DG epi DEUTSCHE GESELLSCHAFT FÜR EPIDEMIOLOGIE

IBS Deutsche Region

gmds Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.

Deutsche Gesellschaft für Sozialmedizin und Prävention

Workshop

Causal Inference & Estimands

21./22. November 2019

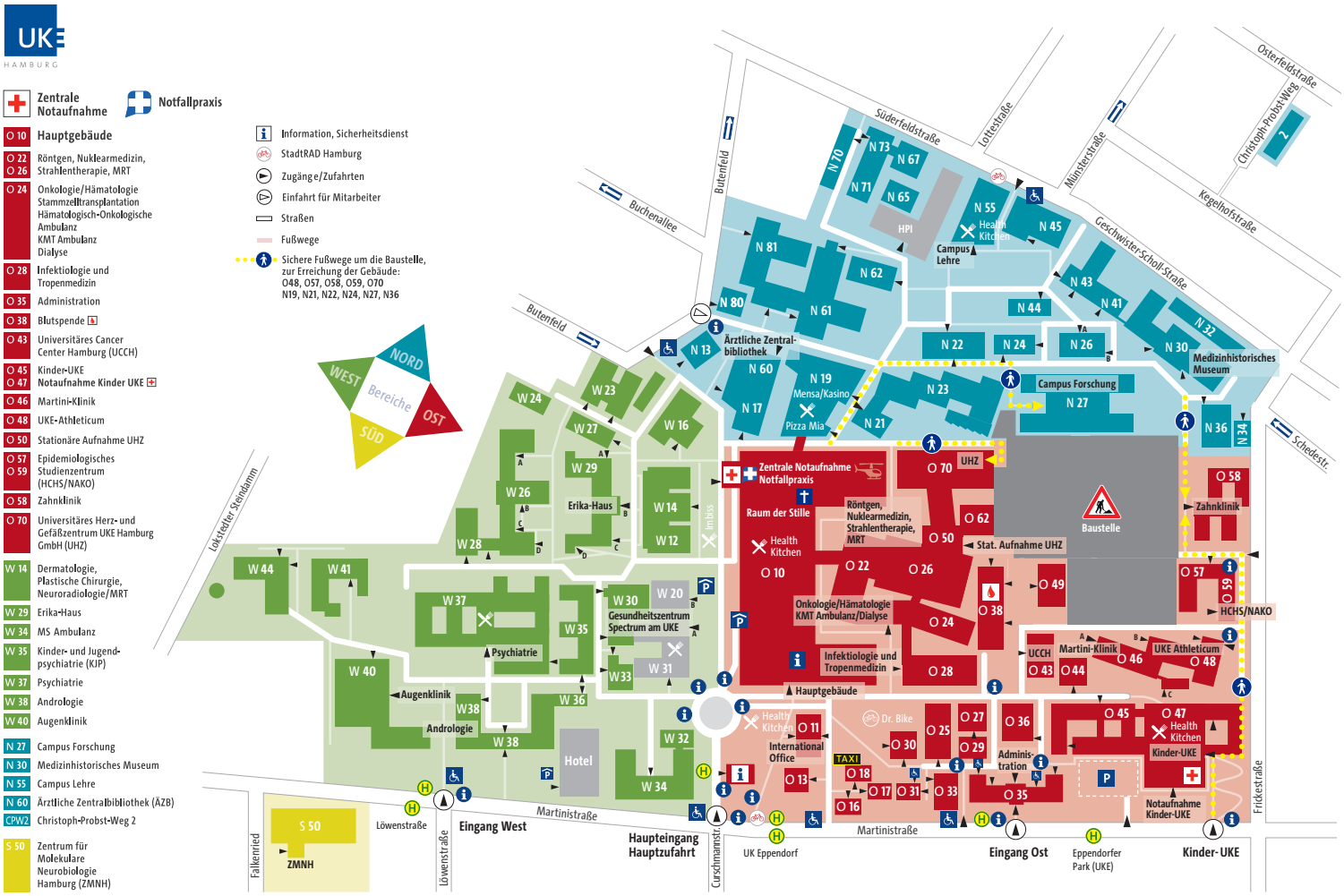
der Arbeitsgruppen “Statistische Methoden in der Medizin“ (IBS-DR), „Statistische Methoden in der Epidemiologie“ (IBS-DR, DGEpi, DGSMP), „Statistische Methoden in der klinischen Forschung“ (GMDS), „Epidemiologische Methoden“ (DGEpi, GMDS, DGSMP)

Gebäude W30, Hörsaal

Veranstaltungsort:

