiences with the CodeClub, we present ideas how to structure peer-based programming learning and how to engage with researchers. We emphasize the need to provide hands-on tutorials and to include researchers with different levels of coding experience. Lastly, we discuss our approaches to code review and provide a code review checklist at <https://doi.org/10.5281/zenodo.8432945>.

More information about the CodeClub is available at

<https://ev.campussource.de/publikationen/csa2022/gewinner.html#sonderpreis>.

2024-03-01 15:20 - 15:40, AM 2

SES-36: Data Sharing and Reproducibility

Simple tips for writing and publishing clear code to ensure reproducible results

Hornung, Roman

Institute for Medical Information Processing, Biometry and Epidemiology, LMU Munich, Munich, Germany

Amidst the ongoing efforts towards Open Science, there is widespread recognition of the need to make research results in biomedical and biometrics research reproducible. The most reliable approach to achieving reproducibility is to publicly share the code used to derive the research results. However, to truly ensure reproducibility without substantial effort, factors beyond simply providing access to the code need to be taken into account. Drawing on the Biometrical Journal's reproducibility standards and my personal experiences reviewing code as one of the Reproducible Research Editors at the journal, this talk will offer straightforward advice on how to write and organize code. The goal is to make it readily understandable to individuals without expertise and to facilitate effortless reproducibility.

Key to ensuring reproducibility are practices such as specifying the software used and the versions of the add-on packages, setting the seed for the random number generator, and verifying the executability of the code shortly before publication. For analyses that demand substantial computational resources, sharing intermediate results, such as those from the individual runs of a simulation, is extremely beneficial. This practice enables spot checks for reproducibility and allows changes in the evaluation of results at a later stage without necessitating reanalysis. Fundamental to code clarity is adequate commenting, providing a concise yet informative README file, and maintaining a clear folder structure. Often overlooked but crucial aspects include limiting the number of code files and ensuring that the names of the generated results match those presented in the publication. Finally, the talk will provide guidance on how to publish code to ensure its long-term availability.

By adhering to the advice provided in this talk, researchers will be better equipped to render their results reproducible and their code more accessible and reusable for their community.

2024-03-01 15:40 - 16:00, AM 2

SES-36: Data Sharing and Reproducibility

Addressing researcher degrees of freedom through adjustment for the multiplicity of analysis strategies

Mandl, Maximilian M^{1,2}; Becker-Pennrich, Andrea S^{1,4}; Hinske, Ludwig C^{4,5}; Hoffmann, Sabine^{1,3}; Boulesteix, Anne-Laure^{1,2}

¹Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität München

²Munich Center for Machine Learning

³Department of Statistics, Ludwig-Maximilians-Universität München

⁴Department of Anaesthesiology, LMU University Hospital

⁵Institute for Digital Medicine, University Hospital of Augsburg

In recent years, the scientific community has become aware that there is high analytical variability when analysing empirical data, i.e. there are plenty of sensible ways to analyse the same dataset for addressing a given research question. If this phenomenon is combined with selective reporting, it may lead to an increased rate of overoptimistic results, e.g. – depending on the context – false positive test results, inflation of effect sizes, or exaggerated measures of predictive performance. Hoffmann et al. (1) argued that this variability of results can be explained by six sources of uncertainty that are omnipresent in empirical research regardless of the respective discipline: sampling, measurement, model, parameter, data pre-processing, and method uncertainty. Failure to take these various uncertainties into account may lead to unstable, supposedly precise, but overoptimistic and thus potentially unreplicable results.

Even relatively simple settings may lead to many combinatorial possibilities, e.g. for the model selection in a multivariable regression setting we can come up with 2^p possible specifications - with p being the number of covariates. Also in cases where we only focus on reasonable specifications we can observe a substantial analytical variability due to a combination of different sources of uncertainties (2). In the long run this has devastating consequences on the credibility of research findings, in particular if independent teams of researchers publish contradicting results on the same data set.

In this study, we tackle this issue by viewing it as a multiple testing scenario. We suggest to address researcher degrees of freedom through a multiple testing procedure. As the hypotheses of the different analysis strategies are usually highly correlated, a naive approach such as the Bonferroni correction is inappropriate because it leads to an unacceptable loss of power. Instead, we propose using the “minP” adjustment method (3), which approximates the underlying null distribution through a permutation-based proce-

dure and thus achieves more power than simpler approaches while ensuring a weak control of the family-wise error rate.

We assess the properties of this “minP” procedure for addressing researcher degrees of freedom using simulations and illustrate its use by applying it to a study on the impact of perioperative paO₂ on post-operative complications after neurosurgery. Our results suggest that “minP” adjustment is a practical valid approach that allows the selective reporting of the most favorable result while preventing type 1 error inflation in the presence of analytical uncertainty.

References

- (1) Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix A-L. The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science, 8(4):201925, 2021.
doi: <https://doi.org/10.1098/rsos.201925>.
- (2) Silberzahn R, Uhlmann EL, Martin DP, et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3):337-356, 2018, doi:10.1177/2515245917747646.
- (3) Westfall PH, Young SS, Paul Wright S. On adjusting p-values for multiplicity. Biometrics, 49(3):941–945, 1993. doi: <https://doi.org/10.2307/2532216>.

2024-02-29 13:20 - 13:40, AM S1

SES-40: Machine Learning 1

Achieving explainable machine learning by functional decomposition of black-box models into explainable predictor effects

Köhler, David¹; Rügamer, David²; Schmid, Matthias¹

¹University Bonn, Germany

²Department of Statistics, LMU Munich, Germany

Advances in statistics, mathematics and computing have led to a rise of computationally expensive, high accuracy methods for prediction or modeling. Many of these methods, such as deep neural networks or random forests, often lack interpretability due to their blackbox nature, hindering their applicability in many applications such as healthcare, ecology or economy. Functional decomposition is a well explored tool that improves the interpretability of black-box models by splitting the prediction function into a sum of main and interaction effects. This makes the model understandable by providing an intuitive feature effect visualization. However, implementations of existing methods are often computationally infeasible, especially when analyzing higher dimensional continuous data. Here we present a novel method for deriving a functional decomposition of arbitrary continuous prediction functions inspired by Hooker's generalized Functional ANOVA. This is done by fitting a neural additive model (NAM) with DNN-based main-effect and interaction submodels using the model predictions as outcome variable. The submodels are orthogonalized against higher-order terms to ensure interpretable and identifiable low order feature effects. By having minimal prerequisites on DNN architecture and model fitting, the method can be widely applied without constraining the learning algorithm and model predictive performance. Our empirical results demonstrate the algorithm's ability to correctly identify the shape and size of the contributions of single features, yielding insights into the contribution of features on model predictions.

References

Hooker G. Generalized functional anova diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*. 2007 Sep 1;16(3):709-32.

2024-02-29 13:40 - 14:00, AM S1

SES-40: Machine Learning 1

Red-light crossing or bank robbery? On the bias in model performance estimates resulting from incorrect optimization of algorithm and preprocessing hyperparameters

Sauer, Christina^{1,2}; Hanßum, Luzia¹; Hodiamont, Farina³; Bausewein, Claudia³; Boulesteix, Anne-Laure^{1,2}; Ullmann, Theresa⁴

¹Institute for Medical Information Processing, Biometry and Epidemiology, LMU Munich, Munich, Germany

²Munich Center for Machine Learning (MCML), Munich, Germany

³Department of Palliative Medicine, LMU University Hospital, Munich, Germany

⁴Medical University of Vienna, Center for Medical Data Science, Section for Clinical Biometrics, Vienna, Austria

In recent years, many clinical research fields have experienced an increase in newly proposed prediction models that were developed using supervised machine learning algorithms. Generating such a model usually requires a number of specifications that impact its predictive performance. This includes both the hyperparameters of the algorithm (e.g., for a tree-based algorithm, the minimum number of observations in any terminal node) and the hyperparameters of the preprocessing (e.g., the percentage of outliers to be removed or which variant of a certain feature variable should be used).

In terms of choosing the optimal hyperparameter configuration of the algorithm, a widely used approach is to employ automated tuning procedures based on resampling schemes such as (nested) cross-validation. While this approach would technically also be suitable for optimizing preprocessing hyperparameters, it is rarely implemented in practice. This is probably because the set of preprocessing hyperparameters is often application dependent and automated procedures for tuning preprocessing hyperparameters are thus not readily available in standard software. As a consequence, researchers may configure the preprocessing manually by successively trying different options and choosing the one that yields the best predictive performance. This procedure can lead to a biased performance estimation of the model if it violates the separation between train and test data, i.e., if the preprocessing parameters are tuned on the same dataset that is used to estimate the performance of the model. Compared to the optimization of algorithm hyperparameters, where this issue has been explicitly addressed in literature, researchers who incorrectly select the best preprocessing hyperparameters are more likely to either be unaware of the introduced bias or consider it comparable to crossing a red light: not exactly correct, but only a misdemeanor (Simmons et al., 2018).

Abstracts of Contributed Talks

In our study, we aim to raise awareness for incorrect hyperparameter optimization by illustrating different variants of subtly or substantially flawed procedures to optimize both algorithm and preprocessing hyperparameters and investigate to which extent they lead to biased predictive performance estimates. For this purpose, we consider the task of generating a decision tree to predict the costs of patients receiving palliative care using data from the COMPANION project. By comparing the model performance estimated through the investigated procedures to the performance estimated using a hold-out dataset, we quantify the introduced bias and demonstrate that some procedures, metaphorically speaking, rather correspond to robbing a bank than crossing a red light.

References

- Simmons JP, Nelson LD, Simonsohn U (2018). "False-Positive Citations". Perspectives on Psychological Science, 13(2), 255-259.
<https://doi.org/10.1177/1745691617698146>

2024-02-29 14:00 - 14:20, AM S1

SES-40: Machine Learning 1

Using background knowledge from previous studies in model building: the good, the bad and the ugly

Hafermann, Lorena¹; Heiko, Becher²; Herrmann, Carolin¹; Klein, Nadja³; Kammer, Michael⁴; Rauch, Geraldine¹; Heinze, Georg⁴

¹Charité - Universitätsmedizin Berlin, Germany

²Universitätsklinikum Heidelberg

³Technische Universität Dortmund

⁴Medizinische Universität Wien

The question of which variables to include in a statistical model is of key importance. Thereby, the integration of background knowledge into the modelling process is an issue that is gaining increasing interest. Often background knowledge is based on results from previous studies. Its quality can vary substantially, particular with regard to the methods used to select variables, raising concerns about its utility for model building in a new study. We found that in the context of descriptive and explanatory linear regression models, background knowledge from previous studies that employed inappropriate model building strategies such as univariable selection distorted the specification of an appropriate predictor set. We also investigated if prediction models estimated by Random Forests benefitted from the incorporation of appropriate background knowledge on the set of predictors. Surprisingly, there was no benefit in terms of discrimination, but the calibration of the model improved by well-informed restriction of the candidate predictor set. Therefore, we recommend to use background knowledge with care and to critically appraise the methodological quality of any previous studies from which it was derived.

2024-02-29 14:20 - 14:40, AM S1

SES-40: Machine Learning 1

Random forests more data hungry than logistic regression models? A confirmatory, large-scale, real-data study on the link between the number of events per variable and prediction performance

Lange, F. Julian D.^{1,2}; Boulesteix, Anne-Laure^{1,2}

¹Institute for Medical Information Processing, Biometry, and Epidemiology, LMU Munich, Germany

²Munich Center for Machine Learning (MCML), Munich, Germany

Binary classification is a common task in many applied research fields. In the development of clinical prediction models, a frequently considered concept is the number of events per variable (EPV), which, for binary classification, is defined as the ratio of the number of observations in the minority class of the outcome variable to the number of predictor variables. Previous studies on the needed number of EPV are mostly simulation studies that focus on regression modeling and do not examine prediction performance. In contrast, we present a real-data study that investigates the predictive performance of logistic regression and random forests in relation to the number of EPV and aims to confirm two results from an existing simulation study by van der Ploeg et al. (1). Noteworthily, following best practices from clinical research and other applied disciplines, a comprehensive research protocol was written before the start of the benchmark experiment to increase transparency and ensure the confirmatory nature of the study. In the conducted large-scale benchmark experiment involving 75 datasets, models are trained with data subsets corresponding to different numbers of EPV (ranging from 5 EPV to 500 EPV), and their performance is measured by the area under the receiver operating characteristic curve (AUC) through repeated 5-fold cross-validation. The results of the study indicate strong support for the pre-specified hypotheses and, thus, roughly confirm the findings by van der Ploeg et al. (1). Specifically, it shows that, for the analyzed datasets, untuned random forests (a) require more EPV than logistic regression to realize their predictive performance potential and (b) are prone to overfitting even when generated with a large number of EPV, such as 500 EPV. Therefore, the study provides new confirmatory evidence on the relationship between the number of EPV and the predictive performance of logistic regression and random forests. In the context of research on the topic of EPV in general, it is the first large-scale, real-data benchmark experiment and, thus, provides insights from a different perspective than the many existing simulation studies in this area. Additionally, an extensive sensitivity analysis of the primary results shows that some design and analysis choices had great impact on the study's findings, demonstrating how important the public pre-specification of hypotheses and plans is for studies that are meant to be confirmatory and how pre-registration can prevent selective

reporting and over-optimistic conclusions.

References

- van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014;14:137.

