Invited talk 1:

Do you believe in causes? The distinction between causality and causal inference

Miguel Hernan

Harvard T.H. Chan School of Public Health

The terms “causality” and “causal inference” are often used interchangeably. However, these two terms can be naturally used to differentiate two forms of causal research: (i) causality as the identification of causes and (ii) causal inference as the quantification of causal effects. Both forms of causal research generally require the combination of data and untestable assumptions but, as discussed here, the distinction between causality and causal inference has implications for the identification, estimation, and interpretation of causal estimates. Importantly, causal inference requires sufficiently well-defined “causal effects” but remains agnostic about the meaning of the word “cause”.

Emulating a randomised trial from registry data to evaluate the effect of surgical treatment on survival among older lung cancer patients: results and methodological challenges

Clémence Leyrat, Camille Marine C, Sara Benitez Majano, Matthew Smith, Aimilia Exarchakou, Bernard Rachet, Aurélien Belot

Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

Older cancer patients often have suboptimal cancer treatment and poorer cancer outcomes than younger patients: non-small cell lung cancer (NSCLC) patients diagnosed after 70 years of age in England show a hugely reduced probability of receiving major surgery compared to younger patients [1]. As older cancer patients are generally excluded from clinical trials, the evidence supporting aggressive cancer management is scarce and has to rely on non-randomised studies that are more likely to be prone to bias. For instance, when using observational data to estimate the causal effect of surgery on survival, immortal-time bias is an issue because of the waiting-time period between diagnosis and surgery.

To measure the causal effect of receiving a major surgery for lung cancer patients aged 70+ at diagnosis on 1-year survival probability, we emulated a randomised trial using population-based cancer registry data, linked to electronic health records [2].

The inclusion criteria led to the selection of 1535 patients. The intervention arm was surgery within six months following diagnosis compared to no surgery within 6 months. At diagnosis, patients were cloned and entered both arms. Observations were censored when they deviated from the protocol, and this dependent censoring was accounted for using inverse-probability-of-censoring weights [3] when estimating non-parametrically survival probabilities.

We will illustrate the performance of this approach to estimate the causal effect of receiving surgery on 1-year survival. Results based on different models for estimating the censoring weights will be provided and compared to a g-computation method based on flexible hazard-based regression models.

Handling time-dependent exposures and confounders when estimating attributable fractions – bridging the gap between multi state and counterfactual modeling

Johan Steen¹, Stijn Vansteelandt²,³, Johan Decruyenaere¹

¹Department of Intensive Care, Ghent University Hospital, Belgium
²Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
³Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

The population attributable fraction (PAF) expresses the proportion of events that could have been prevented by eliminating a certain exposure in a certain population. In some contexts, this quantity can be strongly time-dependent because either exposure incidence or the excess risk changes over time. Moreover, in hospital epidemiology, competing events (such as hospital discharge) may prevent the outcome of interest (in-hospital death) and occurrence of either of these events may, in turn, hinder the exposure of interest (hospital-acquired infection; HAI). Estimation approaches thus need to carefully account for the timing of events in such highly dynamic settings. The use of multi state models (MSMs) has been widely encouraged to meet this need so as to eliminate preventable yet common types of bias (e.g. failure to appropriately account for competing endpoints and so-called immortal time bias). However, accurate assessment of the PAF hinges on unbiased estimation of the counterfactual exposure-free cumulative incidence (the cumulative incidence that would have been observed if, contrary to the fact, HAI could have been prevented in all patients). We illustrate, using a toy example as well as a substantive example from hospital epidemiology, that also suggested MSM approaches fail to produce unbiased estimates of this quantity, even in the absence of confounding. This can most easily be understood by framing estimation as a three step procedure which involves cloning, censoring and weighing observations. Closer inspection reveals that these MSM approaches may even re-introduce a new (albeit weaker) form of immortal time bias.
Separable effects: New estimands for causal inference in competing risk settings

Mats J. Stensrud, Jessica G. Young, Vanessa Didelez, Miguel A. Hernan

Department of Epidemiology, Harvard T.H. Chan School of Public Health
Department of Biostatistics, University of Oslo

In time-to-event settings, a competing event is any event that makes it impossible for the event of interest to occur. In this setting, the total effect of a treatment on the event of interest is hard to interpret as it may partially capture the treatment effect on the competing event. Therefore, direct effects that do not capture treatment effects on the competing event are of great interest. Previous definitions of direct (and indirect) effects in this setting have problematic interpretations, either requiring conceptualizing cross-world counterfactuals or intervention on competing events. Here we propose the new separable effects, which are inspired by Robins and Richardson’s extended graphical approach (1). The separable direct effect is the treatment effect on the event of interest not mediated by its effect on the competing event. The separable indirect effect is the treatment effect on the event of interest only through its effect on the competing event. The separable direct and indirect effects add up to the total effect, which equals the treatment effect on the cumulative incidence of the event of interest. To identify the separable effects, we assume that the treatment can be decomposed into distinct components that exert effects through distinct causal pathways, similar to Didelez (2). Separable effects are not defined by cross-world contrasts or hypothetical interventions to prevent competing events. As an illustration, we implement our approach in a study of estrogen therapy in patients with prostate cancer, using data from a randomised clinical trial.

A conceptual framework for the population-attributable fraction in the presence of time-dependent exposures and competing risks

Maja von Cube, Martin Schumacher, Jean-Francois Timsit, Martin Wolkewitz on behalf of COMBACTE-MAGNET Consortium

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Stefan-Meier Str. 26, 79104 Freiburg, Germany

Recent studies on antimicrobial resistance imply that a large proportion of hospital deaths could have been prevented if the risk of infection caused by multi-resistant bacteria was eliminated. The statistical quantity that defines the number of attributable death cases is the population-attributable fraction (PAF). The PAF, quantifying the public health impact of a potentially harmful exposure, is of significant importance to patients, clinicians and public health officials. Nevertheless, an estimand accommodating typical data arising in hospital epidemiology is not clearly defined.

By applying different approaches that account for time-dynamic processes, we explain how to define, identify and estimate the PAF for a binary internal time-dependent exposure and data that is subject to competing outcomes. The different approaches result in different estimands of the PAF. By explicitly differentiating between effect measures defined with observable quantities and causal effect measures defined with counterfactual ones, we develop a clear concept of the PAF for complex time-to-event data.

First, we explain the different implicit assumptions of the estimands using directed acyclic graphs. Then, we discuss and compare their interpretation and performance based on a simulation study. Finally, the approaches are applied to a large French database to estimate the health impact of ventilator-associated pneumonia caused by the specific pathogen Pseudomonas aeruginosa.

Our findings cover the often used approach of disregarding the timing of exposure and outcome to estimate the PAF. Moreover, they apply equally to other effect measures such as the relative risk, the odds ratio or the risk difference.
Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions

Michael Schomaker¹, Miguel Angel Luque-Fernandez², Valeriane Leroy³, Mary-Ann Davies¹

¹ University of Cape Town, Centre for Infectious Disease Epidemiology & Research, South Africa
² Biomedical Research Institute of Granada. Noncommunicable and Cancer Epidemiology Group. Andalusian School of Public Health, University of Granada, Granada, Spain
³ INSERM, Unite 1027, Toulouse, France

When estimating treatment effects in the context of longitudinal observational data, time-dependent confounding with treatment-confounder feedback is often a concern. For example, when estimating the effect of different antiretroviral treatment (ART) initiation rules on HIV positive children’s growth, a child’s CD4 count influences both the probability of ART initiation and growth - but is affected by past treatment as well. Both the parametric g-formula and longitudinal targeted maximum likelihood estimation (LTMLE) are promising estimation strategies that can appropriately deal with such situations (if, given the assumptions of a structural causal model and the observed data, the respective target quantity can be identified). However, both estimators have not often been used and compared in the context of complex longitudinal settings with long follow-up time, gradually declining sample size, dynamic interventions rules (i.e. where treatment assignment is based on confounders), limited support for some intervention rules of interest and a very large set of potential adjustment variables, increasing both the need and the challenge of integrating appropriate machine learning methods.

The appropriate application of these estimators is demonstrated in a sophisticated and topical example from HIV treatment research and complemented by a simulation study which targets the performance of LTMLE in a challenging longitudinal setting. A critical assumption for identification, namely positivity, is discussed and diagnostic measures are explored in both the simulation study and the data example.
Difference-in-differences estimation with high-dimensional common trend confounding

Michael Zimmert

University of St. Gallen (SEW)

In this study, we consider difference-in-differences models with a common trend assumption that is only valid after conditioning on covariates. We suggest estimators that allow the covariates to enter the model in a very flexible form. In particular, we propose estimation procedures that involve supervised machine learning methods. We derive asymptotic results for new semiparametric and linear model based estimators for repeated cross-sections and panel data and show that they have desirable statistical properties like asymptotic normality and double robustness. Further, we establish a semiparametric efficiency bound for panel difference-in-differences estimation. The proposed semiparametric estimator attains this bound. The usability of the methods is assessed by replicating a study on an employment protection reform. We demonstrate that the notion of high-dimensional common trend confounding has implications for the economic interpretation of the policy evaluation results. Notably, measured reform effects are substantially decreased or even reversed when covariates are included in a data-driven manner.
Direct and spillover effects using the synthetic control method in the presence of interference

Marco Mariani\textsuperscript{2}, Giulio Grossi\textsuperscript{2}, Patrizia Lattarulo\textsuperscript{2}, Alessandra Mattei\textsuperscript{1}, Özge Öner\textsuperscript{3}

\textsuperscript{1} University of Florence – Dept. of Statistics, Informatics, Applications
\textsuperscript{2} IRPET – Regional Institute for Economic Planning of Tuscany
\textsuperscript{3} University of Cambridge – Dept. of Land Economy

The Synthetic Control Group (SCG) method has gained widespread popularity with both methodological and applied scholars. In recent years, there has been a growing number of methodological contributions extending the SCG method, e.g., to studies involving multiple treated units, and comparing it to alternative standard or new approaches for program evaluation. To our best knowledge, however, the existing literature on SCG methods always relies on the assumption of non-interference. In this paper we propose to generalize the SCG method to studies where interference between the treated and the untreated units cannot be ruled out. Building on recent methodological works on causal inference with interference in the potential outcomes framework, we formally define relevant unit-level direct and spillover effects, investigate the assumptions under which we can identify and estimate them, and show how they can be estimated using the SCG method. We illustrate our approach using an observational study where the focus is on assessing the causal effects of a new tramway line recently built in Florence (Italy) on the commercial vitality of the street where it was built. We are interested in evaluating both the direct effect of the new tramway line as well as the spillover effect originating from the tramway on a number of close-by streets.
Invited talk 2:

**Synthetic difference In differences**

Guido Imbens

*Stanford Graduate School of Business*

We present a new perspective on the Synthetic Control (SC) method as a weighted least squares regression estimator with time fixed effects and unit weights. This perspective suggests a generalization with two way (both unit and time) fixed effects, and both unit and time weights, which can be interpreted as a unit and time weighted version of the standard Difference In Differences (DID) estimator. We find that this new Synthetic Difference In Differences (SDID) estimator has attractive properties compared to the SC and DID estimators. Formally we show that our approach has double robustness properties: the SDID estimator is consistent under a wide variety of weighting schemes given a well-specified fixed effects model, and SDID is consistent with appropriately penalized SC weights when the basic fixed effects model is misspecified and instead the true data generating process involves a more general low-rank structure (e.g., a latent factor model). We also present results that justify standard inference based on weighted DID regression. Further generalizations include unit and time weighted factor models.