Universitätsklinikum Hamburg-Eppendorf
Martinistraße 52
20246 Hamburg

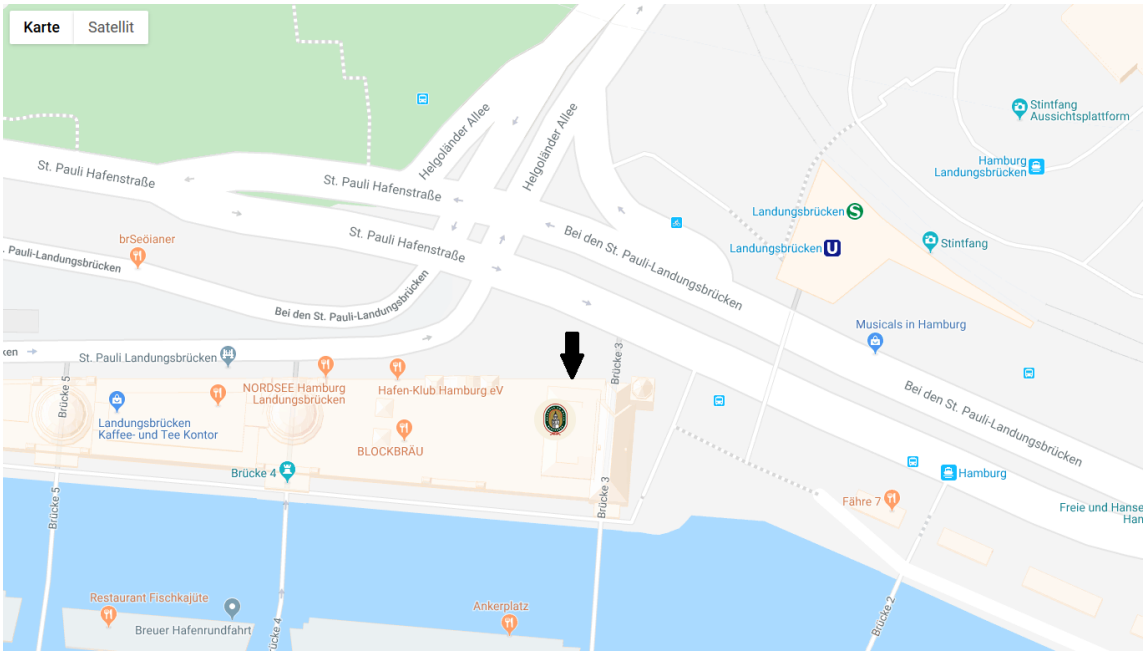
iMBE Universitätsklinikum Hamburg-Eppendorf


UKE
HAMBURG
Zentrale Notaufnahme
Notfallpraxis
O 10 Hauptgebäude
O 22 O 26 Röntgen, Nuklearmedizin, Strahlentherapie, MRT
O 24 Onkologie/Hämatologie Stammzelltransplantation Hämatologisch-Onkologische Ambulanz KMT Ambulanz Dialyse
O 28 Infektiologie und Tropenmedizin
O 35 Administration
O 38 Blutspende
O 43 Universitäres Cancer Center Hamburg (UCCH)
O 45 O 47 Kinder-UKE Notaufnahme Kinder UKE
O 46 Martini-Klinik
O 48 UKE-Athleticum
O 50 Stationäre Aufnahme UHZ
O 57 O 59 Epidemiologisches Studienzentrum (HCHS/NAKO)
O 58 Zahnklinik
O 70 Universitäres Herz- und Gefäßzentrum UKE Hamburg GmbH (UHZ)
W 14 Dermatologie, Plastische Chirurgie, Neuroradiologie/MRT
W 29 Erika-Haus
W 34 MS Ambulanz
W 35 Kinder- und Jugend-psychiatrie (KJP)
W 37 Psychiatrie
W 38 Andrologie
W 40 Augenklinik
N 27 Campus Forschung
N 30 Medizinhistorisches Museum
N 55 Campus Lehre
N 60 Ärztliche Zentralbibliothek (ÄZB)
CPW2 Christoph-Probst-Weg 2
S 50 Zentrum für Molekulare Neurobiologie Hamburg (ZMNH)
Information, Sicherheitsdienst
StadtRAD Hamburg
Zugänge/Zufahrten
Einfahrt für Mitarbeiter
Straßen
Fußwege
Sichere Fußwege um die Baustelle, zur Erreichung der Gebäude: O48, O57, O58, O59, O70 N19, N21, N22, N24, N27, N36
NORD
WEST
OST
SÜD
Bereiche
Süderfeldstraße
Lottestraße
Osterfeldstraße
Münsterstraße
Christoph-Probst-Weg
2
Kegelhofstraße
Geschwister-Scholl-Straße
Buchenallee
Butenfeld
Butenfeld
Schedestr.
Lokstedter Steindamm
Falkenried
Löwenstraße
Löwenstraße
Martinistraße
Martinistraße
Curschmannstr.
Fricktstraße
Eingang West
Haupteingang Hauptzufahrt
Eingang Ost
Kinder-UKE
UK Eppendorf
Eppendorfer Park (UKE)
HPI
Campus Lehre
Health Kitchen
Ärztliche Zentral-bibliothek
Mensa/Kasino
Pizza Mia
Campus Forschung
Medizinhistorisches Museum
Zentrale Notaufnahme Notfallpraxis
Raum der Stille
Röntgen, Nuklearmedizin, Strahlentherapie, MRT
Onkologie/Hämatologie KMT Ambulanz/Dialyse
Infektiologie und Tropenmedizin
Hauptgebäude
UHZ
Stat. Aufnahme UHZ
Baustelle
Zahnklinik
HCHS/NAKO
UCCH
Martini-Klinik
UKE Athleticum
Dr. Bike
International Office
TAXI
Administration
Notaufnahme Kinder-UKE
Erika-Haus
Psychiatrie
Augenklinik
Andrologie
Gesundheitszentrum Spectrum am UKE
Hotel
ZMNH
Intass

Abendessen im BLOCKBRÄU

Bei den St. Pauli-Landungsbrücken 3, 20359 Hamburg

https://www.block-braeu.de/



Programm:

Donnerstag, 21. November

12:30-13:15 Snacks und Registrierung

Causal Inference *(Chair: Antonia Zapf)*

13:15-13:30 Begrüßung und Vorstellung der beteiligten AGs

13:30-14:30 Vanessa Didelez (Bremen): Causal Estimands and Inference for Time-Varying Treatments