2024-02-29 15:00 - 15:20, AM S1

SES-41: Machine Learning 2

Adversarial random forests for imputing missing values

Golchian, Pegah^{1,2}; Kapar, Jan^{1,2}; Blesch, Kristin^{1,2}; Watson, David S.³; Wright, Marvin N.^{1,2,4}

¹Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany

²University of Bremen

³King's College London

⁴University of Copenhagen

When applying machine learning to real datasets, the problem of missing values often occurs. However, in order to make a statement about a population, it is necessary to fill these values with so-called imputation methods, which are categorized into single and multiple imputation (Van Buuren, 2018). In a statistical setting, multiple imputation is recommended to use, since it accounts for uncertainty. It reaches its limits when the number of variables is large and the sample size is moderate, leading to computational costs and overparameterization (Tang and Ishwaran, 2017). Moreover, it has problems in tackling complex interactions and nonlinearity of variables (Tang and Ishwaran, 2017). Here, machine learning could lead to improvements. Novel machine learning imputation methods, such as MissForest (Stekhoven and Bühlmann, 2012), are convenient for outputting a single complete dataset on the one hand but do not account for uncertainty on the other. We suggest an algorithm based on generative machine learning that offers single and multiple imputation. It is based on the method adversarial random forest (ARF) (Watson et al., 2023) for density estimation and data synthesis. ARF uses a recursive variant of unsupervised random forests (Shi and Horvath, 2006), influenced by the idea of generative adversarial networks (GANs) (Goodfellow et al., 2014). We extend ARF by using conditional sampling on the non-missing values and call the method Miss-ARF. For comparison with other competing imputation methods, we use the realdatalab benchmark OpenML-CC-18 (Bischl et al., 2021) and simulate the three missing data mechanisms missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) for a range of missingness proportions. We compare the imputed dataset with the real dataset by normalized root mean squared error (NRMSE) and proportion of falsely classified entries (PFC). Since we consider a generative modeling setting, we also compare the performance of machine learning models with the imputed dataset and the real dataset by using machine learning efficacy/utility (Choi et al., 2017), which is a suitable measure for synthetic tabular data.

References

Van Buuren S. Flexible imputation of missing data. 2nd ed. CRC press; 2018.

Tang F, Ishwaran H. Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2017;10(6):363-77.

Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-8.

Watson DS, Blesch K, Kapar J, Wright MN. Adversarial random forests for density estimation and generative modeling. In: *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. vol. 206. PMLR; 2023. p. 5357-75.

Shi T, Horvath S. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics*. 2006;15(1):118-38.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. vol. 27; 2014. p. 2672-80.

Bischl B, Casalicchio G, Feurer M, Gijsbers P, Hutter F, Lang M, et al. OpenML benchmarking suites. In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*; 2021. .

Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multilabel discrete patient records using generative adversarial networks. In: *Proceedings of the 2nd Machine Learning for Healthcare Conference*. vol. 68. PMLR; 2017. p. 286-305.

2024-02-29 15:20 - 15:40, AM S1

SES-41: Machine Learning 2

Confidence intervals for tree-structured varying coefficients based on parametric bootstrap

Spuck, Nikolai; Schmid, Matthias; Berger, Moritz

Institute of Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn, Germany

Varying-coefficient models first introduced by Hastie and Tibshirani (1993) extend the class of linear regression models, as they allow that coefficients of covariates change with the value of other variables, the so-called effect modifiers. This model structure assumes a specific type of interaction between covariates and effect modifiers. The tree-structured varying-coefficient model proposed by Berger et al. (2019) applies recursive partitioning techniques to identify relevant effect modifiers such that the varying effects of the covariates are described by tree structures. Specifically, their approach yields a separate tree for the linear effect of each covariate, which is modified by one or several other variables, where each leaf node contains a partition-specific coefficient.

To quantify uncertainty of the estimated partition-specific coefficients, we propose a procedure for constructing confidence intervals. Asymptotic normal distribution-based confidence intervals like Wald intervals are not suitable as they assume that the tree structures determining the varying coefficients are predefined and neglect the statistical uncertainty induced by the data-driven tree building process. Therefore, these confidence intervals tend to be too short. The application of classical bootstrap approaches is further aggravated as the tree structures change with each random sample, which might even strongly deviate from the true data generating process. Consequently, the question of how to determine the true effect of a covariate in a subset of the population arises. Given a fitted tree, we define the true coefficient in a leaf node of the tree as the expectation of the coefficient truncated to the subset of the population that comprises this leaf node. Based on this definition we propose a parametric bootstrap approach for constructing confidence intervals, where in each replication each partition-specific coefficient is calculated as the average coefficient of the subset of observations contained in the corresponding partition.

Coverage probabilities for 95% and 90% confidence intervals were assessed in a simulation study and compared to the coverage probabilities of standard Wald intervals. Preliminary results indicate coverages close to the nominal levels in particular in scenarios with higher error variance. In an application to real-world data, a prediction model with tree-structured varying coefficients was fitted to data from patients with acute odontogenic

infection and confidence intervals for the resulting coefficients were constructed using the proposed approach.

References

- Berger, M., Tutz, G. & Schmid, M. Tree-structured modelling of varying coefficients. *Stat Comput* 29, 217–229 (2019). <https://doi.org/10.1007/s11222-018-9804-8>
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4), 757–796.
<http://www.jstor.org/stable/2345993>

2024-02-29 15:40 - 16:00, AM S1

SES-41: Machine Learning 2

On the handling of method failure in comparison studies

Wünsch, Milena^{1,2}; Boulesteix, Anne-Laure^{1,2}

¹Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, LMU Munich, Munich, Germany

²Munich Center for Machine Learning, Munich, Germany

Neutral comparison studies aim to compare methods in an evidence-based manner and thus to provide reliable guidance to data analysts in the selection of a suitable method for their application. To be of true value, they require methodological researchers to carefully design, implement, and report their analyses, while failure to do so can have tremendous consequences. However, since these three steps are typically subject to a high degree of flexibility, resulting in variability in the results of the comparison, the difficulty lies precisely in their right choice. In this work, we focus on a particular aspect of the design of comparison studies, namely the handling of missing performance values which is often inadequate and, moreover, insufficiently reported. Missing performance values occur when a method fails to output a result for a specific (simulated or real) data set, e.g. due to non-convergence or other computational issues. Simply disregarding these missing values in the manner of complete-case analysis can potentially lead to serious biases in the results. A well-considered approach to this issue is therefore necessary. Based on various examples, we discuss different approaches to handling these missing performance values, assess their impact on the evaluation of the methods, and provide preliminary guidance in this context depending on the study's goal. Our examples include different types of comparison studies (simulation studies and real data-based benchmarking), different types of methods (regression modelling, statistical testing, and machine learning), and different types of data (low- and high-dimensional data). The results suggest that the handling of missing performance values has a substantial impact on the conclusions of comparison studies in many cases and that simple approaches often do not suffice to capture the complexity of the issue.

2024-02-29 16:00 - 16:20, AM S1

SES-41: Machine Learning 2

An empirical study of the performance of semi-supervised machine learning methods in systematic review tools for abstract and title screening

Kutil, Raoul¹; Borgelt, Christian¹; Hirlaender, Simon¹; Zimmermann, Georg²

¹IDA Lab Salzburg, Austria

²PMU Salzburg, Austria

Background: Evidence-based medicine makes decisions based on relevant and up-to-date evidence from systematic research by including the most reliable and relevant evidence like randomized controlled trials, which are subsequently synthesized in systematic reviews. Screening research papers and selecting relevant ones for a systematic review requires substantial effort. If automatic document relevance ranking with software tools could be fully trusted, the workload could be significantly reduced.

Methods: Subsets of 450 labeled documents were ranked with three systematic review tools and a support vector machine as well as combinations of the individual rankings. Negative predictive value (NPV) was used to assess how well irrelevant documents could be excluded if only the documents in the top half of a ranking were considered in order to halven the workload.

Results: Larger training set size slightly increases performance metrics like negative predictive value (NPV). Different ranking tools perform similarly, reaching 86% NPV with the largest training sets. Combining rankings does not improve the NPV.

Conclusions: Systematic review tools can only aid a human in the decision-making procedure and accelerate the process. Most of these tools use traditional methods that are fast, but also limited in recognizing the context in which words and sentences appear in the abstract/document. To obtain better results, one could harness recent advancements in machine learning and natural language processing to develop semi-supervised labeling algorithms.

Abstracts of Contributed Posters

Poster-Session

2024-02-29 11:40 - 12:40, Foyer Audimax

including Poster-Speed-Session from 11:40 - 12:05 in the AM 1 (Audimax)

Comparison of genetic maps from different cattle breeds (Poster ID 01)

Wittenburg, Dörte¹; Ding, Xi¹; Melzer, Nina¹; Schwarzenbacher, Hermann²; Seefried, Franz R.³

¹Research Institute for Farm Animal Biology (FBN), Germany

²ZuchtData GmbH, Austria

³Qualitas AG, Switzerland

The proximity of loci on a genome can be measured in physical (base pairs) or in genetic distance units (Morgan), where the latter is of special importance for animal breeders. It is expected that one crossover event occurs on average during meiosis at one Morgan distance. Hence, the genetic distance between loci allows drawing inferences, for instance, on genetic variation of not yet born progeny and provides valuable information when searching for top breeding animals.

Genetic diversity among cattle breeds and different breeding objectives for meat, dairy or dual purpose require breed-specific genetic maps. We analysed genotype data from seven commercial cattle breeds with sample size ranging from 4,181 to 298,850. Since various assays were used for genotyping the animals, we standardised the data preparation and analysed the data with our pipeline "hsrecombi". We investigated the frequency of paternal recombination events and derived genetic-map coordinates of about 50K SNP markers. Additionally, estimates of recombination rate between intra-chromosomal pairs of markers enabled the localisation of further putatively misplaced markers or regions in the bovine genome assembly ARS-UCD1.2. Estimates of map length varied from 23.99 M to 27.36 M between breeds. Two to 49 misplaced candidates were detected in each breed, mostly overlapping among breeds. Furthermore, a genomewide association study on the number of recombination events among progeny revealed 13 significant SNPs in total. A subset of these hits was located in or near two genes with known impact on recombination activity providing options to counteract genetic erosion in breeding populations in future. To explore recombination activity interactively and to evaluate differences between breeds, we implemented all results in an R Shiny app "CLARITY".

Bayesian borrowing using mixture prior: frequentist operating characteristics (Poster ID 02)

Weru, Vivienn; Calderazzo, Silvia; Wiesenfarth, Manuel; Kopp-Schneider, Annette

Division of Biostatistics, German Cancer Research Center (DKFZ), Germany

In clinical trials with small populations it is sometimes desired to borrow information from external sources to increase efficiency of the current study. Bayesian methods have seen widespread application in this context, due to the ease in incorporating such external information via the specification of informative prior distributions. Various robust Bayesian approaches have been proposed to mitigate the effect of potential heterogeneity between external and current trial information, i.e., prior-data conflict. One such approach is the robust mixture prior (1) 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 additional 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. We assess impact on frequentist operating characteristics (Type 1 error/power, MSE) of various choices with respect to the location, dispersion and prior weight of the robust component, also in relation to different current sample sizes. The results reflect the fact that if the robust component is too vague it can lead to a mixture prior that does not adapt well to prior-data conflict. A less well-known behavior, however, is that, especially with small sample sizes, the unit-information prior can still be quite informative even in presence of significant prior-data conflict and can lead to undesirable operating characteristics like unbounded Type 1 error rate for some choices of the parameters.

References

1. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014 Dec;70(4):1023-32.

A multi-omics differentiation pattern analysis of CCl4-treated mice data (Poster ID 03)

Heiner, Jonas¹; Hengstler, Jan²; Groll, Andreas¹

¹TU Dortmund, Germany

²Leibnitz-Institut für Arbeitsforschung an der TU Dortmund, Germany

Proteins are encoded by an organism's DNA and result from transcription and translation of the gene. In this way, every protein can be uniquely associated to an origin gene, resulting in standalone gene-protein-pairs. Thus, it is a general assumption that RNA expression and protein levels are positively correlated, and thereby are assumed to be affected by a treatment of the organism in a similar way.

The objective of this study is the investigation of divergence from this assumption with the goal to obtain insights of how and why certain divergence patterns occur, and whether there is an underlying mechanism behind these. For this, a real data application comprising a transcriptomics and a proteomics data set originating from mice that were induced carbon tetrachloride (CCl4) as a treatment is conducted.

The analysis of the assumed relation between the two types of omics data can be initialized with a so-called differentiation pattern (DiPa) plot (Nell et al., 2022) and associated analyses. In particular, the treatment to control fold changes of both the RNA and protein data are put in relation and can be clustered into subgroups (DiPa groups) of similar relations. As the value of these fold changes presumably depends on the baseline expression level, this motivates a breakdown of the analysis into quantile groups based on the baseline expression. Proceeding from this, within each quantile group, the distribution of gene-protein-pairs in the DiPa groups is analyzed in order to investigate the respective fold change behavior based on the baseline expressions. For a slightly different perspective, the same analysis is made based on protein baseline expressions and their partitioning in quantile groups in order to better classify the previously gained insights.

Finally, the omics relation is additionally investigated via regression modeling with simultaneous inclusion of the previous grouping mechanisms to consolidate previous findings on the one hand and to improve predictions of relations on the other hand.

References

Nell P, Kattler K, Feuerborn D, Hellwig B, Rieck A, Salhab A, et al. Identification of an FXR-modulated liver-intestine hybrid state in iPSC-derived hepatocyte-like cells. *Journal of Hepatology*. 2022;77(5):1386–98.

Combining recurrent and terminal events into a composite endpoint may be problematic (Poster ID 04)

Liu, Xiaofei; Koch, Armin

Institute for Biostatistics, Hannover Medical School, Germany

Chronic heart failure (CHF) in its different stages is a seriously debilitating disease with a substantial mortality rate. The current European Medicines Agency (EMA) guideline on clinical investigation of medicinal products for the treatment of chronic heart failure (1) recommends either mortality or time to the composite of hospitalisation for worsening heart failure or death (whichever occurs first) as the primary efficacy endpoint in confirmatory CHF trials. Motivated by lower event rates in recent CHF trials, there are proposals to use endpoints accounting for recurrent heart failure hospitalisation (HFH) events (e.g. recurrent HFH events alone or plus death as an additional event) as the primary endpoint to better quantify disease burden and to improve trial efficiency. However, analysis and interpretation of recurrent event endpoints may be complicated by the presence of terminal events (e.g. death), which limit the number of events per subject. In consequence, a lower number of events in one as compared to the other treatment group may simply be the result of an increased mortality in the former treatment group.