(joint work with Dmitry Arkhangelsky, Susan Athey, David Hirshberg and Stefan Wager)
Estimating effects from hypothetical treatment regimes in survival analysis with stochastic differential equations

Kjetil Røysland

University of Oslo

We will present a method for estimating the effects one would see if various hypothetical treatment regimes had been applied to observational survival studies. We show that if we have a causal model for the observational data, then various effect measures, such as cumulative incidence, relative survival etc. in the hypothetical scenario can be estimated by solving certain stochastic differential equations. This enables us to measure the effect of many hypothetical treatment regimes in a same way. Especially, it allows us to implement generic software that can handle different situations without doing a lot of technical model building considerations before carrying out an analysis. By using existing theory for stochastic differential equations, we are able to show that our method is consistent, and that the asymptotic co-variance can be partially identified by solutions of relatively simple ordinary differential equations. These equations translate into recursively defined estimators that can be easily implemented on a computer to calculate confidence intervals etc. Marginal structural modelling in continuous time with stabilized weights turn out to be a special case of this. The methodology is fundamentally non-parametric, as the models are only specified in terms of causal local independence graphs, and two instances of Aalen's additive hazard model. We will present applications to clinical studies.
Policy-relevant interventional effects for evaluating interventions on multiple mediators: Application to adolescent self-harm and risk of later financial hardship

Margarita Moreno-Betancur\textsuperscript{1,2}, Paul Moran\textsuperscript{3}, Denise Becker\textsuperscript{2}, Carolyn Coffey\textsuperscript{2}, George Patton\textsuperscript{1,2}, John B Carlin\textsuperscript{1,2}

\textsuperscript{1}University of Melbourne, Melbourne, Australia
\textsuperscript{2}Murdoch Children's Research Institute, Melbourne, Australia
\textsuperscript{3}University of Bristol, Bristol, U.K.

Causal mediation approaches have been primarily developed for the goal of “explanation”, that is, to understand the pathways that lead from a cause to its effect. A subtly different goal is to evaluate the impact of interventions on mediators, for example in epidemiological studies seeking to inform public health policies for improving the life course of those who are vulnerable or ill in early life. Often, it is infeasible to define interventions that would prevent some early life experiences while it might be more plausible to devise strategies to prevent their consequences. While there has been some work in the literature focused on evaluating mediator interventions, there has been no proposal that explicitly defines the target estimands in terms of the hypothetical randomised trials that one might seek to emulate. In this paper, we propose a novel decomposition of the total causal effect into so-called interventional effects devised explicitly for evaluating mediator interventions in the setting with multiple interdependent mediators. We define these effects in terms of a target randomised trial evaluating a number of realistic mediator interventions in the context of limited resources. The interventions were motivated by a study comparing alternative policies for improving the financial outcomes of adolescent self-harmers. We determined the assumptions required to identify these effects from observational data and developed a \textit{g}-computation estimation method. Application to the self-harm example using data from the Victorian Adolescent Health Cohort Study illustrated how these effects can provide relevant evidence to inform the prioritisation alternative courses of action.
You may think that statistical causal inference is about inferring causation. You may think that it cannot be tackled with standard statistical tools, but requires additional structure, such as counterfactual reasoning, potential responses or graphical representations. I shall try to disabuse you of such woolly misconceptions by locating statistical causality firmly within the scope of traditional statistical decision theory. From this view point, the enterprise of “statistical causality” could fruitfully be rebranded as “assisted decision making”.
Modelling time-to-event outcomes in an observational setting with a regression discontinuity design

Mariam Adeleke, Aidan O’Keeffe, Gianluca Baio

Department of Statistical Science, University College London, United Kingdom.

Regression discontinuity designs (RDD) provide a method to estimate a treatment effect in observational studies where treatment is allocated (partially) according to a decision rule linked to an assignment variable. RDDs have been typically used when the outcome of interest is continuous. We explore methods to estimate treatment effect when the outcome is a time-to-event, under an accelerated failure time (AFT) assumption.

Structural AFT (SAFT) models have been shown to produce an unbiased estimate of a treatment effect (acceleration factor) in observational studies. We propose an estimator of the acceleration factor that is based on the assumptions of the RDD. Simulation studies were carried out to mimic data from real observational studies. Results from simulation studies show that when there are no confounders, both methods produce unbiased estimate of the treatment effect. However, when there is confounding, the SAFT becomes biased. The proposed RDD estimator, on the other hand, provides better coverage of the treatment effect in the presence of confounding. Finally, we applied the methods to a real data on the effect of statin prescription on time to cardio vascular disease in UK primary care.
General discontinuity designs using covariates

Juan D. Díaz¹, José R. Zubizarreta²

¹ University of Chile
² Department of Health Care Policy, and Department of Statistics, Harvard University

Regression discontinuity designs or, simply, discontinuity designs (as we call them in this paper), are widely used for causal inference in observational studies. Motivated by a case study of the impact of grade retention on juvenile crime, we propose a new framework for discontinuity designs. Unlike existing frameworks, this framework can naturally incorporate both multiple running variables and multi-valued treatments. Importantly, the running variables may be discrete. In this framework, the observed covariates play a central role for identification, estimation, and generalization. Identification essentially relies on a local unconfoundedness assumption. Estimation proceeds as in any observational study under the unconfoundedness assumption — yet in a neighborhood of the cut-off’s of the running variables. We discuss estimation approaches based on matching and weighting, including additional regression adjustments in doubly robust estimators. We present and discuss assumptions for generalization; this is, for identification and estimation of average treatment effects for target populations. We present a new approach to select the neighborhood for the analyses and assess the plausibility of the assumptions. Here again, the role played by observed covariates is central. The proposed framework is flexible and facilitates principal stratification and, e.g., analyses for intermediate outcome variables. Our case study is of independent interest. We use a unique administrative data set with extensive educational and criminal records of the same students, observed for 15 years of their lives. We find that grade retention in school has no impact on juvenile crime.
Invited talk 4:

Challenges in applying RDD

Kate Tilling

University of Bristol

The regression discontinuity design is deceptively simple in concept, but more challenging to apply in practice. I will present two recent examples using RDD to address causal questions in health research, and discuss the practical challenges including: selection; missing data; model fitting; sensitivity analyses; hypothesis-free searches for causal effect.
Discovering causal mechanisms via contamination mixture modelling

Stephen Burgess\textsuperscript{1,2}, Christopher Foley\textsuperscript{1}

\textsuperscript{1} MRC Biostatistics Unit, University of Cambridge, UK
\textsuperscript{2} MRC/BHF Cardiovascular Epidemiology Unit, University of Cambridge, UK

Distinguishing between correlation and causation is fundamentally important when trying to understand disease mechanisms using population-based data. Instrumental variable (IV) analysis is an approach to assess whether a risk factor has a causal effect on an outcome based on observational data.

We assume that we have multiple candidate instruments, each of which may or may not be a valid IV. This scenario is particularly common for Mendelian randomization, the use of genetic variants as IVs. We here introduce a method for obtaining valid causal inferences with some invalid IVs that models the variant-specific estimates using a mixture distribution, with one component of the mixture being the distribution of the causal estimate for a valid instrumental variable, and the other component being the distribution for an invalid instrumental variable. Our proposal has a number of advantages over previously proposed approaches: it is asymptotically consistent estimates under the ‘plurality of valid instruments’ assumption, fully likelihood-based, computationally scalable to large numbers of candidate instruments, and implemented using summarized genetic data that are widely available from large consortia. A feature of the proposed method is that it can identify groups of genetic variants with similar causal estimates. If multiple such groups are identified, this suggests that there may be several causal mechanisms associated with the same risk factor that affect the outcome.

We illustrate the use of the method in an applied example considering the causal effect of high-density lipoprotein cholesterol on coronary heart disease risk, demonstrating a bimodal distribution of the variant-specific estimates, and investigate factors that may give rise to this distribution.
Average causal effect estimation via instrumental variables: the no simultaneous heterogeneity assumption

Fernando Pires Hartwig\textsuperscript{1,2}, Jack Bowden\textsuperscript{2,3}, George Davey Smith\textsuperscript{2,3}, Bernardo Lessa Horta\textsuperscript{1}, Neil Martin Davies\textsuperscript{2,3}.

\textsuperscript{1}Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil. \textsuperscript{2}MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom. \textsuperscript{3}Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom.

Instrumental variables can be used to assess whether a treatment $X$ has a causal effect on an outcome, $Y$. An instrument, $Z$ is valid if it satisfies the three core instrumental variable assumptions of relevance, independence and the exclusion restriction. Even if the instrument satisfies these assumptions, further assumptions are required to estimate the average causal effect (ACE) of an exposure on an outcome. Sufficient assumptions for this include: homogeneity in the causal effect of $X$ on $Y$; homogeneity in the association of $Z$ and $X$; or no effect modification (NEM) by the instrument of the causal effect of $X$ on $Y$ within levels of $X$. These assumptions can identify the average causal effect or the average causal effect amongst the treated. However, these assumptions are relatively strong, and often implausible, for example, the effects of a treatment on a binary outcome are unlikely to be homogenous across the population.

We describe the NO Simultaneous Heterogeneity (NOSH) assumption, which requires the heterogeneity in the $Z$-$X$ association and heterogeneity in the $X$-$Y$ causal effect to be independent. We show that, if NOSH holds, conventional instrumental variable methods are consistent for the ACE even if both homogeneity assumptions and NEM are violated. Therefore, NOSH allows identification and interpretation of instrumental variable estimates under weaker assumptions.

We illustrate these ideas using simulations and by demonstrating how they affect interpretation of randomized trials (Sommer et al, 1988), natural experiments (Angrist 1990) and Mendelian randomization (Davey Smith and Ebrahim 2003). In conclusion, we demonstrate how instrumental variable estimates can be interpreted as estimates of the average causal effect under weaker and therefore more plausible assumptions.
Detecting heterogeneous treatment effect with instrumental variables

Michael Johnson, Hyunseung Kang

University of Wisconsin - Madison

There is an increasing interest in methods estimating heterogeneity in causal effects in randomized and observational studies. However, little research has been conducted to understand heterogeneity in an instrumental variables study. In this work, we present a method to estimate heterogeneous causal effects using an instrumental variable approach. The method has two parts. The first part uses subject-matter knowledge and interpretable machine learning techniques, such as classification and regression trees and penalized regression, to partition the data set into potential subgroups with heterogeneous treatment effects. The second part tests for heterogeneous treatment effect within these partitions. To control for the concerns of using one data set to find partitions and make statistical inference, sample-splitting and a recent method of taking the absolute value of the outcome with Bonferroni correction and closed testing are applied to strongly control for familywise error rate. We conducted this method on a real data set example on the effect of malaria on stunted child growth, which showed evidence of heterogeneity in children with and without mosquito nets for protection.
Invited talk 5:

Heterogeneous treatment effects using continuous instrumental variables

Anirban Basu

University of Washington

This talk will overview the conceptual underpinnings related to the identification of marginal treatment effects (MTEs) using local instrumental variables approach. It will then illustrate the use of these estimated MTEs to form Person-centered Treatment (PeT) effects that are conditioned on the person’s observed characteristics and averaged over the potential conditional distribution of unobserved characteristics that lead them to their observed treatment choices. PeT effects are more individualized than conditional treatment effects from a randomized setting with the same observed characteristics. PeT effects can be easily aggregated to construct any of the mean treatment effect parameters and, more importantly, are well-suited to comprehend individual-level treatment effect heterogeneity. Use of PeT effects will be illustrated using an example of ICU transfer in hospitals and its effect on mortality.
Comparing covariate prioritization via matching to machine learning methods for causal inference using five empirical applications