14:30-14:45 *Kaffeepause*

Causal Inference in Anwendung *(Chair: Juliane Hardt)*

14:45-15:15 Felicitas Kühne (Innsbruck): Potentielle systematische Fehler bei der Analyse von Routinedaten in der Versorgungsforschung: Ein Vergleich zwischen g-Methods mit Target Trial Emulation und traditionellen Methoden am Beispiel der Zweitlinientherapie des Ovarialkarzinoms

15:15-15:45 Tim Filla (Düsseldorf): Z-Balancing – Maximizing Covariate Balance in Propensity Score Analyses by Minimizing Weighted Z-Differences

15:45-16:15 Irene Schmidtmann (Mainz): Vergleich zwischen vier Therapien für das Leberzellkarzinom bei Vorliegen einer Portalvenen-Tumorthrombose – eine Propensity Score Analyse von Registerdaten

16:15-16:45 *Kaffeepause*

Analyse nicht-randomisierter Studien *(Chair: Daniela Adolf)*

16:45-17:15 Elke Schmitt (Frankfurt): Comparing different statistical analysis methods for a non-randomized controlled observational retrospective cohort study

17:15-17:45 Sven Kleine Bardenhorst (Münster): A critical methodological review of analytic strategies in microbiome studies

17:45-18:15 Dirk Schomburg (Magdeburg) Ansätze für die funktionelle Magnetresonanztomographie-Dekodierung in Echtzeit mit multivariaten Methoden

Gemeinsames Abendessen ab 19:30 Uhr im BLOCKBRÄU (Bei den St. Pauli-Landungsbrücken 3, 20359 Hamburg; Direkt bei der Haltestelle „Landungsbrücken“)

Hinweis zur Anreise: Aufgrund von Umbaumaßnahmen fährt die U3 derzeit nicht die Haltestelle „Landungsbrücken“ an. Daher empfohlene Verbindung ab „UK Eppendorf“: Bus 25 bis „Kellinghusenstraße“ – U1 ab „Kellinghusenstraße“ bis „Jungfernstieg“ – S3 ab „Jungfernstieg“ bis „Landungsbrücken“

Mehr Informationen auch unter: *https://www.hvv.de/de*

Freitag, 22. November

8:30-9:15	Gemeinsame AG-Sitzung inkl. Sprecherwahl

Estimands *(Chair: Sigrid Behr)*

9:15-10:15	Mouna Akacha (Basel): Estimands in Clinical Trials – Broadening the Perspective
10:15-11:15	Panel-Diskussion Causal Inference & Estimands: Mouna Akacha, Vanessa Didelez, Lars Beckmann, Uwe Siebert
11:15-11:30	*Kaffeepause*

Statistische Methoden in der Medizin *(Chair: Philipp Mildenberger)*

11:30-12:00	Edgar Brunner (Göttingen): Win-Ratio und Mann-Whitney-Odds
12:00-12:30	Maria Stark (Hamburg): Neuberechnung der Fallzahl in einer zweiarmig gepaarten Diagnosegütestudie
12:30-13:00	Eric Bibiza (Hamburg): Seamless study design in diagnostic studies
13:00-13:30	Veranstaltungsabschluss mit Snacks und Getränken

Causal Estimands and Inference for Time-Varying Treatments

Vanessa Didelez

Leibniz Institute for Prevention Research and Epidemiology – BIPS and Faculty of Mathematics & Computer Science, University of Bremen

The field of causal inference deals with approaches, models and methods, for investigating the effects of (specific or hypothetical) interventions, typically based on data where these interventions have not actually, or only imperfectly, been carried out. Logically, the first step of such an investigation is to clearly define the target of inference, aka *causal estimand*. In this presentation, I will focus on the fact that, in biomedical and epidemiological research, most treatments or exposures are time-varying. Hence relevant target interventions will often need to be time-varying, too. I believe that this applies to many RCTs, in particular when faced with intercurrent events, and it certainly applies to most epidemiological studies using observational data. I will give an overview of and discuss the ensuing issues, as well as ways to address them.

It is relatively easy to decide on a target of inference for a binary treatment/exposure; often this is simple contrast like the average causal effect or the causal risk ratio, with further refinements to subgroup effects or comparison of survival curves. In case of time-varying treatments/exposures, specifying the target of inference is typically more challenging. Even if we simply wished to compare hypothetical interventions like, say, “always-treat” versus “never-treat”, thus apparently reverting to a binary treatment, the fact that over time patients will not comply with these two options, possibly for good reasons, must be taken into account. So, while ideally, the choice of target of inference should be dictated by the research question or the decision problem at hand, we may often wish to allow for what is actually feasible in practice. The translation of the research question into a target of inference, or estimand, is a key issue, and should be an explicit part of any investigation (experimental or observational). I will argue and illustrate how, in many situations, it makes sense to consider dynamic or adaptive treatment strategies, and we may even aim at finding the optimal strategy. Though less obvious, this is often also relevant in epidemiological studies with time-varying exposures. Formulation of, and inference for, such treatment strategies is not new: see Robins (1986) and the many follow-up papers since, e.g. Dawid & Didelez (2010); also see, for instance, the monograph by Chakraborty & Moodie (2013) on optimal adaptive treatments. Renewed interest in this topic is due to the debate around the (draft) ICH E9 addendum on estimands, by which the analysis of clinical trials has been opened up to causal inference approaches originally aimed at observational data. On the observational studies side, interestingly, there is a recent push to use the “target trial” principles for analyzing observational data (Cain et al., 2016); these principles can be regarded as a systematic guide to formulating sensible causal estimands, and then carrying out the appropriate inference. Both developments appear like two sides of the same coin: an increasing awareness of the general need to be explicit about the target of inference, to design the study towards this aim, and to use suitable models and methods. This ensures a principled and coherent statistical analysis with practically useful results, where avoidable biases are avoided while plausibly minimizing biases that are not avoidable.