Recently, a request for EMA qualification opinion on recurrent event endpoints (2) and the related articles (3-5) were published. In the previous work, the operating characteristics of analysis methods for recurrent event endpoints were investigated. Based on simulations it was shown that in case of a detrimental treatment effect on mortality, analyses of recurrent HFH events alone using the negative binomial (NB) model and the Lin-Wei-Yang-Ying (LWYY) model were associated with an inflated type I error rate, whereas the type I error rate was well-controlled for analyses of recurrent HFH events plus death as an additional event using the same models. Based on these observations, the authors concluded that the type I error rate seems to be controlled for the composite endpoint of recurrent HFH events plus death in settings which are realistic for CHF trials. As these results were based on limited scenarios, we extended the previous simulations and identified situations where the type I error rate is no longer controlled for the analysis of the composite recurrent event endpoint using the LWYY model, despite a neutral effect on HFH and a detrimental effect on death of the experimental treatment, which would be the most concerning situation for regulatory bodies and patients as consumers of licensed drugs. Therefore, depending on clinical settings, the LWYY model should be applied and interpreted with more caution.

References

1. European Medicines Agency. Guideline on clinical investigation of medicinal prod-

- ucts for the treatment of chronic heart failure - Revision 2. 2017.
2. Akacha M, Binkowitz B, Bretz F, Fritsch A, Hougaard P, Jahn A, et al. Request for CHMP Qualification Opinion: Clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses. Recurrent Event Qualification Opinion Consortium. 2018.
3. Schmidli H, Roger JH, Akacha M. Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event. *Stat Biopharm Res.* 2021.
4. Wei J, Mütze T, Jahn-Eimermacher A, Roger J. Properties of Two While-Alive Estimands for Recurrent Events and Their Potential Estimators. *Stat Biopharm Res.* 2021.
5. Fritsch A, Schrömer P, Mendolia F, Mütze T, Jahn-Eimermacher A. Efficiency Comparison of Analysis Methods for Recurrent Event and Time-to-First Event Endpoints in the Presence of Terminal Events—Application to Clinical Trials in Chronic Heart Failure. *Stat Biopharm Res.* 2021.

Quantification of prior impact in terms of effective sample size targeting test decisions (Poster ID 05)

Wiesenfarth, Manuel; Kopp-Schneider, Annette; Calderazzo, Silvia

German Cancer Research Center, Germany

Bayesian methods allow borrowing historical information through prior distributions but domination of the prior information on posterior inference is commonly unwanted. To quantify and communicate prior informativeness, the prior effective sample sizes (ESS) has gained widespread use, equating information provided by a prior to a sample size. The common approach equates the prior ESS to the number of samples in a (hypothetical) historical data set. However, this measure is independent from the newly observed data, and does not capture potential negative impact on posterior inference induced by the prior in case of prior-data conflict.

In previous work (1), we built on Reimherr et al (2) to relate prior information to a number of (virtual) samples from the current data model. Thereby, the number of samples needed to obtain posterior inference under a baseline prior equivalent to that under the prior of interest is quantified. To this end, an inferential target measure of interest needs to be defined and an MSE type measure was investigated.

In this work, we argue that an impact measure should be tailored to the inferential question of interest which is often a test decision in the context of clinical trials. The integrated risk under the 0-k loss (e.g. 3) is a possible target measure in this situation. We illustrate that tailoring the effective sample size to the test decision problem provides desirable behavior, for example by showing a decreasing impact of an informative prior in case the data is clearly in the null or alternative hypothesis. In this case the data may be evidently conclusive such that the prior does not play any role for the test decision.

Furthermore, we illustrate how such ESS measure can be helpful in interpreting and selecting mixture weights in robust mixture priors (4) which adaptively discard prior information in case of prior-data conflict. It is well known that the rate at which information is discarded depends on both the mixture weight and the dispersion of the vague component. This makes interpretation of the mixture weight difficult creating needs for an intuitive impact measure for such priors.

An R implementation for the beta-binomial and normal-normal model is provided.

References

- 1 Wiesenfarth, M., & Calderazzo, S. (2020). Quantification of prior impact in terms

- of effective current sample size. *Biometrics*, 76(1), 326-336.
- 2 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.
- 3 Calderazzo, S., Wiesenfarth, M., & Kopp-Schneider, A. (2022). A decision-theoretic approach to Bayesian clinical trial design and evaluation of robustness to prior-data conflict. *Biostatistics*, 23(1), 328-344.
- 4 Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4), 1023-1032.

BACE2 polymorphisms are associated with memory impairment in a general population cohort SHIP-TREND (Poster ID 06)

Bonk, Sarah¹; Kirchner, Kevin¹; Garvert, Linda¹; Völzke, Henry^{2,3}; Grabe, Hans Jörgen^{1,4}; Van der Auwera, Sandra^{1,4}

¹Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany

²DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, Greifswald, Germany

³Institute for Community Medicine, University Medicine Greifswald, Germany

⁴German Center for Neurodegenerative Diseases DZNE, Site Rostock/ Greifswald, Germany

Dementia affects millions of predominantly elderly individuals worldwide and Alzheimer disease (AD) represents the most common form. Based on the amyloid cascade hypothesis neurodegeneration in AD is caused by abnormal accumulation of amyloid beta (A) plaques in the brain parenchyma (Barage und Sonawane 2015). Meanwhile, additional components related to the central amyloid cascade hypothesis have been discovered, and one of these is a protease called -site APP cleavage enzyme 2 (BACE2). In contrast to the related BACE1, BACE2 is more considered as a neuroprotective candidate for AD (Yeap et al. 2023). As BACE2 is located on chromosome 21, it is also often studied in the context of Down Syndrome.

Motivated by these neuroprotective findings, we investigated the effect of BACE2 Single Nucleotide Polymorphisms (SNPs) on cognition in a population based cohort that is not just restricted to AD or Down Syndrome cases; the second cohort of the Study of Health in Pomerania (SHIP-TREND, Völzke et al. 2022). We used data from 3700 individuals from which genetic information and results from the Nuremberg Age Inventory (NAI) memory test are available. The analysis was adjusted for different lifestyle parameters and for current depressive symptoms (PHQ-9). We further investigated a possible APOE4 carrier status stratification as this is mentioned in the literature (Huentelman et al. 2019).

Summarizing our results, we found BACE2 SNPs that have an effect on immediate- and delayed memory performance in a population-based cohort. Which particularly stress that this effect is not restricted to AD cases and Down Syndrome.

References

Barage, Sagar H.; Sonawane, Kailas D. (2015): Amyloid cascade hypothesis: Pathogenesis and therapeutic strategies in Alzheimer's disease. In: *Neuropeptides* 52, S. 1–18. DOI: 10.1016/j.npep.2015.06.008.

Huentelman, Matt; Both, Matthew de; Jepsen, Wayne; Piras, Ignazio S.; Talboom, Joshua S.; Willeman, Mari et al. (2019): Common BACE2 Polymorphisms are Associated with Altered Risk for Alzheimer's Disease and CSF Amyloid Biomarkers in APOE 4 Non-Carriers. In: *Scientific reports* 9 (1), S. 9640. DOI: 10.1038/s41598-019-45896-4.

Völzke, Henry; Schössow, Janka; Schmidt, Carsten Oliver; Jürgens, Clemens; Richter, Adrian; Werner, André et al. (2022): Cohort Profile Update: The Study of Health in Pomerania (SHIP). In: *International journal of epidemiology* 51 (6), e372-e383. DOI: 10.1093/ije/dyac034.

Yeap, Yee Jie; Kandiah, Nagaendran; Nizetic, Dean; Lim, Kah-Leong (2023): BACE2: A Promising Neuroprotective Candidate for Alzheimer's Disease. In: *Journal of Alzheimer's disease : JAD* 94 (s1), S159-S171. DOI: 10.3233/JAD-220867.

Paradoxes of Inter-Rater Reliability Measures of skewed ordinal Data (Poster ID 07)

Mönch, Maximilian¹; Grittner, Ulrike¹; Unger, Nina²; Keller, Theresa¹; Breitenstein, Caterina³; Schulze, Daniel¹; Pigorsch, Mareen¹

¹Universitätsmedizin Charité Berlin, Germany

²Universitätsmedizin Greifswald, Germany

³Universitätsklinikum Münster, Germany

Based on the known paradox of commonly used reliability coefficients, especially Krippendorf's alpha, which yield biased estimates in the case of unbalanced data but yet high agreement, we evaluate the performance of different coefficients in different scenarios motivated by practical experience.

From this practical point of view, we include the consideration of similar but not identical ratings, e.g. given three different categories Rater 1 chooses category 3 and Rater 2 chooses category 2 . With this extended definition of accuracy allowing for slight but not critical deviation, the known paradoxes for Krippendorf's Alpha are more likely to occur. Thus, not only we can consider agreement but also divergence by chance and lower distinguishability of categories.

We simulate this by categorization of bivariate normal distributed random variables and given correlation as well as forcing high agreement situations with fixed marginal distributions of multinomial random variables.

Further, we consider scenarios with three, four and five rating categories as well as a focussed view on agreement rates between 90% and 99%. Our results aim to further the issue of finding adequate measures of reliability for ordinal data depending on the distribution and true agreement between raters. If feasible, we provide further suggestions on thresholds of agreement and skewness for using measures like Krippendorf's alpha, Gwet's AC2 or Brennan-Prediger's corrected kappa.

References

- (1) Tran D, Dolgun A, Demirhan H. Weighted inter-rater agreement measures for ordinal outcomes. *Commun Stat Simul Comput.* 2020;49(4): 989-1003.
- (2) Vach W, Gerke O. Gwet's AC1 is not a substitute for Cohen's kappa—A comparison of basic properties. *MethodsX.* 2023;10: 102212.
- (3) Yilmaz AE, Demirhan H. Weighted kappa measures for ordinal multi-class classification performance. *Appl Soft Comput.* 2023;134(C): 110020.

- (4) Yilmaz AE. Inter-rater agreement and adjusted overall degree of distinguishability for ordered categories. *Chil J Stat.* 2022; 13(2): 165–86.

Performance of different Interpolation Methods on self-reported symptoms in the context of Digital Allergology (Poster ID 08)

Hernandez-Toro, Camilo Jose^{1,2}; Grittner, Ulrike²; Caminiti, Lucia³; Charpin, Denis⁴; Delgado, Luís^{5,6,7}; Dramburg, Stephanie¹; Kalpaklioglu, Fusun⁸; Nieto, Antonio⁹; Papadopoulos, Nikolaos G.^{10,11}; Pelosi, Simone¹²; Potapova, Ekaterina¹; Priftanji, Alfred¹³; Travaglini, Alessandro^{14,15}; Tripodi, Salvatore^{16,17}; Matricardi, Paolo Maria¹

¹Department of Pediatric Respiratory Care, Immunology and Intensive Care Medicine, Charité –Universitätsmedizin Berlin, Berlin, Germany.

²Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin Berlin, Berlin, Germany.

³Department of Pediatrics- Allergy Unit, University of Messina, Messina, Italy.

⁴Department of Pneumonology and Allergy, La Timone Hospital, APHM, Aix-Marseille University, Marseille, France.

⁵Basic and Clinical Immunology Unit, Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal.

⁶CINTESIS@RISE, MEDCIDS, Faculty of Medicine of the University of Porto, Porto, Portugal

⁷Allergy Unit, Instituto & Hospital CUF Porto, Porto, Portugal

⁸Kırıkkale University School of Medicine, Department of Immunology and Allergic Diseases, Turkey

⁹Pediatric Allergy and Pneumology Unit, Children's Hospital La Fe Health Research Institute La Fe, Valencia, Spain

¹⁰Allergy Department, 2nd Pediatric Clinic, Athens General Children's Hospital ""P&A Kyriakou"", University of Athens, Athens, Greece

¹¹Division of Infection, Immunity & Respiratory Medicine, Royal Manchester Children's Hospital, University of Manchester, Manchester, UK

¹²TPS Production srl, Rome, Italy

¹³Department of Allergology and Clinical Immunology. UHC Mother Teresa, Medical University Tirana, Tirana, Albania

¹⁴Department of Biology, Tor Vergata University, Rome, Italy

¹⁵Italian Aerobiology Monitoring Network - Italian Aerobiology Association, Rome, Italy

¹⁶Pediatric Allergy Unit, Sandro Pertini Hospital, Rome, Italy

¹⁷Allergology Service, Policlinico Casilino, Rome, Italy

The application of digital health technologies in allergology has gained importance in recent years. One example is frequent monitoring of allergic symptoms (e.g., daily), especially in seasonal allergic rhinitis (SAR). Frequent symptom reporting might assist the identification of the eliciting allergen and help managing symptoms. Nowadays multiple digital platforms (e.g., Mask-Air(1), Patient Hay-Fever Diary(2), AllergyMonitor(3)) allow SAR patients to record their own symptoms. A common problem in data collected

through digital Apps is the high prevalence of missing values since analysis of those data might be biased.

Here we used data from a multicenter study, the @IT.2020 multicenter study, in which symptoms of SAR patients across Southern Europe were collected daily via the app AllergyMonitor (3, 4). Using data from a subgroup of patients with high adherence to symptom reporting and by artificially creating missing values (missing completely at random, MCAR) the performance of different interpolation methods was evaluated. Two simulation approaches were followed: i) missing values were generated in each segment using a binomial random variable with specific probability of missing values ($pmiss=0.1, 0.2, 0.3, 0.4, 0.5$) ("simple approach"); ii) only stretches of consecutive missing values ("gaps") of fixed lengths ($lgap=[1,10]$) were allowed ("fixed gap length approach"). After the generation of the simulated segments, the amount of bias inserted by the interpolation was estimated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

Missing values were uniformly distributed across simulated segments, with a slight under/overrepresentation of these in the start/end, respectively, of the fixed gap length simulated segments. The fixed gap length constraint lead to differences between the real and expected proportions of missing values ($pmiss$), where larger $lgap$ resulted in larger deviations from the expected proportion of missing values.