Luke Keele, Dylan S. Small

University of Pennsylvania

When investigators seek to estimate causal effects, they often assume that selection into treatment is based only on observed covariates. Under this identification strategy, analysts must adjust for observed confounders. While basic regression models have long been the dominant method of statistical adjustment, more robust methods based on matching or weighting have become more common. Of late, even more flexible methods based on machine learning methods have been developed for statistical adjustment. These machine learning methods are designed to be black box methods with little input from the researcher. Recent research used a data competition to evaluate various methods of statistical adjustment and found that black box methods outperformed all other methods of statistical adjustment. Matching methods with covariate prioritization are designed for direct input from substantive investigators in direct contrast to black methods. In this article, we use a different research design to compare matching with covariate prioritization to black box methods. We use black box methods to replicate results from five studies where matching with covariate prioritization was used to customize the statistical adjustment in direct response to substantive expertise. For the analysis, we selected a diverse set of applications with different underlying properties. We find little difference across the methods. We conclude with advice for investigators.
Interventional effects models for mediation: analysis with multiple mediators

Wen Wei Loh¹, Beatrijs Moerkerke¹, Tom Loeys¹, Stijn Vansteelandt²,³

¹ Department of Data Analysis, Ghent University, Gent, Belgium
² Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium
³ Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

In settings that involve repeatedly measured mediators or multiple causally-connected mediators, approaches focusing on fine-grained decompositions of natural (in)direct effects are only valid under strong assumptions. In particular, the assumptions are known to be violated when – as often – the structural dependence between the multiple mediators is unknown. In contrast, interventional (in)direct effects, introduced by VanderWeele, Vansteelandt, and Robins (2014), can be identified under much weaker conditions than natural (in)direct effects, but have the drawback of not adding up to the total effect. Vansteelandt and Daniels (2017) adapted their proposal to achieve an exact decomposition of the total effect, and generalized the interventional effects to the multiple mediator setting. In this article, we introduce interventional effects models that allow for simultaneous and parsimonious modeling of the interventional effects when there are multiple mediators. The parameters in the interventional effects models encode the path-specific effects of an exposure on an outcome that are mediated by distinct mediators, even when the direction of the causal effects between the mediators is unknown, or the mediators are manifestations of an underlying latent process, or there may be unmeasured common causes of the mediators. The mediators and outcome can be continuous or non-continuous. Estimation proceeds via Monte Carlo integration and only requires a working joint distribution of the mediators and a working outcome model.
A weighting method for simultaneous adjustment for confounding and joint exposure-outcome misclassifications

Bas B.L. Penning de Vries, Maarten van Smeden, Rolf H.H. Groenwold

Leiden University Medical Center

Joint misclassification of exposure and outcome variables can lead to considerable bias in epidemiological studies of causal exposure-outcome effects. We present a new maximum likelihood based estimator for the marginal causal odds-ratio that simultaneously adjusts for confounding and several forms of joint misclassification of the exposure and outcome variables. The proposed method relies on validation data for the construction of weights that account for both sources of bias. The weighting estimator, which is an extension of the exposure misclassification weighting estimator proposed by Gravel and Platt (Statistics in Medicine, 2018), is applied to reinfarction data. Simulation studies were carried out to study its finite sample properties and compare it with methods that do not account for confounding or misclassification. The new estimator showed favourable large sample properties in the simulations. Further research is needed to study the sensitivity of the proposed method and that of alternatives to violations of their assumptions. The implementation of the estimator is facilitated by a new R function in an existing R package.
Graphical criteria for efficient total effect estimation via adjustment

Leonard Henckel\textsuperscript{1}, Emilija Perkovic\textsuperscript{2}, Marloes Maathuis\textsuperscript{1}

\textsuperscript{1} ETH Zurich
\textsuperscript{2} University of Washington

When the underlying causal graph is known, there exists a sound and complete graphical criterion for when a covariate set allows for consistent total effect estimation. However, typically a large number of sets fulfills this criterion. Restricting ourselves to the causal linear structural equation model setting with arbitrary errors, we introduce graphical criteria that allow one to identify in many cases which of two valid adjustment sets provides the smaller asymptotic variance. Even though this result only induces a partial ordering, it can be used to identify a valid adjustment set that always provides the optimal asymptotic variance, while being easily computable. We further introduce a pruning procedure that given any valid adjustment set outputs a subset that is still valid and guaranteed to provide a smaller or equal asymptotic variance.
On efficient adjustment in causal graphs

Janine Witte, Vanessa Didelez

Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

In typical analyses of epidemiological or biomedical data we wish to estimate the causal effect of an exposure or treatment $X$ on an outcome $Y$ while adjusting for a set of covariates/confounders. There are different valid ways of choosing the adjustment set, depending on the amount of back-ground knowledge. Ideally, but unrealistically, we can fully specify the underlying causal directed acyclic graph (DAG). For this case, Henckel et al. 2019 provide a graphical criterion for reading off the ‘optimal’ adjustment set in the sense that, given a linear model, adjusting for this set results in the most efficient OLS estimator. In this talk, we propose an alternative, equivalent definition of the optimal adjustment set for DAGs that provides additional insight into the intuition underlying its optimality. Our definition is based on the latent projection (Shpitser et al. 2014) of a DAG over the so-called forbidden set of variables. In the more common setting, without sufficient background knowledge, one may fall back on data-driven covariate selection algorithms, such as stepwise selection of regression models. Here, we investigate the assumptions under which popular algorithms select the optimal adjustment set. As a third approach, the IDA algorithm (Maathuis et al. 2009) combines causal search and graphical adjustment criteria in order to estimate possible causal effects under uncertainty about the causal structure. IDA has, for instance, been successfully applied to genetics data (Stekhoven et al. 2012). Here we show that optimal adjustment can easily be integrated into the IDA algorithm to further enhance its performance.

References:


Unmeasured confounding, bias amplification and model selection

Ian Shrier\(^1\), Tyrel Stokes\(^2\), Russell J Steele\(^2\)

\(^1\) Centre for Clinical Epidemiology, Lady Davis Institute, McGill University
\(^2\) Department of Mathematics and Statistics, McGill University

In ignorability approaches to identifying Average Causal Effects, we seek to condition on a set of variables such that the potential outcomes are conditionally independent of the treatment. In non-experimental settings, however, the entire set of variables required to satisfy the ignorability condition is often not available and confounding paths remain. Recent theoretical work by Wooldridge, Pearl and Ding has explored an important class of variables which can further amplify existing unmeasured confounding bias when conditioned on. This is contrary to the popular wisdom that conditioning on additional observables reduces unmeasured confounding bias. We argue for adopting a matrix notation framework to more easily generalize bias amplification. Specifically, using the matrix notation framework we take four concrete steps towards building a more comprehensive accounting of bias amplification. First, we extend existing results to include non-standardized variables and offer an interpretation of bias amplification in terms of the geometry of ordinary least squares and variance of the treatment. Second, we generalize to a larger class of causal structures, including two new causal Directed Acyclic Graphs to demonstrate that amplification fundamentally concerns the relative strength of confounding pathways. Third, we build upon the work by Middleton to show that the amplification term is identifiable under broad assumptions and causal structures in the probability limit. Finally, we propose a counter-factual based simulation method which allows us to define the parameter space and visualize the amplification bias in a comprehensive manner.
Bayesian model averaging for two-sample summary data mendelian randomisation in the presence of pleiotropy

Chin Yang Shapland\textsuperscript{1,2}, Jack Bowden\textsuperscript{1,2}

\textsuperscript{1} MRC Integrative Epidemiology Unit at the University of Bristol, U.K.
\textsuperscript{2} Population Health Sciences, University of Bristol, U.K.

Two-sample summary data Mendelian randomisation (MR) is a popular method for assessing causality in epidemiology, by using multiple genetic variants as instrumental variables (IVs). If genetic variants exert pleiotropic effects on the outcome not through the exposure of interest, this leads to heterogeneous (and potentially biased) estimates of causal effect. It is possible to detect and remove outlying variants \cite{1} but this can lead to an under-estimation of the standard error and conflate weak instrument bias with pleiotropy \cite{2}. Alternatively, mode-based strategies \cite{3,4} focus on the largest subset of variants yielding a homogeneous causal estimate. However, searching all \(2^L\) subsets of \(L\) genetic variants is not computationally feasible when \(L\) is large.

Rather than detecting and removing outlying estimates or attempting to search all possible subsets, we investigate the use of Bayesian model averaging (BMA) to preferentially search the space of models with the highest posterior likelihood. We develop a bespoke Metropolis-Hasting algorithm to perform the search and use the profile likelihood of Zhao et al \cite{5} to define a posterior distribution that accounts for pleiotropic and weak instrument bias. In keeping with the Bayesian framework, our method also allows prior knowledge to influence on the validity of each variant to be seamlessly included.

We use Monte Carlo simulations and real data examples to illustrate our approach and compare it to several related approaches, highlighting its relative strengths and weaknesses in outlier detection and causal estimation.

4. Burgess, S., et al., Modal-based estimation via heterogeneity-penalized weighting: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. IJE 2018
Application of the instrumental inequalities to a Mendelian randomization study with multiple proposed instruments

Elizabeth W. Diemer¹, Jeremy Labrecque², Henning Tiemeier¹,³, Sonja A. Swanson²,⁴

¹Department of Child Psychiatry, Erasmus MC, Rotterdam, the Netherlands
²Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
³Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, USA
⁴Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA

Mendelian randomization (MR), an application of the instrumental variable model proposing genetic variants as instruments, is increasingly popular. Investigators often attempt to support the validity of MR via subject matter knowledge. However, the instrumental variable model implies certain inequalities, offering an empirical method of falsifying (but not verifying) the underlying assumptions. While these inequalities are said to detect only extreme assumption violations in practice, they have not been used in settings with multiple proposed instruments.

Methods: We applied the instrumental inequalities to an MR analysis of the relationship between maternal Vitamin D in pregnancy and offspring attention deficit hyperactivity disorder and persistent developmental problems, proposing four maternal genetic variants as instruments. We assessed whether the four proposed instruments jointly satisfied the instrumental inequalities, and whether the inequalities were satisfied for each variant separately or any combination of two or three of the variants.

Results: The instrumental inequalities were satisfied (i.e., we did not falsify the MR model) when considering each variant separately as a possible instrument. However, the inequalities were violated when considering the four variants jointly, as well as for some combinations of two or three variants.