Cain et al. (2016). Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. International Journal of Epidemiology.

References:

Chakraborty, Moodie (2013). Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference and Personalized Medicine. Springer.

Dawid, Didelez (2010). Identifying the consequences of dynamic treatment strategies: A decision theoretic overview. Statistics Surveys.

Robins (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker effect. Mathematical Modelling.

Potentielle systematische Fehler bei der Analyse von Routinedaten in der Versorgungsforschung: Ein Vergleich zwischen g-Methods mit Target Trial Emulation und traditionellen Methoden am Beispiel der Zweitlinientherapie des Ovarialkarzinoms

Felicitas Kühne[1] (Doktorandin), Uwe Siebert[1] (Betreuer)

Danksagung für die Mitwirkung an: M. Arvandi, L. Hess, D.E. Faries, R. Matteucci Gothe, H. Gothe, J. Beyrer, A.G. Zeimet, C. Marth

[1]Institut für Public Health, Medical Decision Making und HTA

Ziel: Ziel der Studie war die Untersuchung von typischen systematischen Fehlern (Bias), die bei der Analyse von großen Routinedatensätzen auftreten. Als Fallbeispiel diente die Zweitlinientherapie des Ovarialkarzinoms nach Progression.

Methoden: Als Datenmaterial dienten ein Versicherungs-Routinedatensatz US-amerikanischer Frauen mit Ovarialkarzinom (n=1.581) sowie publizierte Ergebnisse einer randomisierten klinischen Studie (RCT) (Rustin et al. 2010).

Zunächst wurden a-priori die für Routinedatenauswertungen typischen Biastypen festgelegt: Informationsfehler/fehlende Daten, Selektionsfehler, zeitunabhängiges/zeitabhängiges Confounding, Immortal-Time-Bias, unklare Zuordnung von Strategie(-start). Anschließend wurde spezifisch für das Fallbeispiel ein gerichteter azyklischer Graph (DAG) als Kausaldiagramm erstellt. Anhand des DAGs wurden die Parameter festgelegt, die bei der kausalen statistischen Analyse berücksichtigt werden sollten. Um die Existenz, Richtung und Größe möglicher Biases in Abhängigkeit von der statistischen Analysestrategie zu untersuchen, wurden die Ergebnisse verschiedener traditioneller und g-methods-basierter Analysestrategien mit dem RCT-Ergebnis verglichen. Die untersuchten Analysestrategien umfassten „traditionelle“ Cox-Modelle (nicht adjustiert, mit Baseline-Confoundern, mit zeitabhängigen Kovariablen) sowie einer Kausalanalyse mit einem Marginal Structural Cox-Modell (MSCM) mit zeitabhängigen Kovariablen basierend auf dem Target Trial Ansatz mit „Replikationen“ und inverse-probability of censoring weighting (IPCW). Das betrachtete Effektmaß war die Hazard Ratio (HR) mit 95%-Konfidenzintervall (95%KI) für den Vergleich mit vs. ohne Zweitlinientherapie.

Ergebnisse: Bei der rohen und baseline-adjustierten Cox-Analyse ergaben sich HRs von 0,57 (95%KI 0,50-0,65) und 0,54 (95%KI 0,47-0,61). Durch Einschluss zeitabhängiger Kovariablen zur Kontrolle des Immortal-Time-Bias erhöhten sich die HRs substanziell auf 1,67 (95%KI 1,46-1,90) und 1,68 (95%KI 1,41-2,01) bei roher und baseline-adjustierter Analyse. Das kausale MSCM mit „Target Trial Approach“ und IPCW zur Kontrolle zeitabhängiger Confounder ergab eine HR von 1,07 (95%KI 1,02-1,12), was dem Ergebnis des RCT mit 1,01 (95%KI 0,82-1,25) sehr nahe kommt und im 95%KI des RCT liegt.

Schlussfolgerung: Bei der Analyse von Routinebeobachtungsdaten kann die Anwendung von DAGs bei der Visualisierung und Identifizierung von potenziellen Biases unterstützen sowie die Entscheidung erleichtern, welche Variablen und strukturellen Ansätze bei der Datenanalyse berücksichtigt werden sollten und welche nicht. Das Fallbeispiel zeigt, dass je nach Analysestrategie verschiedene substanzielle Biases mit verschiedenen Richtungen auftreten können und dass mit einer Kausalanalyse mit Target Trial Ansatz und g-Methods das Ergebnis einer entsprechenden randomisierten klinischen Studie gut angenähert werden kann, sofern alle wesentlichen Confounder erfasst wurden.

Z-Balancing – Maximizing Covariate Balance in Propensity Score Analyses by Minimizing Weighted Z-Differences

Tim Filla[1], Oliver Kuss[1,2]

[1]Institute of Medical Statistics, Medical Faculty, Heinrich Heine University Düsseldorf
[2]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf

Propensity score (PS) analyses are becoming the quasi-standard for analyzing non-randomized studies of treatment effects, which is due to their statistical and epistemological advantages as compared to standard outcome regression. PS analyses are performed in two steps. In the first step, the propensity score, defined as the probability of the treatment conditional on the subject's covariates, is estimated. In the second step, this PS is used for estimating the treatment effect, which is the actual quantity of interest.