In general, larger $pmiss$ and larger $lgap$ resulted in larger bias and a larger variability in the bias measures. The LOCF (last observation carried forward) imputation resulted in slightly larger bias than other interpolation methods. In addition, the nature of the best and worst performing patient reporting periods was studied. Non-surprisingly, the best performing reporting periods were less variable than the worst performing reporting periods.

References

1. Bousquet J, Anto JM, Bachert C, Haahtela T, Zuberbier T, Czarlewski W, et al. ARIA digital anamorphosis: digital transformation of health and care in airway diseases from research to practice. *Allergy*. 2021;76(1):168-90.
2. Berger UE, Jäger S, Bergmann K-C. PHD, the electronic patient's hayfever diary. Eur Respiratory Soc; 2011.
3. Tripodi S, Giannone A, Sfika I, Pelosi S, Dramburg S, Bianchi A, et al. Digital technologies for an improved management of respiratory allergic diseases: 10 years of clinical studies using an online platform for patients and physicians. *Italian Journal of Pediatrics*. 2020;46:1-11.
4. Lipp T, Acar Şahin A, Aggelidis X, Arasi S, Barbalace A, Bourgoin A, et al. Heterogeneity of pollen food allergy syndrome in seven Southern European countries: The@ IT.

Abstracts of Contributed Posters

2020 multicenter study. Allergy. 2021;76(10):3041-52.

A MAP prior approach for piecewise constant hazards and competing risks (Poster ID 09)

Stemke, Alexander^{1,2}; Sailer, Oliver²

¹UlM University, Germany

²Boehringer Ingelheim GmbH & Co. KG, Germany

Historical information can be included in a clinical trial data analysis in order to improve efficiency. Historical information on time-to-event outcomes can be leveraged by using a meta analytic predictive (MAP) prior approach (1) for piecewise constant hazards (2).

Here we are interested in the hazards and probabilities of adverse events in a clinical trial for a setting with piecewise constant hazards. In a competing event setting, the adverse event of interest might not be observed due to study discontinuation. If this is not taken into account, the estimation of survival probabilities will be biased. To avoid such bias the MAP prior will be extended to account for cause-specific hazards.

Using counting process likelihood, we allow for general censoring schemes such as event driven trials (3), but do not require a common Poisson assumption (e.g. 4) that conditions on follow-up times.

We simulate datasets with staggered study entry and censoring after a fixed number of observed events. We look into a setting with two competing risks. Different scenarios are chosen to examine operational characteristics of the approach.

Moreover, the approach is demonstrated using historical data from idiopathic pulmonary fibrosis trials to enhance the safety analysis for a recent trial and be compared to an analysis ignoring the availability of historical data. The data is provided by Boehringer Ingelheim Pharma GmbH & Co. KG.

References

- (1) Heinz Schmidli, Sandro Gsteiger, Satrajit Roychoudhury, Anthony O'Hagan, David Spiegelhalter, and Beat Neuenschwander (2014). "Robust meta-analytic-predictive priors in clinical trials with historical control information". In: Biometrics 70.4, pp. 1023–1032.
- (2) Satrajit Roychoudhury, and Beat Neuenschwander (2020). "Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint". In: Statistics in Medicine 39.7, pp. 984–995.

Abstracts of Contributed Posters

- (3) Jasmin Rühl, Jan Beyersmann, and Sarah Friedrich (2022). “General Independent Censoring in event-driven trials with staggered entry”. In: Biometrics 79.3, pp. 1737–1748.
- (4) Ming-Hui Chen, Mário de Castro, Miaomiao Ge, Yuanye Zhang (2013). “Bayesian Regression Models for Competing Risks”. In: Handbook of Survival Analysis, pp. 179–198.

Using early or baseline data in a trial with missingness in a continuous primary endpoint (Poster ID 10)

Basu, Joydeep; Stallard, Nigel

University of Warwick, United Kingdom

In interrupted trials, we often encounter missingness of observations in a monotone way, such that we lose data for some patients in the follow-up stage. We consider a special case of this problem in which we have primary outcomes missing for some patients for whom early outcome or baseline covariate data are available. Galbraith and Marschner (2003) show how to improve the precision of an estimator for the mean of the primary endpoint by constructing the MLE utilising all available observations from each follow-up stage under the multivariate normality assumption of the observation vector for each patient. Van Lancker et al. (2020) discuss how the precision of treatment effect can be improved by incorporating baseline covariates and short-term endpoints.

In this work, we propose an imputation procedure by fitting a regression line assuming the conditional distribution of primary endpoint given baseline as normal. This allows us to achieve flexibility over the constraint of multivariate normal assumption of the observation vector and it gives the same result in spite of relaxed assumptions.

We perform Monte-Carlo simulations to compare our approach with that of Galbraith and Marschner. Limitations and further possible extensions of the method are also discussed.

References

- [1] Galbraith, S., Marschner, I.C.; Interim analysis of continuous long-term endpoints in clinical trials with longitudinal outcomes; *Statistics in Medicine*, 22, 2003, 1787-1805.
- [2] Van Lancker, K., Vandebosch, A., Vansteelandt, S.; Improving interim decisions in randomized trials by exploiting information on short-term endpoints and prognostic baseline covariates; *Pharmaceutical Statistics*, 19, 2020, 583-601.

Simultaneous inference of multiple binary endpoints in biomedical research: small sample properties of multiple marginal models and a resampling approach (Poster ID 11)

Budig, Sören; Schaarschmidt, Frank

Leibniz Universität Hannover, Germany

In biomedical research, the simultaneous inference of multiple binary endpoints may be of interest. This type of data occurs, for example, in toxicology in carcinogenicity studies, where it is investigated whether an increasing dose of a chemical affects the occurrence of tumours. Or when testing new drugs, the occurrence of several types of adverse events can be examined. In such cases, an appropriate multiplicity adjustment is required, that controls the family-wise error rate, which represents the probability of making incorrect test decisions. In our study, we investigate two approaches that perform single-step p-value adjustments that also take into account the possible correlation between endpoints. A rather novel and flexible approach known as multiple marginal models is considered, which is based on stacking the parameter estimates of the marginal models and deriving their joint asymptotic distribution. We also investigate a non-parametric vector-based resampling approach and we compare both approaches with the Bonferroni method by examining the family-wise error rate and power for different parameter settings, including low proportions and small sample sizes. The results show that the resampling-based approach consistently outperforms the other methods in terms of power, while still controlling the family-wise error rate. The multiple marginal models approach, on the other hand, shows a more conservative behaviour. However, it offers more versatility in application, allowing for more complex models or straightforward computation of simultaneous confidence intervals. The practical application of the methods is demonstrated using a toxicological dataset from the NTP.

Kirstine.jl: A Julia Package for Bayesian Optimal Design of Experiments (Poster ID 12)

Sandig, Ludger

Technische Universität Dortmund, Deutschland

With nonlinear regression models, good experimental design is important for obtaining precise parameter estimates, but existing software packages for finding designs are lacking either speed or flexibility. Due to the nonlinearity, it is necessary to specify some prior knowledge about the unknown model parameters. Specifying a full distribution results in more robust designs than giving only a single best guess, but evaluating the objective function then requires integrating a functional on information matrices numerically. This makes the objective function computationally expensive to maximize, especially in programming languages such as R, where function arguments are passed by value. Computing the thousands of information matrices, that are required for even just one evaluation of the objective function, in this manner is a performance bottleneck. Therefore, existing design packages on CRAN implement the performance critical parts in C or C++, but only for a few specific regression models, which limits their flexibility. While some packages allow arbitrary user-supplied models, these are typically implemented in R, which limits their speed. Hence a package is missing in the design software ecosystem that allows finding Bayesian optimal designs for arbitrary nonlinear regression models efficiently, preferably without having to write code in two different languages.

Kirstine.jl is a Julia package for nonlinear Bayesian design that attempts to be both flexible and efficient. Julia is a dynamic scientific programming language with a user-extensible type system and multiple dispatch. This makes it possible to write both concise high-level and efficient low-level code in the same language. Users of Kirstine.jl can specify a model by defining model and parameter subtypes and implementing methods for functions that dispatch on these. Since Julia passes function arguments by reference, the package-internal code can then work with pre-allocated matrices and pass them to efficient BLAS/LAPACK routines. This setup provides flexibility without sacrificing performance. The package currently supports the D- and A-criterion, vector-valued observations, particle-swarm optimization, and box-shaped design regions of arbitrary dimensions. Locally optimal design is supported implicitly as well. The package is written in a modular way such that users can add drop-in extensions, e.g. for different design criteria or regions, without having to modify the package internals. Kirstine.jl is distributed under the GPL-3.0 license as free and open source software.

cases - an R package for simultaneous evaluation of multiple diagnostic tests or prediction models regarding co-primary endpoints sensitivity and specificity (Poster ID 13)

Westphal, Max¹; Zapf, Antonia²

¹Fraunhofer Institute for Digital Medicine MEVIS

²Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf

In diagnostic accuracy studies and evaluation studies of binary prediction models, sensitivity and specificity are often considered as co-primary endpoints. This implies that a (superiority or non-inferiority) claim has to be demonstrated regarding both endpoints but also that no adjustment for multiplicity is needed (intersection-union test) (1).

However, when multiple candidate tests or models are evaluated on the same dataset (union-intersection test), an adjustment for multiplicity is required to avoid an inflation of the family-wise error rate (FWER). Due to the particular co-primary endpoint scenario, readily available multiple comparison procedures (MCPs) are too conservative and need to be adapted. In a recent work, five applicable approaches for this scenario have been compared in a simulation study (2).

These MCPs have all been implemented in a new R package cases that is available on CRAN (3). The package functionality includes calculation of multiplicity adjusted p-values and test decisions. In addition, results can be visualized as (adjusted) rectangular comparison regions, a relatively novel technique to quantify uncertainty in the specific co-primary endpoints context (2, 4). Calculation of adjusted classical confidence intervals is also possible. In this poster presentation, we will illustrate the overall package functionality by providing exemplary code and output of a real-world data analysis.

References

- (1) European Medicines Agency, CHMP. Guideline on clinical evaluation of diagnostic agents, 2010.
- (2) Westphal M, Zapf A. Statistical inference for diagnostic test accuracy studies with multiple comparisons. arXiv [stat.ME]. 2021. Available from: <http://arxiv.org/abs/2105.13469>
- (3) Westphal M. cases: Stratified Evaluation of Subgroup Classification Accuracy. R package, 2023. Available from: <https://CRAN.R-project.org/package=cases>
- (4) Eckert M, Vach W. On the use of comparison regions in visualizing stochastic uncertainty in some two-parameter estimation problems. Biom J. 2020;62(3):598–609. Available from: <http://dx.doi.org/10.1002/bimj.201800232>

mldesign - an R package to conduct meaningful data splitting in applied machine learning (Poster ID 14)

Westphal, Max

Fraunhofer Institute for Digital Medicine MEVIS

Data splitting is of utmost importance in applied machine learning (ML) projects, primarily to avoid overfitting and over-optimism. Frequently employed techniques (hold-out, (nested) cross-validation, bootstrapping) usually rely on random splitting of the available data (1, 2). In addition, several (grouped, stratified) variations of standard procedures and further special approaches (e.g. for time-series data) do exist.

In effect, ML practitioners often have a hard time choosing the most appropriate data splitting method for their specific ML task. In a recent work, we connect this problem to the definition of an adequate estimand (3). Besides important considerations of prediction task, intended use, prediction model (pipeline), performance metrics and intercurrent events, our framework primarily focuses on a specification of constraints (inclusion- or exclusion criteria) on (a) test observations, (b) training data and (c) the relation between the two. These constraints describe what kind of generalization or transferability (out-of-distribution performance) will be estimated by the ML experiment.

In this work, we focus on the algorithmic implementation of the estimand framework in the R package mldesign. We illustrate that our main algorithm is capable to translate any estimand specification (in the above sense) into a concrete data split, given that such a split is possible for the available dataset. By recursion, nested data splits are also easily possible. In practice, unless vast amounts of data are available, our framework may result in impractical splits (e.g. training sets that are too small). This can be resolved either by relaxing the estimand definition (i.e. enforce fewer constraints) or to couple the outer estimand-based split with classical techniques (e.g. a random hold-out set) for the inner data split. We illustrate these trade-offs and the overall package functionalities by training and evaluating ML models on the International Stroke Trial dataset in different ways. In the future, we will focus on runtime optimization and to allow exporting of results for direct compatibility with frequently used ML libraries.

References

- (1) Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. arXiv [cs.LG]. 2018. Available from: <http://arxiv.org/abs/1811.12808>
- (2) Borra S, Di Ciaccio A. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Comput Stat Data Anal. 2010;54(12):2976–89. Available from:

Abstracts of Contributed Posters

- <https://www.sciencedirect.com/science/article/pii/S0167947310001064>
- (3) Alpers R, Westphal, M. An estimand framework to guide model and algorithm evaluation in predictive modelling. Manuscript in preparation. 2023.
- (4) Westphal M. mldesign: Meaningful Data Splitting in Applied Machine Learning. R package. Available from: <https://github.com/maxwestphal/mldesign>

Multivariate modelling of water quality parameters in nigeria (Poster ID 15)

Dosumu, Ebun Adegbola

Bowling Green State University, United States of America

Inadequate safe water remains a threat to human health in Ibadan Metropolis of Oyo State, Nigeria. The Asejire reservoir is considered a major source of water in Ibadan and water pollution is pronounced as a result of growing population and urbanization in the area it serves. In this study, We studied and analysed water quality parameters of treated and untreated water samples from Asejire Reservoir of Oyo state, Nigeria. We examined the conformance of the sample parameters to the W.H.O 2017 standards for safe and acceptable drinking water parameters. Data were obtained from the Oyo State water corporation which contained values on water quality parameters such as turbidity, colour, pH, alkalinity, etc. Correlations between values of the pollutants were examined for collinearity before estimating linear discriminant functions (LD) that helped to classify water samples into any of "safe and acceptable", "safe but unacceptable", "unsafe yet acceptable" and "unsafe and unacceptable". The estimated discriminant function had a efficiency of up to 84.3% in correctly predicting the class of water sample based on "appearance", "colour", "turbidity", "taste and odour", "alkalinity", "hardness", "chloride", "iron", "manganese" and "total dissolved solids". Results also showed that of all the 41 water samples which were regarded as final and ready for distribution; 7.3% were potentially unsafe and unacceptable for consumption as they were either polluted with iron or dissolved solids from domestic, industrial, agricultural wastes, insecticides and pesticides.