Conclusions: We present an example where the instrumental inequalities detected violations of the MR assumptions for genetic variants jointly proposed as instruments, though the instrumental inequalities were satisfied when considering each proposed instrument separately. This underscores previous calls for the use of the instrumental inequalities, and suggests that assessing these inequalities can eliminate clearly invalid analyses in settings with many proposed instruments.
Multivariable Mendelian randomisation (MVMR) is a form of instrumental variable estimation which utilises genetic variants to estimates the direct causal effects of multiple exposures on an outcome, as an extension to a standard univariate analysis. A key assumption required for consistent MVMR estimation is that the genetic variants used as instruments can strongly predict all of the exposures. This means that the genetic variants must predict all exposures in the model, conditional on each other. In analysis using individual level data this assumption can be tested using the Sanderson-Windmeijer conditional F-statistic.¹ In this paper we derive a summary F-statistic for weak instruments in the two-sample summary data MVMR setting, which uses SNP-trait associations from publicly available GWA studies. We show that it has the same distribution as its analogous individual level conditional F-statistic, but that it depends on summary information that is generally not reported or available. We use simulations to explore the sensitivity of our summary data F-statistic to this missing information, and explore strategies for estimation when this data is not available. Finally, we apply this test to an MVMR analysis of the effects of education and cognitive ability on BMI in UK Biobank.

An interval estimation approach to selection bias in IV studies

Matthew Tudball, Jack Bowden, Rachael Hughes, Kate Tilling

MRC Integrative Epidemiology Unit, University of Bristol

Instrumental variables (IV) estimation is frequently used in applied studies to recover causal effects of interest. Sensitivity to selection bias (e.g. by non-random participation or drop-out), however, is seldom examined in practice. A common approach for handling selection bias, inverse probability weighting, relies on correct specification of the individual weights, which may not be possible in practice. To address this problem, Aronow and Lee (2013) proposes an interval estimator for population means in settings where these probability weights are unobserved but are known to be bounded.

We extend that estimator to derive bounds for IV estimates. To accomplish this, we develop a computational procedure for calculating the interval bounds and prove asymptotic sharpness under relatively weak assumptions. We also develop methods to incorporate a variety of information related to sample selection in order to tighten the interval estimate. Specifically, we show how to incorporate three types of information that are commonly available to researchers. 1) the study response rate, 2) variables that are only observed in-sample, 3) variables that are observed among all individuals who were invited to enter the study. We evaluate the performance of this method in practice using simulations and an applied example of estimating the effect of education on BMI in UK Biobank.

Given that commonly-used datasets such as UK Biobank are known to suffer from non-random sample selection, our method provides a flexible way for researchers to check the sensitivity of their conclusions to plausible sample selection mechanisms.
The confidence interval method for selecting valid instruments for instrumental variables estimation

Frank Windmeijer, Xiaoran Liang, Fernando Hartwig, Jack Bowden

University of Bristol

We consider a setting with many potential instruments for estimating the causal effect of an exposure on an outcome, but were some of the instruments may be invalid in the sense that they don't satisfy the exclusion restriction. An MR application we consider is the effect of BMI on blood pressure, with 96 available SNPs as instruments, where some of these could be invalid due to e.g. pleiotropy. Due to the large number of instruments, we need to use dimension reduction techniques. Use of the Lasso has been proposed before, and a hard thresholding method with voting has been shown by Guo et al. (2018) to have oracle properties. Here we show that a simple and fast method that selects as the set of valid instruments the largest number with overlapping confidence intervals of the per instrument causal effect estimates has oracle properties. Combining it with a downward testing procedure based on the test of overidentifying restrictions results in an estimator with good properties, confirmed by Monte Carlos simulations.
Marginal structural models in the presence of multiple treatments, with application to the analysis of export promotion programs

Chiara Bocci¹, Marco Mariani², Alessandra Mattei¹, Fabrizia Mealli¹

¹Dept. of Statistics, Computer Science, Applications “G. Parenti” - University of Florence
²IRPET – Regional Institute for Economic Planning of Tuscany

Marginal structural models (MSMs) enjoy wide popularity with methodological and applied scholars. Under appropriate sequential ignorability assumptions that address the issue of dynamic confounding, MSMs allow to analyze longitudinal, observational studies characterized by repeated treatment over time and to recover the causal effects of the treatment sequences of interest. Surprisingly, the approach based on MSM has not yet been generalized to settings with longitudinal, multivalued treatments, e.g. with multiple discrete or even continuous treatments. Our paper aims to start to fill the gap. We generalize the assumptions needed to identify causal effects in longitudinal settings characterized by the possibility of the repeated intake of multiple discrete treatments. We adopt a MSM approach and we show how the effects of interest can be estimated using inverse-probability-of-treatment weighting. We apply our approach to the analysis of a real Italian case study of export promotion policy. The program consists of the provision of multiple services and aids by specialized agencies, including consultancy, support to the participation in international fairs and business-to-business meetings, of which firms can take advantage either simultaneously or at different moments in time.
Bayesian inference for a principal stratum estimand – Methods and sensitivity analyses

Björn Bornkamp, Baldur Magnusson

Novartis Pharma AG

The treatment effect in a specific subgroup is often of interest in randomized clinical trials. When the subgroup is characterized by the absence of certain post-randomization events, a naive analysis on the subset of patients without these events may be misleading. The principal stratification framework allows one to define an appropriate causal estimand in such settings. Statistical inference for the principal stratum estimand hinges on scientifically justified assumptions, which can be included with Bayesian methods through prior distributions. Our motivating example is the setting of drug development in multiple sclerosis, discussed in detail in Magnusson et al. (2018), where the outcome as well as the intermediate variable are binary.

In this presentation we will outline strategies for sensitivity analyses with respect to the chosen prior distributions, by (i) modifying the utilized prior distributions within plausible ranges as well as (ii) determination of the identification region (without utilizing prior distributions) based on a generic computational approach.

Efficient, doubly robust estimation of the effect of dose switching for switchers in a randomised clinical trial

Kelly Van Lancker¹, An Vandebosch², Stijn Vansteelandt¹,³

¹Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium
²Janssen R&D, a division of Janssen Pharmaceutica NV, Beerse, Belgium
³Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

This work is motivated by a clinical trial conducted by Janssen Pharmaceuticals in which a flexible dosing regimen is compared to placebo. In order to understand whether flexible dosing is potentially beneficial for switchers in the treatment arm (i.e., patients who were switched to the higher dose), it is of interest to evaluate how they would have fared had they been kept on the low dose. Comparing these patients’ responses with those of patients who stayed on the low dose does not likely entail a satisfactory evaluation because the latter patients are usually in a better health condition and the available information is too scarce to enable a reliable adjustment. In view of this, we will transport data from a fixed dosing trial that has been conducted concurrently on the same target, albeit not in an identical patient population.

In particular, we will propose a doubly robust estimator, which relies on an outcome model and a propensity score model for the association between study and patient characteristics. The proposed estimator is asymptotically unbiased if either model is correctly specified and achieves the semi-parametric efficiency bound (under the model defined by the restrictions on the propensity score) when both models are correctly specified. Theoretical properties are also evaluated through Monte Carlo simulations and the method will be illustrated based on the motivating example.
Addressing treatment contamination in the design and analysis of trials of complex interventions

Sabine Landau¹, Nicholas Magill¹², Paul McCrone¹, Khalida Ismail¹

¹ Institute of Psychiatry, Psychology and Neuroscience, King’s College London, UK
² London School of Hygiene and Tropical Medicine, UK

Participants from the control arm receiving the active intervention is a concern in trials of complex interventions. Cluster randomisation with clusters defined by the level at which such contamination is thought to operate, is the most commonly proposed solution [1]. However, where measured the amount of contamination was found to be low (median 13% participants) and alternative estimators exist that can estimate treatment efficacy in the presence of non-adherence with randomised treatments.

We carried out a simulation study to evaluate the performance of two competing design and analysis options: (A) cluster randomisation combined with an estimator of the Average Treatment Effect that accounts for clustering versus (B) individual level randomisation and measurement of treatment contamination combined with an instrumental variables estimator of the Complier Average Causal Effect. We simulated the contamination process of most concern [1]: therapists being trained in both conditions delivering the active treatment in the control arm. The study confirmed that both options provide consistent estimators of respective causal estimands, and showed that their relative efficiency was driven by the strength of the clustering and the size of the contaminator stratum. For example, for moderate/large (10 or 20) cluster sizes, moderate/large (0.05 or 0.1) intra-cluster correlation coefficients and contamination ≤30% design option B was more efficient.

The simulated relative efficiency ratios were implemented as an online decision support tool, see https://nicholasmagill.shinyapps.io/shiny_app/. We will demonstrate how to use this tool to identify the better trial design option under contamination, with a mental health trial as an exemplar [2].

References:


Sibling comparison designs: A viable path to causal effects?

Anne Helby Petersen¹, Theis Lange¹²

¹ Section of Biostatistics, Department of Public Health, University of Copenhagen
² Center for Statistical Science, Peking University, Beijing, China

Sibling comparison designs are becoming an increasingly popular tool for estimating causal effects in settings where randomized trials are not feasible. By utilizing the natural matching of siblings into family units that may share certain types of confounding effects - for example parental genomes or socio-economic conditions in the childhood - it is possible to estimate confounding-free effects by the use of within-family contrasts. However, such confounding-free effects are not necessarily causal effects, and the distinction between these two concepts has generally received little attention in the sibling comparison literature. A first step towards understanding the causal nature (or lack thereof) of the effects estimated by sibling models is to establish a formal mathematical language in which the sibling models can be described. In this presentation, we propose a counterfactual-based framework for defining and discussing sibling comparison designs that allows us to describe the parameters of these models in terms of hypothetical interventions. More specifically, within this framework, we can address questions like the following: Does there exist a well-defined intervention, such that one could in principle design a randomized trial, perform this intervention and thereby target the same effects as the typical sibling models do? And if so, is this intervention performed at the individual level, or do we need to intervene on the whole sibling group jointly in order to obtain the same effects? We link these theoretical results to typical within-family effect estimation approaches by deriving the asymptotic limits of popular choices of sibling design estimation procedures.
Network deconvolution (ND) is a simple method that aims to distinguish direct from indirect associations. While ND can be applied to any association measure, it recently became a popular approach to process multiple causal effect estimates obtained from Mendelian randomization (MR) studies. Many other methods have been proposed to distinguish direct and indirect effects, ranging from precision matrix estimation to structure learning algorithms for graphical models. The key difference between these methods and ND is that ND assumes that the association measure to be transitive. This transitivity assumption does not hold for most statistical association measures such as correlations, but it does hold for total causal effect estimates.

Despite this apparent fundamental difference, we show that ND is strongly related to precision matrix estimation. Specifically, ND consists of a (1) normalization stage and a (2) deconvolution stage, where stage (2) is mathematically equivalent to computing the precision matrix. This equivalence makes it clear that stage (1) can be seen as an alternative regularization method for precision matrix estimation. We find empirically that ND outperforms precision matrix estimation methods such as graphical lasso on dense networks, whereas it performs equally or worse on sparse networks.

Our results imply that contrary to expectations and current practice, there is in fact no good reason to prefer ND over more established methods such as the graphical lasso, except where the input network is likely to be dense.
Causal learning for linear SDEs

Søren Wengel Mogensen, Niels Richard Hansen

University of Copenhagen

Causal learning is an active field of research and a variety of methods are available, though most methods assume that data is sampled either from time series models or from models with no explicit notion of time. We consider the case of multivariate continuous-time stochastic processes, in particular solutions of linear stochastic differential equations. This class of processes has been used for modeling e.g. in psychology, neuroscience, finance, and survival analysis, and our work is aimed at extending causal learning theory to this class of models.

Several authors have studied the notion of local independence, and we use this concept to describe how a system of continuous-time stochastic processes evolves over time. Such systems can only be sampled in discrete time which complicates inference. We implement local independence tests of discretely sampled data by computing pseudolikelihoods using a simulation-based approach.