The ultimate aim in the first step model is balancing covariates in the treatment groups (1, 2). However, the methods regularly used to this task, standard logistic regression or more evolved machine learning methods, aim for optimizing the prediction for the respective outcome, effectively ignoring covariate balance. As such, methods have been proposed which explicitly aim for minimizing covariate balance in first step models (3, 4), and they have shown to be superior to standard models in simulations.

We here propose z-balancing, a new method which uses the idea of minimizing weighted z-differences (5) as an optimality criterion for covariate balance in first step models. This method improves on previous methods by modelling all covariates on their original (continuous, binary, ordinal, or nominal) scale. In addition, not only standard inverse probability weights can be used (which frequently have problems with extreme weights compromising model fits), but also other weights, e.g., the recently proposed matching weights (6). In the talk we introduce the method and present first simulation results that compare z-balancing with its competing methods.

References:

(1) Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. Pharmacoepidemiol Drug Saf. 2011 Nov;20(11):1115-29.
(2) Kuss O, Blettner M, Börgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl Int. 2016 Sep 5;113(35-36):597-603.
(3) Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Polit Anal. 2012;20(1):25-46.
(4) Imai K, Ratkovic M. Covariate balancing propensity score. J R Statist Soc B. 2014;76(1):243–263.
(5) Filla T, Kuss O. The weighted z-difference can be used to measure covariate balance in weighted propensity score analyses. Submitted.
(6) Li L, Greene T. A weighting analogue to pair matching in propensity score analysis. Int J Biostat. 2013 Jul 31;9(2):215-34.

Vergleich zwischen vier Therapien für das Leberzellkarzinom bei Vorliegen einer Portalvenen-Tumorthrombose – eine Propensity Score Analyse von Registerdaten

Irene Schmidtmann[1], Verena Steinle[2, 3, 4], Aline Mähringer-Kunz[2], Roman Kloeckner[2], Arndt Weinmann[4, 5]

[1]Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin Mainz
[2]Klinik und Poliklinik für Diagnostische und Interventionelle Radiologie, Universitätsmedizin Mainz
[3]Klinik und Poliklinik für Diagnostische und Interventionelle Radiologie, Universitätsklinikum Heidelberg
[4]I. Medizinische Klinik und Poliklinik, Universitätsmedizin Mainz
[5]Clinical Registry Unit, Universitätsmedizin Mainz

Eine Portalvenen-Tumorthrombose (PVTT) ist eine häufige Komplikation des Leberzellkarzinoms (HCC) und führt dazu, dass der Tumor als fortgeschritten klassifiziert wird. Der klinische Standard nach BCLC (Barcelona Clinic Liver Cancer)-Klassifikation für dieses Stadium ist eine palliative systemische Therapie, zumeist mit Sorafenib oder Lenvatinib in der Erstlinie. In der klinischen Realität erhalten diese Patienten jedoch unterschiedliche Therapien, von Resektion über transarterielle Chemoembolisation (TACE) bis hin zu „Best supportive care“ (BSC).

Das HCC-Register an der Universitätsmedizin Mainz dokumentiert seit 1998 alle behandelten Leberzellkarzinome, ihre Diagnose, Therapie und das Follow-Up. Daten aus diesem Register wurden herangezogen, um zu untersuchen, ob Patienten mit PVTT im Hinblick auf das Überleben von Therapien profitieren könnten, die für weniger fortgeschrittene Stadien empfohlen werden. Da es sich hier nicht um eine randomisierte Studie handelt und damit gerechnet werden muss, dass das Ausmaß der PVTT sowie weitere Patientenmerkmale wie die Tumorgröße oder der Grad der Leberzirrhose einen Einfluss sowohl auf die Therapieentscheidung als auch auf das Überleben hatten, ist in geeigneter Weise für Confounding zu adjustieren.

Wir haben mehrere Analyseverfahren verglichen, um den „average treatment effect“ (ATE) zu schätzen:

- klassische Adjustierung für potenzielle Confounder im Cox-Regressionsmodell,
- Cox-Regression mit Therapie als Kovariable und Propensity-Score-Gewichtung, wobei die Wahrscheinlichkeit eine bestimmte Therapie zu erhalten in Abhängigkeit von den Confoundern mittels logistischer Regression für multinomiale Outcomes bestimmt wurde,
- ein doppelt robustes Verfahren, bei welchem sowohl adjustiert als auch gewichtet wurde.

Außerdem haben wir für paarweise Vergleiche zwischen je zwei Therapien den „average treatment effect of the treated“ (ATT) geschätzt – sowohl für alle Patienten, die eine der beiden Therapien erhielten als auch für die durch das PVTT-Stadium definierten Untergruppen.

Ignoriert man das PVTT-Stadium, scheinen die Patienten, die sich einer Resektion unterzogen, bessere Überlebenschancen zu haben also solche, die mit TACE oder Sorafenib behandelt

wurden. Wenn das Ausmaß der PVTT berücksichtigt wurde, relativierte sich dies. Unterschiede zeigten sich vor allem im frühen Stadium der PVTT.

Nicht unerwartet wurden gegenüber einer naiven Analyse die meisten Effekte schwächer (Hazard Ratio näher an 1), wenn für Confounder adjustiert oder eine Propensity-Score-Gewichtung angewandt wurde. Die Propensity-Score-Gewichtung ergab überwiegend schwächere Effekte als die Adjustierung. Die doppelt robuste Analyse ergab zumeist Punktschätzer, die zwischen denen aus den nur adjustierten oder nur gewichteten Analysen lagen. Allerdings waren hier die Konfidenzintervalle am breitesten.