A unified parametric approach to the estimation of dependence and marginal distributions in bivariate competing risks survival data (Poster ID 16)

Zhang, Hyun-Soo; Jung, Inkyung; Nam, Chung Mo

College of Medicine / Yonsei University, Korea, Republic of (South Korea)

In bivariate competing risks survival data where only the minimum of the time-to-events is observed and never both, dependence between the survival endpoints is known to be non-identifiable. If dependence or correlation exists between the time-to-events, cause-specific hazards analysis under independent censoring or inference under incorrectly assumed correlations become biased. Arguably, the most important parameter for estimation when dependence exists is the correlation between the time-to-events. However, maximum likelihood estimation (MLE) is known to be biased with large variance, and no practical methods to estimate the correlation exist. [1]

Using the fact that bivariate normally (BVN) distributed competing risks data is identifiable, [2] we propose a unified parametric approach where the bivariate central limit theorem provides a connection between a given bivariate competing risks data and the identifiable BVN distribution. We demonstrate that the correlation in the given data is estimable by finding a BVN distribution that produces the same sample mean information as that of the given data. Estimating the correlation subsequently enables an unbiased estimation of the marginal survival or hazard functions of the event of interest. Simulations showed that the proposed method works well over various marginal distributions, copulas, and sizes of the correlation.

Our study provides a potential contribution to the existing literature in that the proposed method is applicable to any parametric bivariate competing risks data, requires no covariate information to estimate the correlation, and shows accurate and precise results where the conventional MLE fails to do so. [3] We expect the current study to have further applications in biomedical time-to-event analyses where dependence between the survival endpoints exist, such as in disease etiology research or clinical trials of drug efficacy.

References

1. Basu AP, Ghosh JK. Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis* 1978;8(3):413-429.
2. Chen Y-H. Semiparametric Marginal Regression Analysis for Dependent Competing Risks Under an Assumed Copula. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2010;72(2):235-251.

3. Czado C, Van Keilegom, I. Dependent censoring based on parametric copulas. Biometrika 2022;110(3):721-738.

Sample size re-examination for clinical trials with survival endpoints (Poster ID 17)

Dormuth, Ina¹; Liu, Tiantian²; Chen, Zijian³; Ditzhaus, Marc⁴; Pauly, Markus^{1,5}; Xu, Jin⁶

¹TU Dortmund, Germany

²GlaxoSmithKline R&D Company Limited, China

³Boston University, United States

⁴Otto von Goericke University, Germany

⁵Research Center Trustworthy Data Science and Security, Germany

⁶East China Normal University, China

In design of clinical trials with survival endpoints, the determination of sample size plays a crucial role. The widely employed Schoenfeld formula [1] for the log-rank test necessitates assumptions about hazard rates of the comparing survival functions. In particular, it explicitly defines a presumed hazard ratio and implicitly assumes proportional hazards. The former assumption regarding the magnitude of hazard rates is often associated with a considerable amount of uncertainty. And the latter assumption is often violated, especially in the context of immunotherapy treatments [2].

In light of these considerations, we first survey alternative approaches for sample size calculations, specifically tailored for late effect situation or omnibus tests. Secondly, our study examines the reported sample size calculation based on reconstructed data from a randomly selected set of publications. Higher power than originally intended reveals a possibly overpowered design due to mis-specification of parameters or invalid assumptions. In some cases, a portion of sample size could have been saved if correct specifications and appropriate testing methods were adopted. Moreover, we find sample size re-examination through interim analysis can serve as an effective tool to mitigate the risk of mis-specification at the planning stage.

At last, we offer general guidance on what to take into account when estimating sample sizes for a clinical trial. A free online shiny app will be provided to facilitate sample size calculation for the presented approaches.

References

- [1] Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, 499-503.
- [2] Wu, J., & Wei, J. (2022). Cancer immunotherapy trial design with random delayed treatment effect and cure rate. *Statistics in Medicine*, 41(4), 786-797.
- [3] Guyot, P., Ades, A. E., Ouwens, M. J., & Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival

curves. BMC medical research methodology, 12, 1-13.

Improvement in population-based survival in cutaneous malignant melanoma after the introduction of new therapies (Poster ID 18)

Eisemann, Nora¹; Schumann, Laura¹; Baltus, Hannah¹; Labohm, Louisa¹; Kraywinkel, Klaus²; Katalinic, Alexander¹

¹Institute for Social Medicine and Epidemiology, University of Luebeck, Germany

²German Centre for Cancer Registry Data, Robert Koch Institute, Berlin

Background: In Germany, about 25,000 people are diagnosed with malignant melanoma (MM) each year, and 3,000 people die from this disease. Five-year relative survival (5Y-RS) is highly dependent on the tumour stage, exceeding 100% in UICC I and only about 35% in UICC IV. With the introduction of novel immunotherapies with checkpoint and PD-1 inhibitors, new treatment options specifically for prognostically unfavourable MM are available since 2011. Randomised controlled trials report improvements in survival, but population-level studies are lacking.

Methods: From data of the Centre of Cancer Registry Data, all individuals with MM diagnosis (ICD10: C43) in 2000-19 were included. Individuals who were diagnosed before age 15, lacked a date of death, or were registered based on death certificate only were excluded. 5Y-RS was calculated for four five-year periods (2000-04, 2005-09, 2010-14, 2015-19). To correct for temporal and regional differences, standardisation/stratification was performed for sex, age group and UICC stage. Modelling of melanoma-related excess mortality (adjusted Poisson regression) was used for capturing the significance of a time trend.

Results: We included 301,486 patients. The 5Y-RS improved from 93% (2000-04) to 95% (2015-19). For all subgroups, survival improved or remained similar from 2000-04 to 2015-19. The largest increases were in advanced melanoma between 2010-14 and 2015-19. In UICC IV, 5Y-RS increased from 31% to 36% (+15%). The time trend was increasing and significant ($p<0.001$).

Conclusion: Over the past 20 years, survival of MM patients improved. The largest improvements were seen for advanced tumour stages from 2010-14 to 2015-19. This development fits well with the introduction of the new immunotherapies.

"Initiative Biokybernetik" - ten years later (Poster ID 19)

Mau, Jochen

Heinrich Heine Universität, Germany

IBS Objective "Development of Mathematical Theory in Biosciences" I report on my "Initiative Biokybernetik", announced at 60th Biometric Colloquium in Bremen, suggesting to model human body as a whole mathematically from a systems and automation engineering perspective, integrating medicine, psychology, and other health-factor disciplines.

Structured evaluation of drug prescription data from Schleswig-Holstein in a networked big data context (Poster ID 20)

Schuster, Reinhard¹; Emcke, Timo²; Burmester, Mareike¹

¹Medical Advisory Bord, Germany

²Association of Statutory Health Insurance Physicians

Drug prescription data with a sufficiently high structural depth has the potential to both improve treatment options for patients and to enable meaningful key points for negotiations between the contractual partners with controlling options. Individualized advice offers to resident doctors by pharmacists from the health insurance companies and consultants from the Association of Statutory Health Insurance Physicians have not only achieved savings for many years, but also addressed qualitative improvements for patients. All aspects of data protection have a high priority.

For over 10 years, all drug prescription data of statutory health insured people from Schleswig-Holstein have been analyzed pseudonymously at the Biometric Center North at the Medical Advisory bord of the Statutory Health Insurance North. The pharmaceutical central number as a unique identification of the medicines (there are over 500 thousand of them in Germany, more than in any other EU country) is the key to connecting medicine databases that provide precise information for a structured evaluation.

Diagnoses according to the international ICD classification are relevant for drug treatments. In terms of database technology, matching is possible using an anonymized patient ID. A particular problem arises from the fact that the recorded diagnostic data basically only has a quarterly resolution, while the medication prescription data shows an exact date of prescription and dispensing by the pharmacy.

The patient-related distribution functions in relation to the ATC and ICD classification follow a gamma function with two parameters overall and in relation to extensive filters. In this context, the Channon entropy also plays an important role for further result classification.

With regard to a current structured evaluation, the impact of the corona pandemic plays an important role. Special patient groups with a specific additional treatment need, such as psychiatric problems, are of particular importance. Limited shift effects must be distinguished from general changes. A similar problem arises in antibiotic therapy in terms of benefits and risks.

Graph theoretical models in the dimensions of drug groups, patient groups and doctors provide important impetus both in terms of visualization and the determination of relevant

structural parameters. Mathematica from Wolfram Research is very helpful.

In summary, it can be stated that drug prescription data in the networked big data context is highly relevant both for scientific evaluation and for the design of innovative positions in negotiations between contractual partners and thus for the benefit for patients.

References

- Dormann, H., Sonst, F., Vogler, R., Patapovas, A., Pfistermeister, B., Plank-Kiegle, B., Kirchner, M., Hartmann, N., Burkle, T. Maas, R. (2013). Adverse drug events in older patients admitted as an emergency: The role of potentially inappropriate medication in elderly people (priscus). *Dtsch. Arztebl. Int.*, 110:213–219.
- Emcke, T., Ostermann, T., v. Arnstedt, E., Heidbreder, M. (2017). Comparison of Different Implementations of a Process Limiting Pharmaceutical Expenditures Required by German Law. *Healthinf* 2017.
- Fortin, M., Hudon, C., Haggerty, J., Akker, M., Almirall, J. (2010). Prevalence estimates of multimorbidity: a comparative study of two sources. *BMC Health Serv Res*, 10:111.
- Frank, A. S., Smith, D.E. (2010). Measurement Invariance, Entropy and Probability. *Entropy*, 12:289–303. Glynn, L.G., Valderas, J.M., Healy, P., Burke, E., Newell, J., Gillespie, P., Murphy, A.W. (2011). The prevalence of multimorbidity in primary care and its effect on healthcare utilization and cost. *Family Practice*, 28 (5):516–523.
- Islam, M. M., Valderas, J. M., Yen, L., Dawda, P., Jowsey, T., McRae, I. S. (2014). Multimorbidity and comorbidity of chronic diseases among senior australiens: Prevalence and patterns. *PloS one*.
- Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630.
- Jaynes, E. T. (1957b). Information theory and statistical mechanics II. *Phys. Rev.*, 108:171–190. Modeling (Multi-)Morbidity and (Poly-)Pharmacy in Outpatient Treatment with Gamma Distributions 357 Jaynes, E. T. (2003). Probability Theory: The Logic of Science. Cambridge University Press: New York, NY, USA.
- Jeschke, E., Ostermann, T., Vollmar, HC, Tabali, M., Matthes, H. (2012). Depression, Comorbidities, and Prescriptions of Antidepressants in a German Network of GPs and Specialists with Subspecialisation in Anthroposophic Medicine: A Longitudinal Observational Study. *Evid Based Complement Alternat Med*. 2012, 508623.

Abstracts of Contributed Posters

Johnell, K., Klarin, I. (2007). The relationship between number of drugs and potential drug-drug interactions in the elderly: A study of over 600,000 elderly patients from the swedish prescribed drug register. *Drug Saf.*, 30:911–918.

Maher, R. L., Hanlon, J., Hajjar, E. R. (2014). Clinical consequences of polypharmacy in elderly. *Expert Opin. Drug Saf.*, 13:57–65.

Mitty, E. (2009). Medication management in assisted living: A national survey of policies and practices. *J. Am. Med. Dir. Assoc.*, 10:107–114. Ostermann, T., Schuster, R. (2015). An Informationtheoretical Approach to Classify Hospitals with Respect to Their Diagnostic Diverty using Shannon's Entropy. *Proceedings of the International Conference on Health Informatics (HealthInf)*, pages 325–329.

Salwe, K. J., Kalyansundaram, D., Bahurupi, Y. (2016). A study on polypharmacy and potential drug-drug interactions among elderly patients admitted in department of medicine of a tertiary care hospital in puducherry. *J. Clin. Diagn. Res.*, 10:6–10.

Schuster, R., Emcke, T., Arnstedt, E.v., Heidbreder, M. (2016). Morbidity Related Groups (MRG) for epidemiological analysis in outpatient treatment. IOS Press, pages 783–787.

Schuster, R., Ostermann, T., Heidbreder, M., Emcke, T. (2018). Relations of Morbidity Related Groups (MRG), ICD-10 codes and age and gender structure in outpatient treatment. *Proceedings of the 11th International Joint Conference on Biomedical Engineering and Technologies (BIOSTEC)*, HealthInf.

Subeesh, V. K., Gouri, N., Beulah, E. T., Shivaskankar, V. (2017). A Prospective Observational Study on Polypharmacy in Geriatrics at A Private Corporate Hospital. *Journal of Applied Pharmaceutical Science*, 7(10):162–167.