We present novel results on faithfulness of local independence graphs in the case of stationary Ornstein-Uhlenbeck processes, and we use this to argue for the validity of our approach.

By testing local independences, we implement an algorithm to learn an equivalence class of graphs that represents causal explanations of the observed data, also when the system of stochastic processes is only partially observed. The algorithm can handle an arbitrary and unknown number of unobserved processes. The algorithm is in some ways analogous to the classical Fast Causal Inference (FCI) algorithm; however, fundamental differences also exist.

We apply this new methodology to simulated and real data.
Bayesian inference of dag models for the estimation of causal effects

Federico Castelletti, Guido Consonni

Università Cattolica del Sacro Cuore (Milano)

Directed Acyclic Graphs (DAGs) are widely used in many scientific areas to model and investigate dependencies among a set of random variables. In addition, they provide an effective framework for causal reasoning too.

It is well known that we cannot distinguish between DAGs encoding the same conditional independencies. These can be collected in equivalence classes each represented by a partially directed graph called Essential Graph (EG). However, an EG does not in general provide information on the causal relationship between variables. Consequently, to estimate causal effects we require additional assumptions based on intervention calculus.

We propose a Bayesian methodology which combines DAG model selection and estimation of causal effects. Coherently, our method provides information on two main sources of uncertainty. The first is related to the DAG generating model and represented by the posterior distribution over the EG space, while the second corresponds to the magnitude of the causal effect between covariates and response and is encoded by the posterior distribution of selected DAG model parameters. We apply our methodology on simulation scenarios and real data.
Invited talk 6:

Robust causal structure learning with some hidden variables

Marloes Maathuis

ETH Zurich

We introduce a new method to estimate the Markov equivalence class of a directed acyclic graph (DAG) in the presence of hidden variables, in settings where the underlying DAG among the observed variables is sparse, and there are a few hidden variables that have a direct effect on many of the observed ones. Building on the so-called low rank plus sparse framework, we suggest a two-stage approach which first removes the effect of the hidden variables, and then estimates the Markov equivalence class of the underlying DAG under the assumption that there are no remaining hidden variables. This approach is consistent in certain high-dimensional regimes and performs favorably when compared to the state of the art, both in terms of graphical structure recovery and total causal effect estimation.

(joint work with Benjamin Frot and Preetam Nandy)
Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes

Helene Charlotte Rytgaard

University of Copenhagen

The aim of this talk is to discuss aspects of a continuous-time generalization of the targeted minimum loss-based estimation (TMLE) framework to estimation of time-varying interventional effects in settings where both interventions, covariates and outcome can happen on any arbitrarily fine time scale and particularly at subjects-specific time-points. TMLE provides a general template for constructing regular and asymptotically linear substitution estimators for smooth low-dimensional parameters in infinite-dimensional models, combining flexible ensemble learning and semiparametric efficiency theory in a two-step procedure. Previous longitudinal TMLE methods are developed for data where observations are all made at a few time-points that are the same for all subjects. We consider a continuous-time model based on counting processes and focus on estimation of the intervention-specific mean outcome at the end of follow-up in a nonparametric model. To construct a TMLE algorithm we derive a closed-form expression for the efficient influence curve. Our proposed targeting algorithm is based on fluctuation models that smoothes information across time, iterating over sequential updating steps until convergence. The resulting estimator solves the efficient influence curve equation, providing the basis for establishing double robustness and statistical inference of the estimator.

Our methods have applications in pharmacoepidemiology where large scale registries provide subject-specific information (electronic records), for example, on dates of drug purchases and hospital admissions, and inference is concerned with a health outcome such as the risk of death.
Integrating Experimental and Observational Data through Machine Learning

Johann Gagnon-Bartsch¹, Adam Sales², Edward Wu¹, Anthony Botelho³, Luke Miratrix⁴, Neil Heffernan³

¹University of Michigan, Ann Arbor
²University of Texas, Austin
³Worcester Polytechnic Institute
⁴Harvard University

Two main virtues of randomized experiments are that they (1) do not suffer from confounding and (2) allow for design-based inference, meaning that the physical act of randomization largely justifies the statistical assumptions made. However, sample sizes are often small. Conversely, observational studies typically offer much larger sample sizes at lower costs, but may suffer confounding. Recent work has sought to integrate "big observational data" with "small but high-quality experimental data" to get the best of both worlds. For example, how can one exploit a large database of electronic health records to improve the accuracy of a small clinical trial? Or, how can one use administrative data on hundreds of thousand of students to improve a small experiment testing the effectiveness of a new educational technique? Similar questions arise across many disciplines.

In this talk I will discuss a flexible framework that allows researchers to employ machine learning algorithms to learn from the observational data, and use the resulting models to improve precision in randomized experiments. Importantly, there is no requirement that the machine learning models are "correct" in any sense, and the final experimental results are guaranteed to be exactly unbiased. Thus, there is no danger of confounding biases in the observational data leaking over into the experiment.

The framework is applied to A/B tests of educational software, using an observational administrative database of previous student achievement. Large gains in precision are obtained, in some cases more than tripling the effective sample size.
Machine learning methods for causal inference from complex observational data

Alexander Volfovsky\textsuperscript{1}, Cynthia Rudin\textsuperscript{2}, Sudeepa Roy\textsuperscript{2}, Tianyu Wang\textsuperscript{2}, Awa Dieng\textsuperscript{2}, Yamen Liu\textsuperscript{2}, Harsh Parikh\textsuperscript{2}, Pritam Dey\textsuperscript{2}

\textsuperscript{1}Department of Statistical Science, Duke University
\textsuperscript{2}Duke University

A classical problem in causal inference is that of matching treatment units to control units in an observational dataset. This problem is distinct from simple estimation of treatment effects as it provides additional practical interpretability of the underlying causal mechanisms that is not available without matching. Some of the main challenges in developing matching methods arise from the tension among (i) inclusion of as many relevant covariates as possible in defining the matched groups, (ii) having matched groups with enough treated and control units for a valid estimate of average treatment effect in each group, (iii) computing the matched groups efficiently for large datasets, and (iv) dealing with complicating factors such as non-independence among units. Many matching methods require expert input into the choice of distance metric that guides which covariates to match on and how to match on them. This task becomes impractical for modern electronic health record and large online social network data simply because humans are not naturally adept at constructing high dimensional functions manually. We propose the Fast Large-scale Almost Matching Exactly (FLAME) framework to tackle these problems for categorical covariates. At its core this framework proposes an optimization objective for match quality that captures covariates that are integral for making causal statements while encouraging as many matches as possible. We demonstrate that this framework is able to construct good matched groups on relevant covariates and leverage these high quality matches to estimate conditional average treatment effects (CATEs) in the study of the effects of a mother's smoking status on pregnancy outcomes. We further extend the methodology to incorporate continuous and other complex covariates.
Invited talk 7:

A unifying approach for doubly-robust L1 regularized estimation of causal contrasts

Andrea Rotnitzky

Di Tella University, Buenos Aires, and Harvard School of Public Health

Recently, there has been a spate of papers on methods for the estimation of the average treatment effect (ATE) under a high-dimensional vector of potential confounding factors, of length often greater than the sample size. All of the papers have assumed that the outcome regression and the propensity score functions were exactly or approximately sparse and therefore proposed to estimate them using L1 regularized regression. The papers differ in the estimators proposed, the assumptions made about the data generating process, and the theorems proved or conjectures made about the statistical behavior of their estimators under their assumptions. These papers prove or conjecture that their estimators are doubly robust; however, in the high dimensional setting, there are two different natural definitions of double robustness: model double robustness and rate double robustness, both of which will be defined in this talk. Each paper concentrates on one definition or the other. It has also been recently shown in the econometrics literature that ATE is an instance of a much larger class of functionals with the property that rate doubly robust estimators can be obtained by the estimation of two nuisance functions of covariates, even when the number of covariates exceeds the sample size. In this talk we show that there exists an even larger class of functionals, which contains interesting causal parameters not covered in the aforementioned class, for which this property holds. For functionals in this larger class we propose a unifying methodology that yields an estimator that improves upon previous estimators in being simultaneously doubly robust in both senses, even while imposing assumptions that are weaker than those imposed by any of the other papers.
POSTERS
Mendelian randomization provides evidence for a causal role of dehydroepiandrosterone sulfate in decreased NT-proBNP levels in a Caucasian population

Lyda Z. Rojas\textsuperscript{1,2,3*}; Oscar L Rueda-Ochoa\textsuperscript{1,4}; Eralda Asllanaj\textsuperscript{4*}; Carolina Ochoa Rosales\textsuperscript{4}; Felix Day\textsuperscript{5}; Eliana Portilla Fernandez\textsuperscript{3}; Katerina Trajanoska\textsuperscript{1}; Jana Nano\textsuperscript{1}; Arfan Ikram\textsuperscript{1}; Oscar H. Franco\textsuperscript{1,6}; Marija Glisic\textsuperscript{1,7*}; Taulant Muka\textsuperscript{1,8*}

\textsuperscript{1}Department of Epidemiology, Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands
\textsuperscript{2}Department of Paediatrics, Obstetrics, Gynaecology and Preventative, Universidad Autónoma de Barcelona, Barcelona, España
\textsuperscript{3}Research Center Fundación Cardiovascular de Colombia, Floridablanca, Colombia
\textsuperscript{4}Electrocardiography Research Group, Medicine School, Universidad Industrial de Santander, Bucaramanga, Colombia
\textsuperscript{5}MRC Biostatistic Unit and Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, UK
\textsuperscript{6}Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
\textsuperscript{7}Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany.
\textsuperscript{*}Indicates equal contribution

BACKGROUND: Observational evidence indicates an inverse association between the levels of the most abundant hormones in the human body, dehydroepiandrosterone and its sulfate ester – DHEA and DHEAs and N-terminal pro B-type natriuretic peptide (NT-ProBNP). We aimed to generate estimates of the associations of DHEA and DHEAs (exposures) with NT-proBNP that were free from confounding and reverse causation, and thus to assess the causal role of these endogenous sex hormones.

METHODS: Serum DHEA, DHEAs and NT-proBNP were assessed in 7,390 men and women free of cardiovascular diseases from the prospective population-based Rotterdam study. DHEA, DHEAS and NT-proBNP were naturally log transformed. Regression coefficients and 95% confidence intervals (CI) were calculated from multivariable linear regression models adjusting for confounders to explore the cross-sectional association of DHEA and DHEAs with NT-proBNP. To investigate the causal association between DHEAs and NT-proBNP, we applied the two-stage least squares (2SLS) method using genetic risk score associated with DHEAs (DHEAs GRS) as an instrumental variable (IV). DHEAs GRS was calculated using nine SNPs previously reported from genome wide association studies to have an association with DHEAs. No genome-wide associations have been described for DHEA in literature, and therefore we could not run the MR analysis for this hormone.

RESULTS: In models adjusted for multiple confounders (age, sex, lifestyle and cardiovascular risk factors), high levels of DHEA (β=-0.146, 95%CI: -0.190; -0.101, p<0.001) or DHEAs (β=-0.214, 95%CI: -0.262; -0.166, p<0.001) were associated with lower levels of NT-proBNP. Genetic risk score of DHEAs explained 29.39% variance of the circulating levels of NT-proBNP. The Mendelian Randomization analysis showed an evidence for a causal association between DHEAs and NT-proBNP, with a causal coefficient of -0.450 (95% CI: -0.792; -0.107, p=0.010). When stratified by sex, although, the directions of associations were in line with overall findings, results were statistically significant only in women, which may be due to low power, as we confirmed in the power calculation analysis.