Wir präsentieren die Ergebnisse unserer Analyse und gehen auf die speziellen Probleme bei der Gewichtung ein, wenn mehr als zwei Therapien verglichen werden sollen.

Dieses Abstract enthält Teile der Dissertation von Verena Steinle.

Comparing different statistical analysis methods for a non-randomized controlled observational retrospective cohort study

Elke Schmitt[1, 2], Patrick Meybohm[1], Eva Herrmann[2], Karin Ammersbach[3], Raphaela Endres[1], Simone Lindau[1], Philipp Helmer[1], Kai Zacharowski[1], Holger Neb[1]

[1]Department of Anaesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Frankfurt, Goethe University Frankfurt, Germany
[2]Institute of Biostatistics and Mathematical Modelling, Department of Medicine, Goethe University Frankfurt, Frankfurt, Germany
[3]Division of Software and Information Systems, Department of Information and Communication Technology, University Hospital Frankfurt, Germany

Background: The potential harmful effects of particle-contaminated infusions for critically ill adult patients are yet unclear. A single-centre non-randomized retrospective cohort study was conducted to investigate the effect of in-line filtration of intravenous fluids with finer 0.2 or 1.2 µm vs larger control 5.0 µm filters on the reduction of complications in critically ill adult patients (Schmitt et al, 2019, in press).

Even so randomized controlled trials are considered as optimal for causal inference; observational cohort studies have the advantage to evaluate a more representative patient cohort in clinically relevant settings. Nevertheless, it is necessary to account for systematic differences in baseline characteristics when comparing treatment effects between the patient groups (Kuss, 2016; Austin, 2011). Propensity score methods are of increasing importance for reducing the effects of confounding. However, there is still an ongoing debate whether propensity score analysis should use matched or unmatched statistical tests (Wan, 2019; Austin, 2011).

Methods: We accounted for differences in patient characteristics in the total cohort by propensity score matching with respect to surgery group, sex and age (n = 1506 in the fine filter and in the control filter cohort). In our clinical manuscript, we used unmatched non-parametric tests and multivariable regression models to compare different in-hospital endpoints between the two cohorts. Here, we evaluate whether the results are robust when matched and unmatched tests are used and also analyse the adjustment effects.

Results: The three approaches (unadjusted, adjusted unmatched and matched) were compared and, overall, similar results are received. Adjusted analysis with matched and unmatched tests showed significantly or nearly significantly better therapy effects in inflammation, respiratory dysfunction, pneumonia, sepsis, ICU and hospital stay for the finer filter cohort while no relevant differences could be detected for renal dysfunction, brain dysfunction, mortality and vasoplegia. In addition, unadjusted analysis failed to receive significant results for inflammation.

Conclusions: Matched or unmatched tests can lead to comparable results in propensity score adjusted analysis. In our filter study, the advantage of the finer filters on organ dysfunction and less inflammation is supported in a robust way in critically ill adult patients.

A critical methodological review of analytic strategies in microbiome studies

Sven Kleine Bardenhorst[1], André Karch[1], and Nicole Rübsamen[1]

[1]Institute of Epidemiology and Social Medicine, University of Münster

Introduction: Recent advantages of high-throughput sequencing methods led to an exponentially increasing number of publications focusing on the study of the human microbiome and its associations with various diseases. However, reproducibility is a major issue in microbiome studies, mainly caused by missing consensus about analytic strategies. The complex nature of microbiome data—high-dimensional, zero-inflated and compositional—prohibits the use of classical statistical methods. Furthermore, while research in the context of the human microbiome advances, research questions become more focused and the accompanying study designs increase in complexity. To improve the statistical toolbox for the analysis of microbiome data, it is important to have a clear picture of what kind of questions are asked and which tools are needed to adequately answer these questions.

Methods: We conducted a review considering the 750 most recent publications that focused on the analysis of human microbiome data, among which 164 were identified as clinically relevant (i.e., concerning biosamples collected in a clinical context). Information about research questions, study designs, and analytic strategies was extracted from the selected publications.

Results: The results of our review confirmed the expected shift to more advanced and focused research questions, as one-third of the studies faced the analysis of clustered data (either because of longitudinal sampling or samples from several body sites of the same individual). While most studies investigated group differences, only one study investigated a time-to-event outcome. Although there was a standard workflow (α- and β-diversity followed by differential abundance testing), heterogeneity existed at each stage of this workflow, ranging from 11 alpha diversity measures to 42 different approaches to test for differential abundance.

Discussion: Our study found heterogeneity in analysis strategies among microbiome studies. Increasingly complex research questions magnify the impact of these inconsistencies. Our results point out the need for an extensive evaluation of the strengths and shortcomings of existing methods to guide the choice of proper analytic strategies. The results also serve as a compass for the development of novel methods that are thoroughly designed to address more advanced research questions while taking into account the complex structure of the data.

Ansätze für die funktionelle Magnetresonanztomographie-Dekodierung in Echtzeit mit multivariaten Methoden

Dirk Schomburg[1] und Johannes Bernarding[1]

[1]Institut für Biometrie und Medizinische Informatik, Medizinische Fakultät, Otto-von-Guericke Universität Magdeburg

Einleitung

Die funktionelle Magnetresonanztomographie (fMRT) ermöglicht es, verschiedene durch Stimuli oder Aufgaben stärker aktivierte Hirnareale zu detektieren. Die fMRT basiert auf unterschiedlichen magnetischen Eigenschaften des Blutes vor und nach Sauerstoffverbrauch durch Neuronenaktivität. Das gemessene Signal wird als blood-oxygen-level-dependend (BOLD) Signal bezeichnet. Die Auswertung kann mit angepassten Methoden bereits während der Messung erfolgen (Echtzeit-fMRT).