Subeesh, V. K., Shivaskankar, V., Gouri, N., Sriram, S. (2015). A study on prevalence and impact of potentially inappropriate medication use in geriatrics at a private corporate hospital. *Pharmanest*, 6:2845–2848. HEALTHINF 2019 - 12th International Conference on Health Informatics 358

High Degree of Agreement but Low Measures - Problems of Inter-Rater Reliability Measures for Unbalanced Ordinal Data (Poster ID 21)

Grittner, Ulrike¹; Schulze, Daniel¹; Unger, Nina²; Keller, Theresa¹; Breitenstein, Caterina³; Pigorsch, Mareen¹; Mönch, Maximilian¹

¹Institute of Biometry and Clinical Epidemiology, Charité Universitätsmedizin Berlin, Germany

²Department of Neurology, University Medicine Greifswald, Germany

³Department of Neurology, University of Münster, Germany

Common reliability coefficients such as Cohen's kappa (), Krippendorff's alpha or Scott's yield in biased coefficients in the case of unbalanced ordinal data and true high agreement between raters. This phenomenon is known as the paradoxes of kappa. In extreme cases, reliability coefficients can be 0 or even negative in presence of high agreement between raters. In this talk, we explain the origin of this phenomenon. Alternative measures of reliability have been proposed by Gwet (2008) or Brennan and Prediger (1981). Although these measures provide unbiased measures for extreme cases of agreement and unbalanced ordinal data, they are biased in other scenarios. Presenting examples, simulations and formulas, we will discuss this problem.

References

Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 2008;61(1):29-48.

Brennan RL, Prediger DJ. Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement*. 1981;41(3):687-99.

Time-dependent change in risk through an exposure and possible estimands of interest: A simulation study using two clinically motivated examples (Poster ID 23)

Meiszl, Katharina; Tokic, Marianne C.; Timmesfeld, Nina

Ruhr-Universität Bochum, Abteilung für Medizinische Informatik, Biometrie und Epidemiologie, Bochum, Deutschland

In epidemiologic research, it is often simplistically assumed that exposure has a constant effect on the risk of an event over a given period of time. In the case of recurrent event analyses, this means that a constant relative risk is assumed over this period, i.e. the relative risk is modeled with a step function over time [1].

In our two clinical real-world examples, acute myocarditis following COVID-19 vaccination and multiple sclerosis (MS) relapses during pregnancy, this assumption is not biologically plausible [2-4]. In both examples, it is already known that the risk varies over time. Risk modification is typically described by relative risks (RR) or hazard ratios (HR) interpreted as relative risks.

Using the two examples, we want to explore which measure is estimated when analysis is performed using a step function and how the chosen analysis model influences this measure.

For both examples, data is generated by discrete time simulations [5] and analyzed using Monte Carlo simulation [6]. RRs are simulated as a function of time. In the example with acute myocarditis, simulated data will be analyzed using the self-controlled case series method [7]. For the MS example, primary analysis will be based on a mixed effects Poisson model [8].

In this poster, the observed results of our simulation will be compared with different measures describing the time-depending RR, i.e. arithmetic mean, geometric mean, and median.

References

- [1] Farrington CP. Relative Incidence Estimation from Case Series for Vaccine Safety Evaluation. *Biometrics*. 1995 Mar;51(1):228. Available from: <http://dx.doi.org/10.2307/2533328>
- [2] Hromić-Jahjefendić A, Sezer A, Aljabali AAA, Serrano-Aroca Á, Tambuwala MM, Uversky VN, et al. COVID-19 Vaccines and Myocarditis: An Overview of Current Evidence. *Biomedicines*. 2023 May 17;11(5):1469. Available from: <http://dx.doi.org/10.3390/biomedicines11051469>

- [3] Confavreux C, Hutchinson M, Hours MM, Cortinovis-Tourniaire P, Moreau T. Rate of Pregnancy-Related Relapse in Multiple Sclerosis. *New England Journal of Medicine*. 1998 Jul 30;339(5):285–91. Available from: <http://dx.doi.org/10.1056/nejm199807303390501>
- [4] Langer-Gould A, Smith JB, Albers KB, Xiang AH, Wu J, Kerezsi EH, et al. Pregnancy-related relapses and breastfeeding in a contemporary multiple sclerosis cohort. *Neurology*. 2020 Apr 13;94(18):e1939–49. Available from: <http://dx.doi.org/10.1212/wnl.00000000000009374>
- [5] Tang J, Leu G, Abbass HA. Discrete Time Simulation. *Simulation and Computational Red Teaming for Problem Solving*. 2019 Oct 18;143–55. Available from: <http://dx.doi.org/10.1002/9781119527183.ch8>
- [6] Mooney C. Monte Carlo Simulation. 1997; Available from: <http://dx.doi.org/10.4135/9781412985116>
- [7] Farrington P, Rush M, Miller E, Pugh S, Colville A, Flower A, et al. A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *The Lancet*. 1995 Mar;345(8949):567–9. Available from: [http://dx.doi.org/10.1016/s0140-6736\(95\)90471-9](http://dx.doi.org/10.1016/s0140-6736(95)90471-9)
- [8] Tango T. Repeated Measures Design with Generalized Linear Mixed Models for Randomized Controlled Trials. 2017 Sep 14; Available from: <http://dx.doi.org/10.1201/9781315152097>

Job Advertisements from Sponsors

Account Manager*in (m/w/d)

Über Perfood

„Lass Deine Ernährung Deine Medizin sein.“ Bei Perfood entwickeln wir digitale Therapien basierend auf personalisierter Ernährung. Unser Ursprung liegt in der universitären Forschung. Unsere Stärke liegt in der Verwendung modernster Technologien und wissenschaftlicher Erkenntnisse. Mit unserem Produkt MillionFriends Original haben wir ein innovatives und effektives Ernährungsprogramm zur Prävention geschaffen. Mit sinCephalea haben wir die erste ernährungsbasierte und effektive Digitale Gesundheitsanwendung für die Therapie von Migräne entwickelt. In den kommenden Jahren wollen wir basierend auf unserer zum Patent angemeldeten Technologie eine Pipeline digitaler Gesundheitsanwendungen aufbauen. Dafür arbeiten wir in Forschungs- und Entwicklungsprojekten mit den führenden akademischen Einrichtungen, beispielsweise der Universität zu Lübeck sowie der Fraunhofer und Helmholtz Gesellschaft und der Leibniz Gemeinschaft zusammen. Dabei steht der Patient immer an erster Stelle, denn das Ziel von Perfood ist es die Lebensqualität nachhaltig zu verbessern. Ohne Risiken und Nebenwirkungen.

Wir suchen Teammitglieder, die uns dabei helfen wollen, die Zukunft der Ernährungsmedizin neu zu gestalten, indem sie personalisierte Ernährung mit digitaler Medizin kombinieren und so Patienten helfen. Wir sind auf der Suche nach einem Vertriebsprofi (m/w/d) in Teil-/Vollzeit (unbefristete Anstellung), der daran interessiert ist, die Entwicklung und Erforschung innovativer Therapeutika aktiv und eigenständig mitzugestalten.

Deine Aufgaben

- Akquise von Ärzten als Kunden für die Medizinprodukte von Perfood
- Aktive Ansprache und Beratung von potenziellen Verschreibern im persönlichen Gespräch vor Ort in Arztpraxen, medizinischen Versorgungszentren, Kliniken oder auf Veranstaltungen
- Regelmäßige Teilnahme an deutschlandweiten Fachkongressen und Veranstaltungen lokaler Ärztenetzwerke oder Patientennetzwerke
- Aufbau und Pflege von Kontakten per Außendienst, Telefon, E-Mail, oder Video-Telefonie
- Steuerung und Umsetzung von Strategiekonzepten
- Definition und Überwachung relevanter KPIs
- Stetige Analyse des Vertriebsgebiets mit dem Ziel, potenzielle Kunden zu identifizieren und aktiv zu akquirieren
- Entwicklung, Implementierung, Umsetzung und Evaluation von Vertriebsmaßnahmen
- Dokumentation von Kundenerfahrungen und jeglicher Aktivitäten in unserem CRM, sowie kontinuierlicher Austausch mit dem Team zur Optimierung der Vertriebsaktivitäten
- Zusammenarbeit mit unserem Marketing-, Service- und Content-Team

Was Du mitbringst

- Abgeschlossenes Hochschulstudium oder Ausbildung im Gesundheitsbereich oder kaufmännischem Bereich

- Mind. 2 Jahre Berufserfahrung im Vertrieb bzw. Außendienst, idealerweise im Gesundheitswesen, Pharma- oder MedTech-Bereich Erfahrungen mit digitalen Anwendungen und Kommunikationskanälen
- Bereitschaft Fachwissen in Bereich der Digitalen Gesundheitsanwendungen (DiGA) anzueignen
- Einfühlungsvermögen für die Bedürfnisse und Ziele im Gesundheitswesen
- Grundverständnis für die Struktur und die Rahmenbedingungen der verschiedenen Akteure im Gesundheitswesen
- Hohe Kundenorientierung, starke kommunikative Fähigkeiten und sehr ausgeprägte Netzwerk-Qualitäten
- Selbstbewusstes und professionelles Auftreten und hohe intrinsische Motivation unsere Produkte überzeugend im Dialog mit anderen Experten in den Arztpraxen zu präsentieren
- Gültiger Führerschein und Reisebereitschaft (bis zu 80%)
- Sehr gute Organisationsfähigkeit, selbstständige und strukturierte Arbeitsweise sowie Entscheidungskompetenz
- Empathie, Flexibilität, Zuverlässigkeit und Freude an kollegialer Arbeit im Team
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift
- komfortabel im Umgang mit Laptop, Handy und MS-Office Programmen und Bereitschaft, sich in das CRM Tool einzuarbeiten und dies regelmäßig zu pflegen

Was Dich erwartet

- Die Chance beim Aufbau eines Unternehmens mit einem hoch-innovativem Produkt in einem wachstumsstarken Markt aktiv zu unterstützen
- Ein diverses und kollegiales Team, das Dich bei der Einarbeitung unterstützt
- Freie Einteilung der Arbeitszeit und Auswahl Deines Arbeitsmaterials
- Die Chance die medizinische Versorgung zahlreicher Erkrankungen maßgeblich zu verbessern.
- Die Zusammenarbeit in einem interdisziplinären Team aus den Bereichen Medizin, Ernährungswissenschaften, Data Analytics und Künstliche Intelligenz sowie Rechtswissenschaften und Betriebswirtschaft und die damit einhergehende Möglichkeit Dein bestehendes Wissen über das eigene Aufgabenfeld hinaus zu erweitern
- Teilnahme an spaßigen online und offline Teamevents und kostenlose Nutzung des Produkts MillionFriends

Biostatistician (m/w/d)

You want to combine nutrition with digital medicine and deliver cutting-edge technology to patients? You want to develop and research innovative therapeutics to bring health to the people? And all this in a dynamic, friendly start-up? Then join our dynamic startup team at Perfood GmbH and make a real impact on people's health.

To drive our growth, we are looking for a **Biostatistician (m/f/d)** in part-time or full-time (permanent position). You will play a key role in our interdisciplinary **Research & Development team** and take over full responsibility for the statistical planning and evaluating of our pre-clinical and clinical studies as well as real-world data from our digital products (software as a medical device). With your academic background and your experience in (pharma-) industry research you will be a core driver of our clinical study pipeline.

Your tasks

- Support the R&D team in planning, conducting, evaluating and publishing pre-clinical and clinical studies
- Planning and analysis for all types of statistical data from real-world data to clinical trials, from proof-of-concept to pilot to pivotal studies. This involves preparation of statistical study planning, SAPs and support in the preparation of other essential study documents
- Set-up a real-world data assessment concept for our clinical pipeline
- Develop new statistical concepts expand state of the art methods to ensure higher quality, efficiency and improved compliance
- Responsibility for producing high quality biostatistical analyses for our digital product development
- Contribute to internal education and journal clubs
- Cross-functional work with our Data Science team

What you bring

- Completed Bachelor's and/or Master's degree in (bio-) statistics, biometrics, epidemiology, mathematics or comparable degree with sound basic knowledge in medicine
- A completed doctoral degree or comparable academic experience is desirable
- Professional experience with clinical and non-clinical studies, incl. the legal framework (ICH-GCP, MPG, AMG, among others), ideally experience in (pharma-) industrial research
- Experience with experimental designs and innovative methods (e.g. adaptive study designs) is desirable
- Very good knowledge of German and English, both written and spoken
- Basic programming skills, e.g. in Python and SQL, is helpful for the cross-functional work with the Data Science team but non mandatory
- Very good organizational skills, independent and structured way of working as well as decision-making skills
- Empathy, motivation, flexibility, initiative, independence, communication skills, reliability and enjoy working in a collegial team

How we'll convince you to join us

- The possibility to work fully remote, in home office or in the office in Lübeck
- Flexible working hours, as long as participation in our team meetings is ensured
- A friendly and dynamic team that will support you during the training process
- Team events on a regular basis (virtual and onsite)
- Get our product "MillionFriends" for free as often as you like for yourself and get a discount for family and friends
- Collaboration in an interdisciplinary team from the fields of medicine, nutritional sciences, data analytics, and artificial intelligence as well as law and business administration, and the opportunity to use your existing knowledge beyond your own field of activity
- The opportunity to actively shape the strategy, culture, and organization of a young company
- The chance to help build up a company with a highly innovative product in a fast-growing market
- Office in Lübeck with very good transport connections to Hamburg and the Baltic Sea
- Snacks and drinks at your free disposal in the office in Lübeck
- Beautiful office building in an old brewery with a view of the 7 towers of Lübeck

Clinical Study Management (m/w/d)

Über Perfood

„Lass Deine Ernährung Deine Medizin sein.“ Bei Perfood entwickeln wir digitale Therapien basierend auf personalisierter Ernährung. Unser Ursprung liegt in der universitären Forschung. Unsere Stärke liegt in der Verwendung modernster Technologien und wissenschaftlicher Erkenntnisse. Mit unserem Produkt MillionFriends Original haben wir ein innovatives und effektives Ernährungsprogramm zur Prävention geschaffen. Mit sinCephalea haben wir die erste ernährungsbasierte und effektive Digitale Gesundheitsanwendung für die Therapie von Migräne entwickelt. In den kommenden Jahren wollen wir basierend auf unserer zum Patent angemeldeten Technologie eine Pipeline digitaler Gesundheitsanwendungen aufbauen. Dafür arbeiten wir in Forschungs- und Entwicklungsprojekten mit den führenden akademischen Einrichtungen, beispielsweise der Universität zu Lübeck sowie der Fraunhofer und Helmholtz Gesellschaft und der Leibniz Gemeinschaft zusammen. Dabei steht der Patient immer an erster Stelle, denn das Ziel von Perfood ist es die Lebensqualität nachhaltig zu verbessern. Ohne Risiken und Nebenwirkungen.