CONCLUSIONS: The causal association between DHEAs and NT-proBNP observed in this study suggests a new metabolic pathway linking DHEAs with NT-proBNP. Our results should stimulate future research to evaluate the potential role of DHEAs in prevention and management of chronic heart failure.
Selecting causal risk factors from high-throughput experiments using multivariable Mendelian randomization

Verena Zuber\textsuperscript{1,2}, Johanna Maria Colijn\textsuperscript{3,4}, Caroline Klaver\textsuperscript{3,4,5}, Stephen Burgess\textsuperscript{2,6}

\textsuperscript{1} Department of Epidemiology and Biostatistics, Imperial College London, UK
\textsuperscript{2} MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, UK
\textsuperscript{3} Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
\textsuperscript{4} Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, The Netherlands
\textsuperscript{5} Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands
\textsuperscript{6} MRC/BHF Cardiovascular Epidemiology Unit, School of Clinical Medicine, University of Cambridge, UK

Modern high-throughput experiments provide a rich resource to investigate causal determinants of disease risk. Mendelian randomization (MR) is the use of genetic variants as instrumental variables to infer the causal effect of a specific risk factor on an outcome. Multivariable MR is an extension of the standard MR framework to consider multiple potential risk factors in a single model. However, current implementations of multivariable MR use standard linear regression and hence perform poorly with many risk factors.

Here, we propose a novel approach to multivariable MR based on Bayesian model averaging (MR-BMA) that scales to high-throughput experiments and can select biomarker as causal risk factors for disease. In a realistic simulation study we show that MR-BMA can detect true causal risk factors even when the candidate risk factors are highly correlated. We illustrate MR-BMA by analysing publicly-available summarized data on metabolites to prioritise likely causal biomarkers for age-related macular degeneration.
Is the effect of Mediterranean diet on hip fracture mediated through type 2 diabetes mellitus and body mass index?

Adam Mitchell 1, Tove Fall 2, Håkan Melhus 3, Alicja Wolk 1,4, Karl Michaëlsson 1, Liisa Byberg 1

1. Department of Surgical Sciences, Orthopaedics, Uppsala University, Sweden
2. Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Sweden
3. Department of Medical Sciences, Clinical Pharmacogenomics and Osteoporosis, Uppsala University, Sweden
4. Institute of Environmental Medicine, Division of Nutritional Epidemiology, Karolinska Institutet, Sweden

Background— High adherence to a Mediterranean diet is associated with lower risk of hip fracture. The causes are unclear. We sought to investigate whether the effect is mediated by effects of Mediterranean diet on type 2 diabetes mellitus (T2DM) or body mass index (BMI). Conventional regression approaches may be biased, therefore novel statistical methods were applied.

Methods— In 50,755 participants from two population-based cohorts (the Cohort of Swedish Men (COSM) and the Swedish Mammography Cohort (SMC)), we calculated a modified Mediterranean diet score (MDS) categorized into 3 a priori groups that indicated the relative adherence to a traditional Mediterranean diet. T2DM was defined as incident diabetes using a self-reported diabetes diagnoses. First incident hip fracture occurring after 1st January 2009 and prior to December 31, 2014, was considered as outcome. Using weighting and marginal structural models (MSM) we estimated the controlled direct effect (CDE) of Mediterranean diet on hip fracture risk. Using flexible multiple mediator analysis we calculated the natural direct (NDE), natural indirect (NIE) and partial indirect effects (PIE).

Results— Compared to low adherence, the OR (95% confidence interval) for median and high adherence to MDS on hip fracture were 0.82 [0.71, 0.95] and 0.75 [0.62, 0.91]. The corresponding estimates for the CDE were 0.84 [0.71, 0.95] and 0.63 [0.52, 0.78], and for the NDE 0.82 [0.71, 0.94] and 0.86 [0.79, 0.94]. NIE and PIE were negligible.

Conclusions— Mediterranean diet has a direct effect on the risk of hip fracture via pathways other than through T2DM and BMI.
Identifying causes of effects with mediators

Monica Musio¹, Macartan Humphreys², Philip Dawid³

¹ University of Cagliari
² Columbia University and WZB Berlin
³ Cambridge University

X and Y are binary exposure and response variables, and we have full knowledge of the distribution of Y, given application of X. From this we know that the effect of X on Y is positive on average. We are now interested in assessing, for a case that was exposed and exhibited a positive response, whether it was the exposure that caused the response. The relevant “probability of causation”, PC, typically is not identified exactly by the distribution of Y given X, but bounds can be placed on it, and these bounds can be improved if we have further information about the causal process. Here we consider cases in which PC is not identified, and investigate the effect of observing a sequence of variables directly mediating between exposure and response. When these form a time-homogeneous Markov chain, positive evidence along the chain raises the lower bound on PC, and the longer the chain the higher is the lower bound. However, we never achieve point identification, even with unlimited mediators. The highest lower bound arises from a non-homogeneous chain with a single mediator. In this case PC is identified and is given by Pr(Y=0 | X=0).
The mediating effect of low-grade inflammatory markers on the association between sex and cardiac function and structure


1 Department of Epidemiology & Biostatistics, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, The Netherlands
2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
3 Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium
4 Department of Cardiology, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands
5 Department of Cardiology, Maastricht University Medical Center, Maastricht, The Netherlands
6 Department of Cardiology, CARIM School for Cardiovascular Diseases Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
7 Department of Nephrology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, The Netherlands

The prevalence of heart failure (HF) is increasing and over half of the HF patients have heart failure with preserved ejection fraction (HFpEF). Women are more frequently diagnosed with severe HFpEF. It is unclear which mechanism explains sex differences in HFpEF, but inflammation could account for these sex differences. This study aims to assess the mediating role of markers of low-grade inflammation and endothelial dysfunction on the associations between sex and echocardiographic markers in two prospective population-based cohorts: the Hoorn Study (n=290) and the Flemengho Study (n=353). Causal mediation analyses will be performed with the z-score derived from several inflammatory markers at baseline in tertiles and continuously, as mediator, sex as determinant and left ventricular ejection fraction (LVEF), left ventricular mass index (LVMI) and left atrial volume index (LAVI) at follow-up as outcomes. Additional linear regression analyses will be performed with the z-score as determinants and LVEF, LVMI and LAVI at follow-up as outcomes, adjusted for age, follow-up time, baseline echo values and BMI. This analysis will be tested for interaction between sex and the z-score. All analyses will be performed separately for the two cohorts and will be pooled to increase statistical power. At EUROCIM 2019, the results of the associations between sex and cardiac function and structure will be presented including an estimate of the mediation effect by low-grade inflammation. This study will provide a better understanding of the mechanisms that can explain the sex differences in HFpEF.
Management of colorectal cancer in older patients: Exploring the role of comorbidity and the diagnostic route using mediation analysis

Sara Benitez Majano¹, Stijn Vansteelandt¹ ², Chiara Di Girolamo¹, Sarah Walters¹, Bernard Rachet¹

¹ London School of Hygiene & Tropical Medicine
² Ghent University

Background
Older cancer patients often have fewer staging investigations and inferior treatment than younger patients. This suboptimal cancer management is commonly attributed to comorbidity, which may contraindicate some medical interventions. We aim to disentangle the effect of biological and chronological age on having complete staging investigations and radical surgery for colorectal cancer (CRC), and quantify the proportion mediated by health status and diagnosis route.

Methods
Population-based cancer registries provided information on CRC patients diagnosed in England during 2010-2012. Information on investigations, surgery and comorbidity was derived from national cancer registry and secondary care records. A counterfactual-based mediation analysis, allowing for multiple mediators, quantified the proportion of the age effect on staging investigations, and radical surgery mediated by health status, and by the diagnosis route. Novel sensitivity analysis techniques for multiple mediators were developed to assess the robustness of the findings against unmeasured confounding.

Results
There was a U-shape association with more complete investigations and radical surgery among those aged 60-69 years. The associations between exposure and outcomes (investigation and surgery) were barely mediated by health status, but was partly mediated by being diagnosed through an emergency admission. An important proportion of the age differential in cancer management was not mediated by these factors, especially in older patients.

Discussion
Extreme age is associated with suboptimal CRC management, and this was not explained by health status in elderly, contradicting prevailing beliefs. Although some patients may not benefit from treatment, having complete investigations is essential to plan optimal treatment, regardless of chronological age.
Mediation analysis of maternal depression and child neurodevelopment

Nicola Fitz-Simon¹,², Alina Rodriguez²,³

¹National University of Ireland Galway
²Imperial College London
³University of Lincoln

Maternal depression during pregnancy has been linked to compromised foetal development including the foetal brain, via biological mechanisms, and is related to behavioural, emotional, and cognitive problems, i.e. neurodevelopmental deficits, in childhood. However, depression during pregnancy also predicts depression after pregnancy, which may explain, or be affected by, childhood neurodevelopment. Moreover, outcomes measured by maternal report may be affected by the exposure. In this study we examine alternative mediation pathways between depression during pregnancy and child neurodevelopment deficits.

We analyse data from a Swedish mother-child cohort study, which has information on maternal depression and child neurodevelopment up to age 5, and confounding variables. The outcome, child behaviour score, is available by both mother and teacher report. We consider complex hypothetical interventions on two possible mediation pathways: via maternal depression trajectories post-pregnancy, and via irritability scores in early childhood.

Under the hypothetical intervention, we counterfactually shift the distribution of the mediators in the depressed mothers to that of the non-depressed mothers, conditional on confounders, to estimate an indirect effect.

In this cohort, maternal depression during pregnancy is associated with deficits in child neurodevelopment at age 5 by mother but not by teacher report. For the mother-reported outcome, the association is explained largely by maternal depression after pregnancy. Mediation analysis may help us to understand the observed differences for outcomes reported by mothers and teachers, with implications for design and interpretation of similar studies.
Estimation of natural direct and indirect effects in causal mediation analysis requires a set of assumptions, which includes a “cross-world” assumption of conditional independence between outcome and mediator counterfactuals in which treatment is set at two distinct values simultaneously—one value for the counterfactual outcome and the other for the counterfactual mediator. In the literature, this assumption is commonly described as the absence of confounders of the mediator-outcome relationship that are also affected by treatment. While this has the advantage of being straightforward to understand, others have pointed out that the cross-world assumption is more complex and can be violated even in the absence of such confounders. In this project, we aim to provide applied researchers with new insights and a better understanding of the cross-world assumption in two ways. First, we present and synthesize in one place the different ways the cross-world assumption can be understood, which up to now have been scattered across the literature. Second, we discuss the theoretical requirements of the cross-world assumption in the context of a series of realistic examples taken from gerontology research, starting with the absence of mediator-outcome confounders affected by treatment and ending with more general violations. Overall, we believe that our work will be of interest to any researcher who has done or plans to do a causal mediation analysis, and our hope is that it will improve applied researchers’ understanding of the cross-world assumption’s subtleties, thereby leading to more thoughtful and fruitful causal mediation analyses.

References:

Purpose: Real world evidence (RWE) pose additional methodological challenges for evaluating causality, including confounding, missing/misclassified data, no clear treatment assignment, dynamic treatment regimens, and switching. We aimed to assess the type and impact of biases potentially occurring when analyzing RWE using the case of ovarian cancer treatment.