Mit der Echtzeit-fMRT-Dekodierung soll aus dem Wissen über die bedingt erwartet stärker aktivierten Hirnareale und dem ortsaufgelöst gemessenen BOLD-Signal auf die aktuelle Hirnfunktion geschlossen werden. Eine Vormessung dient in der Regel als Trainingsphase für die Parameter, mit denen die Prädiktions-Funktionen für die fMRT-Dekodierung angepasst werden. Mit multivariaten Methoden soll das Prädiktionsrisiko verringert werden, um auch bei schwachen neuronalen Aktivitäten zu dekodieren.

Wegen der großen Anzahl an Voxeln (>150.000) und der kleinen Stichprobengröße (100-500) kann aus den unveränderten klassischen Regressionsverfahren kein Prädiktor gewonnen werden. Moderne regularisierte Regressionsmethoden benutzen ein inverses Modell, dessen zufällige Fehler nicht zu interpretieren sind. Außerdem führt die dabei nötige Bestimmung der Hyperparameter zu stark erhöhtem Trainingsrechenaufwand. Verschiedene Anpassungen dieser Methoden ermöglichen dennoch deren Verwendung. Ziel der Untersuchung war ein Vergleich verschiedener Methoden.

Zusätzlich sind die Messwerte oft von einem zufälligen Trend überlagert, für den eine lineare Modellierung nicht ausreicht. Übliche Trendbereinigungsmethoden verwenden die gesamte Messzeitreihe; dies ist hier nur in der Trainingphase möglich, während für die Echtzeitdekodierung eine Trendberücksichtigung erforderlich ist, die nur die bis zum Zeitpunkt der Messung akquirierten Messwerte benötigt.

Methoden

In Simulationsrechnungen wurden für die ad-hoc-Methode, für ein stabilisiertes klassisches Verfahren, für ein regularisiertes modernes Verfahren und für die Hauptkomponentenregression jeweils die Rechenzeit und das Prädiktionsrisiko für verschiedene Szenarien berechnet und verglichen.
Zur Trendberücksichtigung wurde das Modell additiv mit einem Gaußschem Random-Walk ergänzt. Restricted-Maximum-Likelihood-Schätzer ergeben die Varianzkomponenten. Damit führt ein Pre-Whitening der Daten zu Daten mit homoskedastischen Fehlern, für die die Regressionsmethoden anwendbar sind. Die Prädiktoren mit den so erhaltenen Modell-Parametern ergeben trendbehaftete Zeitreihen, deren Trendbereinigung mit weiterem Vorwissen über die Stimulationszeitreihen möglich wurde.

Ergebnisse

Als bester Kompromiss zwischen Trainingsrechenzeit und Prädiktionsrisiko hat sich hier die Hauptkomponentenregression gezeigt. Trendberücksichtigungsmethoden für Training und Dekodierung konnten gefunden werden. Simulierte und gemessene Daten zeigten ein gutes Ergebnis für den entwickelten Ansatz zur fMRT-Dekodierung.

Estimands in Clinical Trials – Broadening the Perspective

Mouna Akacha

Novartis Pharma AG, Basel, Switzerland

In this talk we will provide background on the emerging topic of 'estimands' which is at the heart of the draft addendum to the ICH E9 guideline – the holy grail of pharmaceutical statistics. Broadly speaking, an estimand represents 'WHAT' is to be estimated in order to address the scientific question of interest. In contrast, an estimator is a statistical function that represents 'HOW' to estimate the estimand from the data, presumably with reasonable statistical properties.

Established practices in pharmaceutical research suggest that relatively more focus has been paid to the 'HOW' rather than to the 'WHAT'. The aim of the planned addendum is to promote harmonized standards on the choice of so-called estimands (the 'WHAT') in clinical trials.

The discussions around the addendum have also resulted in questioning the traditional so-called intention-to-treat (ITT) principle – the 'steadfast beacon in the foggy vistas of biomedical experimentation' (Efron , 1998) . Inherently, the topic of estimands is closely related to the occurrence of post-randomization events, e.g. non-compliance and discontinuation of study treatment.

In this presentation, we will introduce the estimand framework as presented in the draft ICH E9 addendum and create the link to the ITT principle as well as to causal inference.

Win-Ratio und Mann-Whitney-Odds

Edgar Brunner

Institut für Medizinische Statistik, Universitätsmedizin Göttingen,
Humboldtallee 32, 37073 Göttingen, Germany
E-mail: ebrunne1@gwdg.de

Von Pocock et al. (2012) wurde zur Bewertung von kombinierten Endpunkten das Win-Ratio eingeführt. Dadurch wird ein nichtparametrischer, gut zu interpretierender Effekt definiert, der bei zwei unverbundenen Stichproben einen Behandlungseffekt beschreibt. Bezeichne $X \sim F_1(x)$ den Messwert in der Behandlungsgruppe und $Y \sim F_2(x)$ den Messwert in der Vergleichsgruppe. Dann ist das Win-Ratio definiert als

$$\lambda_{WR} \;=\; \frac{P(X > Y)}{P(X < Y)}\,,$$

also ein Verhältnis der *Gewinn-Wahrscheinlichkeit* zur *Verlustwahrscheinlichkeit*. Weiterhin führten Wang und Pocock (2016) dieses Win-Ratio als Maß für einen Behandlungseffekt bei 'nicht-normalverteilten Daten' ein. Dabei sind Bindungen, d.h. $P(X = Y) > 0$ ausdrücklich zugelassen. Dies ist damit ein nichtparametrisches Effektmaß, das für stetige und diskrete metrische Daten, für ordinale und sogar für dichotome Daten einen sinnvoll, gut zu interpretierenden Effekt beschreiben soll.