Du möchtest personalisierte Ernährung mit digitaler Medizin kombinieren und damit Patienten helfen? Du möchtest aktiv und eigenverantwortlich die Entwicklung und Forschung innovativer Therapeutika gestalten? Du hast Lust auf ergebnisorientiertes Arbeiten in einem dynamischen, freundlichen Start-up?

Wir suchen ab sofort ein erfahrenes Teammitglieder (m/w/d) für den interdisziplinären Bereich "Forschung & Entwicklung" in Teil-/Vollzeit (unbefristete Anstellung).

Deine Aufgaben

- Übernahme der Verantwortung für die Planung und Durchführung unserer Studien
- Sicherung der erfolgreichen Studiendurchführung unserer proof-of-concept-, Pilot-, und Zulassungsstudien unserer Software- und Medizinprodukte, v.a. in den Indikationen Migräne, Typ 2 Diabetes mellitus, aber auch in neuen Geschäftsfeldern
- Übernahme von Personalverantwortung für die Mitarbeitenden im Bereich Trial Management
- Überwachung/Monitoring von Studien und Einhaltung von Planungsvorgaben, regulatorischer und ethischer Vorgaben (z.B. MDR, ICH/GCP) und Qualitätsziele während der gesamten Studie
- Auswahl, Koordination und Überwachung von qualifizierten Prüfzentren, externen Dienstleister (z.B. CROs) zur Studiendurchführung und weiteren Stakeholder wie Ethikkommissionen und regulatorischen Behörden
- Erstellung antragsrelevanter Studienpläne und -dokumente
- Enge Zusammenarbeit in allen Belangen mit dem Chief Medical Officer und den Mitarbeitenden im Bereich Medical Affairs

Was Du mitbringst

- Abgeschlossenes Hochschulstudium
- Mind. 3 Jahre Berufserfahrung (außerhalb einer Promotion) mit klinischen und nicht-klinischen Studien

- Erfahrung in der außeruniversitären industriellen Forschung, beispielsweise aus Pharma, Med Tech oder einer CRO
- Solides Wissen in statistischer Studienplanung und -auswertung
- Bereitschaft Personalverantwortung zu übernehmen
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift
- Gute PC-Kenntnisse, insbesondere im Umgang mit MS-Office Programmen
- Sehr gute Organisationsfähigkeit, selbstständige und strukturierte Arbeitsweise sowie Entscheidungskompetenz
- Empathie, Motivation, Flexibilität, Eigeninitiative, Eigenständigkeit, Kommunikationsfähigkeit und Zuverlässigkeit
- Freude an kollegialer Arbeit im Team
- Erfahrung mit Stoffwechselkrankungen um das Metabolische Syndrom sowie Entzündungserkrankungen sind von Vorteil
- Verständnis von (betriebs-)wirtschaftlichen Rahmenbedingungen und Auswirkungen von F&E-Investitionen auf die Geschäftsentwicklung sind von Vorteil

Was Dich erwartet

- Ein hochinnovatives und effektives Produkt in einer wachstumsstarken Branche
- Ein gutes und kollegiales Team, das Dich bei der Einarbeitung unterstützt
- Die Chance die medizinische Versorgung zahlreicher Erkrankungen maßgeblich zu verbessern
- Die Zusammenarbeit in einem interdisziplinären Team aus den Bereichen Medizin, Ernährungswissenschaften, Data Analytics und Künstliche Intelligenz sowie Rechtswissenschaften und Betriebswirtschaft und die damit einhergehende Möglichkeit Dein bestehendes Wissen über das eigene Aufgabenfeld hinaus zu erweitern
- Die Chance beim Aufbau eines Unternehmens mit einem hoch-innovativem Produkt in einem wachstumsstarken Markt aktiv zu unterstützen
- Snacks und Getränke im Büro zur freien Verfügung
- Schönes Bürogebäude in einer alten Brauerei mit Blick auf die 7 Türme von Lübeck
- Auswahl Deines Arbeitsmaterials
- Stelle kann auch remote besetzt werden
- Teilnahme an spaßigen online und offline Teamevents

Digital Marketing Manager - Healthcare (m/w/d)

Über Perfood

„Lass Deine Ernährung Deine Medizin sein.“ Bei Perfood entwickeln wir digitale Therapien basierend auf personalisierter Ernährung. Unser Ursprung liegt in der universitären Forschung. Unsere Stärke liegt in der Verwendung modernster Technologien und wissenschaftlicher Erkenntnisse. Mit unserem Produkt MillionFriends Original haben wir ein innovatives und effektives Ernährungsprogramm zur Prävention geschaffen. Mit sinCephalea haben wir die erste ernährungsbasierte und effektive Digitale Gesundheitsanwendung für die Therapie von Migräne entwickelt. In den kommenden Jahren wollen wir basierend auf unserer zum Patent angemeldeten Technologie eine Pipeline digitaler Gesundheitsanwendungen aufbauen. Dafür arbeiten wir in Forschungs- und Entwicklungsprojekten mit den führenden akademischen Einrichtungen, beispielsweise der Universität zu Lübeck sowie der Fraunhofer und Helmholtz Gesellschaft und der Leibniz Gemeinschaft zusammen. Dabei steht der Patient immer an erster Stelle, denn das Ziel von Perfood ist es die Lebensqualität nachhaltig zu verbessern. Ohne Risiken und Nebenwirkungen.

Wir suchen Teammitglieder, die uns dabei helfen wollen, die Zukunft der Ernährungsmedizin neu zu gestalten, indem sie personalisierte Ernährung mit digitaler Medizin kombinieren und so Patienten helfen. Wir sind auf der Suche nach einem Online Marketing-Manager (m/w/d) in Teil-/Vollzeit (unbefristete Anstellung), der daran interessiert ist, die Entwicklung und Erforschung innovativer Therapeutika aktiv und eigenständig mitzugestalten.

Deine Aufgaben

- Eigenverantwortliche Konzeption, Umsetzung und Steuerung digitaler Werbekampagnen über verschiedene Kanäle wie SEA und Social-Media (Meta, TikTok, LinkedIn)
- Optimierung der Websites mittels Suchmaschinenoptimierung, um mehr Traffic und Leads zu generieren
- Messung des Erfolgs anhand von Kennzahlen wie z.B. Matomo
- Überwachung und Durchführung von Traffic-Analysen
- Entwicklung und Umsetzung von Performance-Marketing-Strategien zur Förderung von Kundenakquise, Konversion und Umsatzwachstum
- Erstellung von Creatives und Ad Copies in Zusammenarbeit mit Grafik Designern
- Trendanalysen, Markt- und Wettbewerbsbeobachtung
- Kontinuierliche Verbesserung der Online-Marketing-Maßnahmen
- Enge Zusammenarbeit mit verschiedenen Abteilungen zur effizienten Organisation und Durchführung

Was Du mitbringst

- Abgeschlossenes Hochschulstudium oder Ausbildung in den Bereichen Gesundheitswesen, Marketing, Betriebswirtschaft, (Wirtschafts-)Psychologie, Kommunikation oder vergleichbare Qualifikationen
- Mehrjährige Berufserfahrung im Bereich Online-Marketing und der Werbung idealerweise im Gesundheitswesen, Pharma- oder MedTech-Bereich

- Bereitschaft zur Einarbeitung in das Heilmittelwerbegesetz und ähnlichen Regularien
- Kenntnisse der gängigen Performance Marketing Tools (Meta, Google, Tiktok etc.)
- Hohe Affinität zu digitalen Medien
- Begeisterung für digitale Trends, neue Kanäle und Entwicklungen im Online-Marketing
- Erfahrung mit SEA-Kampagnen sowie sicheres Arbeiten mit MS-Office Programmen
- Hohe Kundenorientierung und starke kommunikative Fähigkeiten
- Schnelles Auffassungsvermögen und analytisches Denken
- Sehr gute Organisationsfähigkeit, selbstständige und strukturierte Arbeitsweise
- Empathie, Kreativität, Flexibilität, Zuverlässigkeit und Freude an kollegialer Arbeit im Team
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift

Was Dich erwartet

- Die Chance beim Aufbau eines Unternehmens mit einem hoch-innovativem Produkt in einem wachstumsstarken Markt aktiv zu unterstützen
- Ein diverses und kollegiales Team, das Dich bei der Einarbeitung unterstützt
- Freie Einteilung der Arbeitszeit und Auswahl Deines Arbeitsmaterials
- Die Chance die medizinische Versorgung zahlreicher Erkrankungen maßgeblich zu verbessern
- Die Zusammenarbeit in einem interdisziplinären Team aus den Bereichen Medizin, Ernährungswissenschaften, Data Analytics und Künstliche Intelligenz sowie Rechtswissenschaften und Betriebswirtschaft und die damit einhergehende Möglichkeit Dein bestehendes Wissen über das eigene Aufgabenfeld hinaus zu erweitern
- Teilnahme an spaßigen online und offline Teamevents und kostenlose Nutzung des Produkts MillionFriends

Indication Lead Metabolic (Medical Affairs Manager) (m/w/d)

Über Perfood

„Lass Deine Ernährung Deine Medizin sein.“ Bei Perfood entwickeln wir digitale Therapien basierend auf personalisierter Ernährung. Unser Ursprung liegt in der universitären Forschung. Unsere Stärke liegt in der Verwendung modernster Technologien und wissenschaftlicher Erkenntnisse.

Wir arbeiten in Forschungs- und Entwicklungsprojekten mit den führenden akademischen Einrichtungen, beispielsweise der Universität zu Lübeck sowie der Fraunhofer und Helmholtz Gesellschaft und der Leibniz Gemeinschaft zusammen. Dabei stehen die Patient*innen immer an erster Stelle, denn das Ziel von Perfood ist es die Lebensqualität nachhaltig zu verbessern.

Mit unserem Produkt MillionFriends haben wir ein innovatives und effektives Ernährungsprogramm zur Prävention geschaffen. Mit sinCephalea haben wir die erste ernährungsbasierte und effektive Digitale Gesundheitsanwendung (DiGA) für die Therapie von Migräne entwickelt.

Über glucura

Mit glucura (www.glucura.de) haben wir ein weiteres Medizinprodukt basierend auf unserer zum Patent angemeldeten Technologie für die Behandlung des Typ 2 Diabetes mellitus entwickelt. Erste Studiendaten zeigen eine herausragende klinische Wirksamkeit und das Potential Komplikationen und unerwünschte Arzneimittelnebenwirkungen effektiv zu senken. glucura übersetzt die neuesten wissenschaftlichen Erkenntnisse der personalisierten Ernährung in einfach umsetzbare Empfehlungen für die Betroffenen. glucura ermöglicht die Analyse von individuellen Daten gemessen mit CGM-Sensoren und mithilfe unserer App-Technologie. Patient*innen setzen sich individuelle Therapieziele, werden an Bewegung und Sport herangeführt und lernen ihre eigene Erkrankung und effektive Verhaltensregeln kennen. Ziel ist es, glucura in die Regelversorgung zu bringen und neue Behandlungsperspektiven aufzuzeigen.

Du möchtest personalisierte Ernährung mit digitaler Medizin kombinieren und damit innovative Therapieoptionen für Menschen mit metabolischen Erkrankungen, insbesondere Typ 2 Diabetes mellitus und Metabolisches Syndrom, schaffen?

Du möchtest klinische Studien planen und die wissenschaftliche Weiterentwicklung von glucura aktiv und eigenverantwortlich gestalten?

Du hast Lust auf ergebnisorientiertes Arbeiten in einem dynamischen, freundlichen Start-up?

Wir suchen ab sofort ein erfahrener Teammitglied (m/w/d) für den interdisziplinären Bereich „Forschung & Entwicklung“ in Teil-/Vollzeit (unbefristete Anstellung).