Methods: We compared overall survival (OS) with and without second-line treatment in ovarian cancer patients (n=1581) using retrospective IMS Oncology electronic medical records. We identified potential confounding and other biases using directed acyclic graphs (DAG). To assess the biases, we applied several analytic approaches starting with simple Cox regression with and without baseline variables, including time-dependent covariates to reduce immortal time bias, applying the “target trial” approach, including pseudo-populations and marginal structural models with inverse probability of censoring weighting (IPCW) (causal). We compared hazard ratios (HR) and 95% confidence intervals (95%CI) to assess the bias associated with each of these approaches compared to the causal analysis.

Results: The crude and baseline-adjusted analyses yielded a HR for second-line versus no second-line therapy of 0.565 (95%CI 0.495-0.645) and 0.535 (95%CI 0.468-0.613), respectively. Including treatment as a time-dependent covariate to account for immortal time bias, the corresponding crude and adjusted HR increased to 1.665 (95%CI 1.459-1.901) and 1.683 (95%CI 1.407-2.014). Applying a causal (counterfactual) analysis using IPCW and replication yielded a HR of 1.067 (95%CI 1.020-1.115), which matched the results of a published randomized clinical trial.

Conclusions: When using routine RWE data, DAGs can guide the identification of potential biases and variables that need to be controlled for. In our analysis, potential biases were substantial with different directions. Only the application of the target trial and replication approach in combination with a causal analysis matched data from clinical trials.
Sequential Design of Experiments for Personalized Medicine

Mia Sato Tackney¹, Dave Woods², Kim May Lee³, Ilya Shpitser⁴, Peter Smith¹

¹ Social Statistics and Demography, Social Sciences - University of Southampton, Southampton, SO17 1BJ, United Kingdom,
² Mathematical Sciences - University of Southampton, Southampton, SO17 1BJ, United Kingdom
³ MRC Biostatistics Unit, School of Clinical Medicine - Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge, CB2 0SR, United Kingdom
⁴ Department of Computer Science - Johns Hopkins University, 160 Malone Hall, 3400 North Charles Street, Baltimore, MD 21218-2608, USA

Advances in genomics are making it possible to personalize medicine so that treatments are tailored to patients' genetic information. For example, cancers can now be characterized at the molecular level, and treatments can be targeted at specific genetic and biological mechanisms (biomarkers). Personalized clinical trials need to be able to identify effective treatment-biomarker combinations. We demonstrate a method of designing a sequential experiment with an adaptive treatment allocation scheme which seeks to both find effective treatments and estimate the corresponding treatment-biomarker interactions. This method weights the probabilities of treatment assignment according to an optimality criterion which takes into account the biomarkers, treatment assignment and response of the patients in the trial so far. We provide examples of both myopic and non-myopic sequential strategies. In the former, decisions on which treatment to apply to the current patient ignore any potential information about future patients; in the latter, we account for potential treatment allocations to future patients when choosing the treatment for the current patient. We describe some computational challenges in implementing the non-myopic method and, through simulation studies, we describe possible settings in which it may provide benefit over the myopic approach.
How robustly do we verify the assumptions of the causal inference framework? Qualitative methods can provide a more in-depth and informed assessment

Nadia Sourial, Isabelle Vedel, Tibor Schuster

Department of Family Medicine, McGill University, Montreal, Canada

Objectives
Developments in causal inference have made explicit three theoretical conditions (consistency, positivity and exchangeability) required for valid causal estimation. However, literature on how these conditions should be assessed in practice remains vague. We compared a “usual” vs qualitative approach to assess these conditions in the context of a health policy evaluation in Ontario, Canada.

Methods
The “usual” approach was based on the literature and colleagues/experts. The qualitative approach was based on interviews and focus groups with key stakeholders (patients, clinicians, researchers and managers). Eighteen interviews and one focus group with 14 participants were conducted. A summative content analysis of the transcribed interviews was conducted. Information on causal conditions obtained from the “usual” vs qualitative approach were compared.

Results
While a substantial amount of information was obtained through the “usual” approach, assessment of the plausibility of the causal conditions remained vague or incomplete. A higher degree of precision was attained through the qualitative approach. With regard to consistency, we found additional evidence of heterogeneity in the implementation and functioning of the policy. For positivity, we clarified the selection criteria applied in determining allocation to the exposure. Finally, we uncovered additional potential confounders and predictors of the outcome, leading to an extension of the initial directed acyclic graph.

Conclusion
This study demonstrated a feasible and rigorous qualitative approach to verifying causal conditions, expanding the scope and potential of mixed methods. Results demonstrated how qualitative methods can be used to better inform and strengthen quantitative analyses based on the causal inference framework.
Causal inference with missing values: Treatment effect estimation of tranexamic acid on mortality for traumatic brain injury patients

Imke Mayer¹,²,³, Julie Josse²,³, Jean-Pierre Nadal¹, Tobias Gauss⁴, Jean-Denis Moyer⁴

1: CAMS (École des Hautes Études en Sciences Sociales, PSL University, CNRS)
2: CMAP (École Polytechnique)
3: Inria XPOP
4: Traumabase® Group, Hôpital Beaujon (APHP)

Abstract: In healthcare or social sciences research, prospective observational studies are frequent, relatively easily put in place (compared to experimental randomized trial studies for instance) and can allow for different kinds of posterior analyses such as causal inferences. Average treatment effect (ATE) estimation for instance is possible through the use of propensity scores which allow to correct for treatment assignment biases in the non-randomized study design. However, a major caveat of large observational studies is their complexity and incompleteness: the covariates are often taken at different levels and stages, they can be heterogeneous – categorical, discrete, continuous – and almost inevitably contain missing values. The problem of missing values in causal inference has long been ignored and only recently gained some attention due to the non-negligible impacts in terms of power and bias induced by complete case analyses. We propose several consistent doubly robust average treatment effect estimators which directly account for missing values and compare them to complete case ATE estimators applied on imputed data, i.e. on complete data obtained by replacing every missing value by at least one plausible one, and to the recently proposed method of Kallus et al. [2018]. We assess the performance of our estimators on a large prognostic database containing detailed information about over 15,000 severely traumatized patients in France. Using the proposed ATE estimators and this database we study the effect on mortality of tranexamic acid administration to patients with traumatic brain injury in the context of critical care management.
Doubly robust inference procedures for analyzing the cancer registry data

Sho Komukai, Satoshi Hattori

Division of Biomedical statistics, Department of Integrated Medicine, Graduate School of Medicine, Osaka University

Cancer registrations are useful to various research questions for cancer populations. For example, CONCORD study was reported the comparisons for the impact of the cancer death among nations and among periods. The most important feature of the cancer registration is that the database has no information about cause of death. One of the measures for cancer prognosis in the analyses of the cancer registry data is a net survival which is defined as the survival probability if a patient would not die due to reasons other than the cancer. For the net survival, a doubly robust estimator was proposed in the presence of the covariate-dependent censoring (Komukai and Hattori, 2017). As alternative to the net survival, a relative survival ratio is also used. The relative survival is defined as the ratio between the survival probability of the cancer population and that of the general population. To make inference on these measures in the presence of covariate-dependent censoring, which we often encounter in practice, Kodre and Perme (2013) proposed the inverse probability of censoring method. In this research, motivated by success in causal inference, we propose a doubly robust estimator for the relative survival ratio. Although the inference for the net survival by Komukai and Hattori (2017) relies on untestable assumptions on dependence among time-to-events and censoring, it is possible to verify the underlying assumption in the inference for the relative survival ratio. We also propose a doubly robust test to assess the underlying assumption for relative survival ratio.
Causal inference for trends in disease incidence is important for public health. For example, when we observe a decrease in cancer incidence over time we would want to find out modifiable factors that have led to the decreased incidence. Further we need to inform public health policy makers how they can keep the decreasing trend in the future. However analytical approaches to address this question have not been established yet. One common approach has been using group-level data to find out associations between, e.g., changes in smoking prevalence and changes in cancer incidence. However this is prone to an ecological fallacy where a group-level inference may not match with an individual-level inference. Other approaches used individual-level data but they were mainly limited to building regression-based prediction models for disease incidence with and without a modifiable factor of interest and report whether adding or removing the factor changed the trends in disease incidence.

In this work we present a new approach for causal inference for trends in disease incidence and apply this to a longitudinal observational cohort using marginal structural models. Briefly we define trends in disease incidence as changes in age-specific disease incidence by birth cohort. Then we compare an observed incidence of one birth cohort to a counterfactual incidence of another birth cohort after a hypothetical intervention. We will also address typical issues of the real world observational data including selection bias before and after study entry, competing risks of death, irregular observation intervals and missing data.
Validating optimal treatment regimes using multiple imputation

Thomas Klausch, Peter van de Ven, Johannes Berkhof

Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute

A growing set of methods is developed for estimating individualized treatment decision rules, also called Optimal Treatment Regimes (OTR). A well-known problem of this methodology is that OTR policy values estimated on data used for building the OTR can have optimism bias and insufficient confidence interval coverage. We present a bias-adjusted inference method for static OTR using independent validation data. OTR are estimated from observational training data by Bayesian additive regression trees (OTR-BART) or generalised random forests (OTR-GRF). In a validation set, policy values are expectations of functions of the potential outcomes of treatments assigned by the OTR. These outcomes, however, are only available for those validation patients treated in accordance with the OTR. The emerging missing data problem is addressed by multiple imputation using predictive mean matching (MI-PMM) or BART (MI-BART). In simulations, we varied treatment effect heterogeneity, training and validation set sizes, and number of unrelated noise covariates. Policy values estimated by the validation approach with MI-PMM and MI-BART had negligible bias and smaller MSE than the approach using all data for, both, training and validation. Importantly, MI-derived confidence intervals had nominal coverage. To illustrate we use data from Oropharynx cancer patients treated with radiotherapy or chemo-radiotherapy (n=271). Three-year survival probability was 0.80. Direct estimation with OTR-BART on all data suggested an improvement to .84 (OTR-GRF: .83). Optimism-adjusted estimates for OTR-BART trained on a random half-split and validated on the second half by MI-PMM, however, gave .79 [.70, .87] and by MI-BART .81 [.73, .89] (OTR-GRF estimates similar).
Multi-State models and causal inference: Prevention effect on burden of hospital infections

Derek Hazard, Maja von Cube, Klaus Kaier, Martin Wolkewitz on behalf of the COMBACTE-MAGNET Consortium

Institute of Medical Biometry and Statistics; Division Methods in Clinical Epidemiology
Faculty of Medicine and Medical Center - University of Freiburg
Freiburg Center for Data Analysis and Modeling

Multi-state models are established tools that enable summary predictions in real world settings. Additionally, they can facilitate causal inference through prediction in artificially constructed hypothetical settings[1]. One important application is a hospital setting where mortality rates and lengths of hospital stays can be calculated through synthesizing transition rates among various states (e.g. death, discharge). Multi-state models provide an overview of the entire system and these rates can be manipulated by the researcher to produce counterfactual outcomes[2].

Using this causal approach, one can ascertain what effect the hypothetical (total or partial) elimination of hospital infection would have on hospital mortality and financial costs. We demonstrate how transition rates can be first estimated non-parametrically from hospital data, and then subsequently adjusted to calculate hypothetical summary measures. Three real data sets are used for demonstration. For example, the result of a prevention that cuts the infection rate in half would save 20 lives and reduce the amount of patient-days by 1 585 in a population of 10 000 German intensive care unit patients. Using German healthcare cost data, this entails a savings of EUR 2 269 720.