Das Mann-Whitney-Odds $\lambda_{MW} = (1-p)/p$ für $p = P(X < Y) + \frac{1}{2}P(X = Y)$ wurde zum ersten Mal von Noether (1987) in der Literatur erwähnt und unglücklicherweise als Odds-Ratio $r = (1-p)/p$ bezeichnet.

Inhaltliche Probleme des Win-Ratio λ_{WR} sind bereits mehrfach in der Literatur diskutiert worden. Daher soll hier nicht mehr auf diese Probleme eingegangen werden. Vielmehr sollen methodische und statistische Aspekte des Win-Ratio λ_{WR} und des Mann-Whitney-Odds λ_{MW} als auch deren Eigenschaften als Effektmaße, insbesondere bei Bindungen untersucht werden. Einfache allgemeine Beziehungen zwischen diesen Größen werden aufgestellt. Die asymptotischen Verteilungen von einfachen Schätzern für λ_{WR} und λ_{MW} und die Konstruktion von Konfidenzintervallen werden ebenfalls kurz diskutiert. Das wesentliche Fazit soll in diesem Abstract nicht vorweggenommen werden.

References

Pocock, S. J. et al. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal* **33**, 176–182.

Wang, D., Pocock, S. J. (2016). A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharmaceutical Statistics* **15**, 238–245.

Neuberechnung der Fallzahl in einer zweiarmig gepaarten Diagnosegütestudie

Maria Stark[1] und Antonia Zapf[1]

[1]Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf

In einer zweiarmig gepaarten Diagnosegüstestudie wird ein neuer experimenteller Test mit einem bereits existierenden Komparatortest an denselben Patienten verglichen. Der Goldstandard definiert hierbei den wahren Gesundheitszustand der Patienten. Somit wird jeder Patient insgesamt drei diagnostischen Verfahren unterzogen. Die initiale Berechnung der Fallzahl in einer konfirmatorisch zweiarmig gepaarten Diagnosegüstestudie beruht u.a. auf den Annahmen zur Prävalenz und zum Anteil der diskordanten Ergebnisse zwischen dem experimentellen und dem Komparatortest (Flahault 2005, Miettinen 1968).
Um diese Annahmen im Verlauf der Studie zu überprüfen, wird ein adaptives Design für eine zweiarmig gepaarte Diagnosegütestudie vorgestellt. Es beruht auf der Neuschätzung der Prävalenz sowie des Anteils der diskordanten Ergebnisse und dient zur Neuberechnung der Fallzahl. Somit kann eine unter- bzw. überpowerte Studie vermieden werden. Da die Sensitivität und Spezifität des experimentellen und Komparatortests nicht neu geschätzt werden, handelt es sich um ein verblindetes adaptives Design.
Das adaptive Design wird mit Hilfe einer Beispiel- und Simulationsstudie illustriert. Aufgrund des verblindeten Charakters, wird erwartet, dass keine Adjustierung des Fehlers 1. Art notwendig ist. Der Fehler 1. Art, die Power und die Gesamtfallzahl des adaptiven Designs werden mit denjenigen eines fixen Designs verglichen.

Quellenangaben

Flahault, A., Cadilhac, M., & Thomas, G. (2005). Sample size calculation should be performed for design accuracy in diagnostic test studies. Journal of clinical epidemiology, 58(8), 859-862.

Miettinen, O. S. (1968). The matched pairs design in the case of all-or-none responses. Biometrics, 339-352.

Seamless study design in diagnostic studies

Eric Bibiza[1] and Antonia Zapf[1]

[1]Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf

In clinical research, seamless designs provide a way to test two phases of clinical development within a study. The advantage over a conventional design is that part of the data used in the first phase can also be used in the second phase. Here you can reduce the total number of cases and thus the costs.

Currently, seamless designs are mainly used in pharmacological study theses. In other clinical areas, such as diagnostic studies, this type of study design has so far received little attention, and methodological research is only just beginning. In addition, a simple transfer of the methodology from the pharmacological to the diagnostic studies is not possible, as these differ in some cases considerably in the area of the basic design and the evaluation approach.

The subject of this research project is the question of how two different phases of the diagnostic development can be linked within a single study without compromising the integrity of the analysis. One hurdle here is the possible design diversity of phase III and IV, as well as the different populations in phases II, III and IV, which virtually exclude a blanket method of connection. Therefore, it is necessary to design the possible application of seamless combinatorics depending on the target parameters used and the design template within the two phases. In particular, it must be examined to what extent the two required populations are congruent or how the differences can be mathematically or methodically taken into account so that the bias and the Type I error can be controlled.

Within the thesis different combinations of seamless designs in diagnostic studies will be investigated. Here, various compounds of Phase II and Phase III, as well as Phase III and Phase IV will be analyzed. The selection of design templates follows primarily content-related aspects with regard to the required congruence and the possible correction of incongruences. The selected model is then simulatively examined for the generating bias and the Type I error.

Within the talk, the presentation of the results obtained so far takes place. This includes the possible combinations of the different designs of the phase in question and their modifications to guarantee a congruence of the populations.