Deine Aufgaben

- Für metabolische Erkrankungen (v.a. Typ 2 Diabetes) bist Du für medizinische und wissenschaftliche Fragen die Ansprechperson bei Perfood und unterstützt die Produkt-, Marketing- und Vertriebs-Teams

- Aufbau und Pflege eines Netzwerks mit relevanten klinischen/wissenschaftlichen Expert:innen, Institutionen, Fachgesellschaften und Forschungsgruppen aus dem Bereich Typ 2 Diabetologie
- Übernahme der Verantwortung für die Planung unserer Studien im Bereich metabolische Erkrankungen, primär Typ 2 Diabetologie
- Erstellung antragsrelevanter Studienpläne und -dokumente
- Unterstützung bei der Durchführung unserer klinischen Studien und Überwachung von qualifizierten Prüfzentren und externen Dienstleister (z.B. CROs)
- Enge Zusammenarbeit in allen Belangen mit dem Chief Medical Officer und den Mitarbeitenden im Bereich Medical Affairs und Trial Management

Was Du mitbringst

- Abgeschlossenes Hochschulstudium mit Promotion (Medizin, Naturwissenschaften, Gesundheitswissenschaften, Pharmazie o.ä.)
- Mind. 3 Jahre Berufserfahrung (außerhalb einer Promotion) im Themenbereich metabolische Erkrankungen und insbesondere Typ 2 Diabetes mellitus
- Einschlägige Erfahrung in der klinischen Forschung und profundes Verständnis der klinischen Versorgung von Patient:innen mit Typ 2 Diabetes und anderen metabolischen Erkrankungen inkl. der Behandlungsleitlinien
- Bereitschaft Personalverantwortung zu übernehmen
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift
- Gute PC-Kenntnisse, insbesondere im Umgang mit MS-Office Programmen
- Sehr gute Organisationsfähigkeit, selbstständige und strukturierte Arbeitsweise sowie Entscheidungskompetenz
- Empathie, Motivation, Flexibilität, Eigeninitiative, Eigenständigkeit, Kommunikationsfähigkeit, Zuverlässigkeit und Freude an kollegialer Arbeit im Team
- Verständnis von (betriebs-)wirtschaftlichen Rahmenbedingungen und Auswirkungen von F&E-Investitionen auf die Geschäftsentwicklung sind von Vorteil

Was Dich erwartet

- Ein hochinnovatives und effektives Produkt in einer wachstumsstarken Branche
- Ein gutes und kollegiales Team, das Dich bei der Einarbeitung unterstützt
- Die Chance die medizinische Versorgung zahlreicher Erkrankungen maßgeblich zu verbessern
- Die Zusammenarbeit in einem interdisziplinären Team aus den Bereichen Medizin, Ernährungswissenschaften, Data Analytics und Künstliche Intelligenz sowie Rechtswissenschaften und Betriebswirtschaft und die damit einhergehende Möglichkeit Dein bestehendes Wissen über das eigene Aufgabenfeld hinaus zu erweitern
- Die Chance beim Aufbau eines Unternehmens mit einem hoch-innovativem Produkt in einem wachstumsstarken Markt aktiv zu unterstützen
- Snacks und Getränke im Büro zur freien Verfügung
- Schönes Bürogebäude in einer alten Brauerei mit Blick auf die 7 Türme von Lübeck
- Auswahl Deines Arbeitsmaterials
- Stelle kann auch remote besetzt werden
- Teilnahme an spaßigen online und offline Teamevents

Product Designer UI/Visual (m/w/d)

Join our dynamic startup team at Perfood GmbH and make a real impact on people's health. As a Product Manager, you will play a key role in shaping our innovative personalized nutrition program, using cutting-edge technology to analyze blood glucose measurements providing personalized nutritional recommendations based on the latest scientific findings. With your background in agile software development, you will work closely with our cross-functional teams to design, develop and launch products that can help prevent and treat diseases like migraines or diabetes type 2 and improve overall well-being. Be part of a supportive and friendly work environment, where you can apply your skills and passion to make a meaningful difference.

Your tasks

- Collaborate closely with cross-functional teams of engineers, product managers, product designers and key stakeholders to deliver high-quality solution-oriented and user-centric designs
- Create and iterate on visual design experience, concepts, flows, user stories, prototypes, and wireframes to illustrate world-class user experience for iOS, Android, and Web
- Create visual and interaction design specs
- Develop and iterate on visuals that build our product experience and design language
- Shape, measure and improve your designs based on quantitative and qualitative feedback
- Shape modern UI patterns for digital health therapeutics products
- Integrate Product Design naturally into the agile software development process

What you bring

- At least 3+ years of experience in Product Design (UI) with a focus on mobile apps and an excellent understanding of UI and interaction design patterns
- Experience with product delivery in agile and cross-functional teams, as well as working closely with Product Owners and Engineers
- A track record on successful product experience design in consumer products
- Comfortable in prototyping from sketches to digital wire-framing and hi-fidelity UI in Figma
- Innovative with creative ideas and passion for UI, and Product Design
- Experience with Design Systems, as well as designing, prototyping, and developer hand-off
- Structured with creative thinking and a proactive, initiative, and detail-oriented mindset
- CSS and Frontend Development experience is a plus
- Very good communication skills and collaborative approach
- Fluency in German & English

How we'll convince you to join us

- The possibility to work fully remote, in home office or in the office in Lübeck
- Flexible working hours, as long as participation in our team meetings is ensured
- A friendly and dynamic team that will support you during the training process
- Team events on a regular basis (virtual and onsite)

- Get our product "MillionFriends" for free as often as you like for yourself and get a discount for family and friends
- Collaboration in an interdisciplinary team from the fields of medicine, nutritional sciences, data analytics, and artificial intelligence as well as law and business administration, and the opportunity to use your existing knowledge beyond your own field of activity
- The opportunity to actively shape the strategy, culture, and organization of a young company
- The chance to help build up a company with a highly innovative product in a fast-growing market
- Office in Lübeck with very good transport connections to Hamburg and the Baltic Sea
- Snacks and drinks at your free disposal in the office in Lübeck
- Beautiful office building in an old brewery with a view of the 7 towers of Lübeck

Product Manager (m/w/d)

Join our dynamic startup team at Perfood GmbH and make a real impact on people's health. As a Product Manager, you will play a key role in shaping our innovative personalized nutrition program, using cutting-edge technology to analyze blood glucose measurements providing personalized nutritional recommendations based on the latest scientific findings. With your background in agile software development, you will work closely with our cross-functional teams to design, develop and launch products that can help prevent and treat diseases like migraines or diabetes type 2 and improve overall well-being. Be part of a supportive and friendly work environment, where you can apply your skills and passion to make a meaningful difference.

Your tasks

- Identify and define product requirements for existing and future products
- Create and validate product concepts and lead them to execution
- Prepare new products or new markets with a project management mindset
- Work closely with development teams, medical R&D, marketing & sales, content and other product roles and manage products in the market
- Develop and analyze relevant data and key performance indicators
- Communicate with internal stakeholders and external partners

What you bring

- Minimum 2-3 years of work experience as Product Manager, Product Owner and/or in Analytics. (Startup experience appreciated)
- You're known for your empathy. You think of users first and strive to understand yours better than anyone else
- Should be multifaceted: you can sell your ideas to engineers and executives, you can break down complex concepts into easy tasks and you know how to bring the best of every element of your squad
- Analytical and product mindset
- Hard-working with the right drive to take risks and get things done in a proper manner
- Motivated to make an impact, deliver results quickly, and raise the bar in the ecosystem
- Proficiency in German and English language

How we'll convince you to join us

- The possibility to work fully remote, in home office or in the office in Lübeck
- Flexible working hours, as long as participation in our team meetings is ensured
- A friendly and dynamic team that will support you during the training process
- Team events on a regular basis (virtual and onsite)
- Get our product "MillionFriends" for free as often as you like for yourself and get a discount for family and friends
- Collaboration in an interdisciplinary team from the fields of medicine, nutritional sciences, data analytics, and artificial intelligence as well as law and business administration, and the opportunity to use your existing knowledge beyond your own field of activity
- The opportunity to actively shape the strategy, culture, and organization of a young company

- The chance to help build up a company with a highly innovative product in a fast-growing market
- Office in Lübeck with very good transport connections to Hamburg and the Baltic Sea
- Snacks and drinks at your free disposal in the office in Lübeck
- Beautiful office building in an old brewery with a view of the 7 towers of Lübeck

Software Developer Angular/TypeScript (m/w/d)

Perfood offers a scientifically based program for personalized nutrition. By means of a continuous blood glucose measurement and a determination of the intestinal bacteria, we analyze which foods trigger good and bad reactions in the individual. These insights are the foundation of our personalized nutrition recommendations. According to the latest scientific findings, widespread diseases such as migraine can be treated through individualized nutrition and general well-being can be increased. Due to our role as a medical device manufacturer, we have high standards and requirements regarding data security and privacy as well as smooth operation of all working equipment.

We are looking for team members who are eager to help us reshape the future of nutritional medicine. To drive our growth, we are looking for a part-time or full-time software development team member!

Your mission

- Further development of our medical app products, which guides customers through our digital therapy program
- Agile working in a scrum team with experienced Software Architects, Product Developers, and Data Analysts
- Implement new technologies at the highest security standard to protect the data of our patients
- Continuous optimization of the performance and usability of the project together with the product team
- Keeping an eye on the regulatory requirements of a medical device and incorporate them into the development

What you bring

General:

- At least 2 years of experience working on a software development team
- Degree in computer science or similar field
- Motivation to learn new things and dive into complex issues
- Fun to develop yourself and the team
- You speak German and English, at least one of them fluently

Technical knowledge:

- Angular / TypeScript
- NodeJS
- Ionic
- agile development methods
- confident handling of git
- automated software testing

It would be a big plus if you are experienced in CSS, native iOS/Android App Development, Python, SQL, CouchDB and CI/CD, Docker

How we'll convince you to join us

- The possibility to work fully remote, in home office or in the office in Lübeck
- Flexible working hours, as long as participation in our team meetings is ensured
- A friendly and dynamic team that will support you during the training process
- Team events on a regular basis (virtual and onsite)
- Get our product "MillionFriends" for free as often as you like for yourself and get a discount for family and friends
- Collaboration in an interdisciplinary team from the fields of medicine, nutritional sciences, data analytics, and artificial intelligence as well as law and business administration, and the opportunity to use your existing knowledge beyond your own field of activity
- The opportunity to actively shape the strategy, culture, and organization of a young company
- The chance to help build up a company with a highly innovative product in a fast-growing market
- Office in Lübeck with very good transport connections to Hamburg and the Baltic Sea
- Snacks and drinks at your free disposal in the office in Lübeck
- Beautiful office building in an old brewery with a view of the 7 towers of Lübeck

„Werkstudent*in (m/w/d) IT-Administration“

Info

You want to combine nutrition with digital medicine and deliver cutting-edge technology to patients? You want to support the development and research of innovative therapeutics to bring health to the people? And all this in a dynamic, friendly start-up?

Über Perfood

„Lass Deine Ernährung Deine Medizin sein.“ Bei Perfood entwickeln wir digitale Therapien basierend auf personalisierter Ernährung.

Perfood ist ein innovatives Start-Up, das Menschen bei der Verbesserung der metabolischen Gesundheit unterstützt. So können vielfältige gesundheitliche Leiden vorgebeugt oder behandelt werden. Unser Ursprung liegt in der universitären Forschung. Unsere Stärke liegt in der Verwendung modernster Technologien und wissenschaftlicher Erkenntnisse. Mit unserem Produkt MillionFriends haben wir ein innovatives und effektives Ernährungsprogramm geschaffen. Mit sinCephalea haben wir eine moderne und effektive Digitale Gesundheitsanwendung für die Therapie von Migräne entwickelt. In den kommenden Jahren wollen wir basierend auf unserer zum Patent angemeldeten Technologie eine Pipeline digitaler Gesundheitsanwendungen aufbauen. Dabei steht der Patient immer an erster Stelle, denn das Ziel von Perfood ist es, die Lebensqualität nachhaltig zu verbessern. Ohne Risiken und Nebenwirkungen.

Wir suchen Werkstudent*innen (m/w/d), die Lust haben mit uns die Zukunft der Ernährungsmedizin neu zu gestalten. Hierzu suchen wir ab sofort tatkräftige Unterstützung in dem Bereich IT-Administration.

Wir suchen Dich für:

- 1st & 2nd Level Support (Hard- und Software) für Mitarbeiter im Lübecker Büro und für unsere remote arbeitenden Kolleg:innen
- Incident Management via Zammad Ticketsystem
- Bereitstellung von Hardware (Laptops, Smartphones, Peripherie)
- Vorbereitung Onboardings/Nachbereitung Offboardings
- Benutzerverwaltung/Geräteverwaltung auf Basis AzureAD/MS365
- Inventar und Lagerverwaltung (Inventarverwaltung aktuell im Aufbau)

Dein Profil

Du möchtest dabei sein, große globale gesundheitliche Probleme zu lösen und Menschen wirklich zu helfen, ihre Gesundheit zu verbessern? Dich reizt die Herausforderung, einzigartige digitale Medizinprodukte zu entwickeln und verständlich zu machen? Das Ganze in einem dynamischen, sympathischen Startup in einer aufstrebenden Branche? Wir bei Perfood suchen Verstärkung für unser Team!

Was Du mitbringst

- Eingeschriebener Student (m/w/d) einer Hochschule idealerweise mit Fachrichtung Informatik
- Arbeitszeit von 10 Stunden pro Woche
- Hohes Interesse an Aspekten der Informationstechnologie
- Kenntnisse von Windows & MacOS-Betriebssystemen
- Gute Deutsch- und Englischkenntnisse in Wort und Schrift
- Zuverlässigkeit und Interesse an eigenständiger Arbeit
- Ausgeprägte Serviceorientierung

Was Dich erwartet

- Vergütung auf 520€-Basis
- Ein hochinnovatives und effektives Produkt in einer wachstumsstarken Branche
- Ein diverses und kollegiales Team, das Dich bei der Einarbeitung unterstützt
- Freie Einteilung der Arbeitszeit und Auswahl Deines Arbeitsmaterials
- Die Chance, die medizinische Versorgung zahlreicher Erkrankungen maßgeblich zu verbessern
- Die Zusammenarbeit in einem interdisziplinären Team aus den Bereichen Medizin, Ernährungswissenschaften, Data Analytics und Künstliche Intelligenz sowie Rechtswissenschaften und Betriebswirtschaft und die damit einhergehende Möglichkeit, Dein bestehendes Wissen über das eigene Aufgabenfeld hinaus zu erweitern
- Arbeitsort in Lübeck mit sehr guter Verkehrsanbindung an Hamburg und Ostsee
- Snacks und Getränke zur freien Verfügung
- Schönes Bürogebäude in einer alten Brauerei mit Blick auf die 7 Türme von Lübeck
- Teilnahme an spaßigen online und offline Teamevents

Wenn Du Interesse an der Rolle hast, dann sende uns eine E-Mail mit Deinem Lebenslauf an bewerbung@perfood.de (Betreff „**Werkstudent*in (m/w/d) IT-Administration**“). Wir freuen uns auf die gemeinsame Zusammenarbeit!