Prevention effects can be modelled to increase our understanding of not only the effect of different preventions (e.g. hand hygiene, antibiotic stewardship) on the same population, but also identical preventions within different populations (e.g. countries, hospital units). As a result, one can determine in which settings the preventions would bring the most benefit; a vital guidance for cost-benefit analyses in the fight against hospital infections.


Reducing selection bias: turning the underlying heterogeneity to your advantage

Rik van Eekelen

Centre for Reproductive Medicine, Academic Medical Centre, Amsterdam University Medical Centres, Amsterdam, the Netherlands

When there is underlying heterogeneity between patients in terms of the probability to experience a certain event, a selection process occurs over time in which only those with higher survival probabilities remain in the cohort. A clear definition of the time origin i.e. start of follow up is crucial in this context, but not all (observational) data sources start collecting data at the same time, e.g. the time origin of interest is diagnosis, but registry data only contains treated patients. Measured patient characteristics are often insufficient to explain this selection between diagnosis and the start of treatment. This could cause incomparability of patients from data sources that at first glance seem to have similar characteristics, thereby not allowing for any causal estimands such as treatment effect.

If the heterogeneity distribution is known or estimated, one could collapse all patient characteristics and expected selection due to heterogeneity in a summary score in which the (expected) time between diagnosis and treatment is embedded. Simple conditioning on this score in survival models such as a Cox proportional hazards model could potentially reduce the bias that this selection process induces. We show the motivating example for this approach in the field of fertility, in which we compared a database on expectant management (time origin: diagnosis) to a retrospectively collected database on in vitro fertilisation (time origin: treatment start) in a previous study. We apply our simple conditional approach to prospectively collected fertility data in which the time between diagnosis and treatment was known.
Complete inference of causal relationships in dynamical systems

Zoltán Somogyvári¹,², Zsigmond Benkő¹,², Ádám Zlatniczki¹,³, Dániel Fabó⁴, András Sólyom⁵, Loránd Erőss⁵,⁶, András Telcs¹,³,⁷

¹Department of Computational Sciences, Wigner Research Centre for Physics of the Hungarian Academy of Sciences, Hungary
²János Szentágothai Doctoral School of Neurosciences, Semmelweis University, Hungary
³Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary
⁴Epilepsy Center "Juhász Pál", National Institute of Clinical Neurosciences, Budapest, Hungary
⁵Department of Functional Neurosurgery, National Institute of Clinical Neurosciences, Budapest, Hungary;
⁶Faculty of Information Technology and Bionics, Péter Pázmány Catholic University, Budapest, Hungary;
⁷Department of Quantitative Methods, University of Pannonia, Faculty of Business and Economics, Veszprém, Hungary
⁸Neuromicrosystems ltd., Budapest, Hungary

This paper presents the Dimensional Causality (DC) analysis method devised to detect and quantify the probability of all possible types of causal relationships between two time series observed in deterministic dynamical systems: independence, direct or circular causal connection and particularly the existence of a hidden common cause. To detect these relations between two time series, Takens’ embedding theorem is used to reconstruct the attractors of the underlying systems. The new method is based on the subadditivity of the system’s attractor dimensions, where the key is the dimension of the joint attractor of the two systems. We showed that the relations between the joint and individual dimensions unequivocally determine the causal relations between the dynamical systems. The probability of the different causal relations is obtained via Bayesian inference.

We validated our method on simulated examples of classical chaotic and non-chaotic dynamical systems such as coupled Lorentz-systems, Logistic maps, or Hindmarsh-Rose models and demonstrated its capabilities on human neurophysiological measurements.

The method properly detected the increase of common cause probability between the two hemispheres during flashing light stimulation observing patients’ EEG signals. During presurgical investigation the possible focus of epileptic seizure is identified; an area which drives the others. The universality of our method ensures its applicability in many other fields of science.
A Bayesian multivariate factor analysis model for evaluating an intervention using observational time-series data on multiple outcomes

Pantelis Samartsidis, Shaun R. Seaman, Silvia Montagna, André Charlett, Matthew Hickman and Daniela de Angelis

MRC Biostatistics Unit, University of Cambridge

A problem frequently encountered in many areas of scientific research is that of estimating the impact of a non-randomised binary intervention on an outcome of interest using time-series data on units that received the intervention (‘treated’) and units that did not (‘controls’). One popular estimation method in this setting is based on the factor analysis (FA) model. The FA model is fitted to the pre-intervention outcome data in treated units and all the outcome data on controls units, and the counterfactual treatment-free post-intervention outcomes of the former are predicted from the fitted model. Intervention effects are estimated as the observed outcomes minus these predicted counterfactual outcomes.

In this work, we propose two extensions of the FA model for estimating intervention effects: 1) the joint modelling of multiple outcomes to exploit shared variability, and 2) an autoregressive structure in factors to account for temporal correlations in the outcome. By taking a Bayesian approach, we also account for uncertainty in the number of factors. A Markov chain Monte Carlo algorithm is developed to sample from the posterior distribution of the model parameters.

Using simulation studies, we show that the proposed methodology can reduce the bias and improve the precision of the intervention effect estimates. We apply our method to estimate the impact of stricter alcohol licensing policies on alcohol-related harms.
Inferring causal relations with exact cross-mapping

Marcell Stippinger, Zoltán Somogyvári

Department of Computational Sciences, MTA Wigner Research Centre for Physics, Budapest, Hungary

Convergent cross-mapping (CCM) is a method for inferring causal relations in dynamical systems. It relies on the time delay embedding of a time series to get data points in higher dimensions. The CCM method, as presented in (Sugihara et al., 2012), uses an approximate exponential kernel interpolation for the mapping. The performance of this interpolation is expected to degrade at the boundaries of the embedded manifold. In addition, as the embedding dimension increases, boundary effects get more pronounced as the manifold surface to volume ratio increases. To improve cross-mapping quality, we introduce exact cross-mapping that solves the problem of boundaries. In this exact method we explore the possibility of replacing the interpolation by extrapolation. More precisely, we solve the homogeneous linear system of equations for the $k$ nearest neighbors as seen from the point in question. The solution gives exact weights to use for the neighbors. This linear system, however, is often an ill-posed problem. Therefore, one has to use disciplined methods for the regularization that compensate for the shortcomings of extrapolation. In this setting, the Tikhonov regularization is adequate for balancing the accuracy of the analytic solution and the extremities allowed as weights. We demonstrate the results of exact cross-mapping using coupled model dynamical systems, such as the Logistic map, the Lorenz attractor, and Hindmarsh-Rose systems. Preliminary simulation show that our exact method converges faster or, alternatively, provides more robust results using the same amount of data.
Detecting frequency-dependent cortical interactions with topological causal inference techniques

Zsigmond Benkő¹, Jeroen J. Bos², Martin Vinck², Cyriel M.A. Pennartz², Umberto Olcese², Zoltan Somogyvari¹

¹Department of Computational Sciences, Wigner Research Centre for Physics of the HAS
²Swammerdam Institute for Life Sciences, University of Amsterdam

Advanced experimental techniques made possible the observation of neural dynamics on unprecedented temporal and spatial scale, however one needs to have proper data analysis techniques in order to extract functional information from massive neuronal datasets. Such techniques could be causal inference methods leveraging predictive (Granger causality, Transfer entropy) or topological (Convergent cross-mapping, Dimensional causality) reasoning to reveal causal relations from multivariate neural time series.

Several evidence suggests that slow-oscillatory activity modulates functional connections of fast-oscillatory neuronal populations between brain areas. Thus, to reveal the direction of inter- and intra-areal functional connections, frequency domain causal inference methods are needed.

Recently, Granger causality was shown not to be an appropriate tool to infer causal relationship in datasets with dominantly deterministic dynamics. Therefore, in this study we use topological causal inference methods extended to the frequency domain to infer causal relations between cortical areas.

Multivariate Local Field Potential (LFP) measurements were taken with tetrodes in the Barrel cortex and V1 of rodents during active wakefulness and sleep.

We decomposed the LFP signal into frequency bands (Delta 0.4-4 Hz, Theta 6-10 Hz, Beta 16-30 Hz, low-gamma 30-60 Hz, mid-gamma 60-100 Hz, high-gamma 120-250 Hz), then computed causal connectivity within and between cortical areas among signal frequency components using the Convergent Cross-Mapping (CCM) or the Dimensional Causality (DC) method and compared the connectivity-patterns computed for different conditions.

We found significant cross-spectral causal connections (gamma-beta, beta-theta) and significant differences between the connectivities (gamma to theta, beta to theta coupling) in different behavioral states.
Centrality measures as a proxy for causal influence? A cautionary tale

Fabian Dablander, Max Hinne

Department of Psychological Methods, University of Amsterdam

Network models have become a valuable tool in making sense of a diverse range of social, biological, and information systems. These models marry graph and probability theory to visualize, understand, and interpret variables and their relations as nodes and edges in a graph. For example, depression can be understood from this ‘network’ perspective as arising from the causal interplay between symptoms (which constitute nodes) such as sadness, fatigue, and lack of concentration. Many applications of network models rely on undirected graphs in which the absence of an edge between two nodes encodes conditional independence between the corresponding variables. To gauge the importance of nodes in such a network, various node centrality measures have become widely used, especially in psychology, psychiatry, and neuroscience. The intuition is that central nodes in a network are important in a causal sense, an interpretation applied researchers frequently engage in. Using the dominant causal framework based on directed acyclic graphs (DAGs), we show by simulation that the relation between causal influence and node centrality measures is not straightforward. In particular, the correlation between causal influence and several node centrality measures is weak, except for eigenvector centrality. Our results provide a cautionary tale: if the underlying real-world system can be modeled as a DAG but researchers estimate undirected graphs and subsequently interpret nodes with high centrality as causally important, then this may result in sub-optimal interventions.
Collider and reporting biases involved in the analyses of cause of death associations in death certificates: An illustration with cancer and suicide

Moussa Laanani$^{1,2,3}$, Vivian Viallon$^4$, Joël Costé$^2$, Grégoire Rey$^3$

$^1$CépiDc-Inserm (French Center for Epidemiology on Medical Causes of Death), Le Kremlin-Bicêtre, France
$^2$Biostatistics and Epidemiology Department, Cochin Hospital, AP-HP and Paris Descartes University, Paris, France
$^3$Department of Public Health Studies, French National Health Insurance, Paris, France
$^4$Nutritional Methodology and Biostatistics Group, International Agency for Research on Cancer (IARC), Lyon, France

Background: Data from death certificates have long been studied to explore causal associations between diseases. However, these analyses are subject to collider and reporting biases (selection and information biases, respectively). We aimed to assess to what extent associations of causes of death estimated from individual mortality data can be extrapolated to the general population.

Framework: We used a multistate model to generate populations of individuals and simulate their health states up to death from national health statistics and artificially replicate collider bias. Associations between health states can then be estimated from such simulated deaths by logistic regression and the magnitude of collider bias assessed. Reporting bias can then be approximated by comparing the estimates obtained from the observed data (subject to collider and reporting biases) with those obtained from the simulated data (subject to collider bias only).

Application: The cancer/suicide association was negative when assessed using data from death certificates. Collider bias, due to conditioning inclusion in the study population on death increasingly reversed the associations with cancer lethality. Reporting bias was much stronger than collider bias and depended on the cancer site, but not prognosis.

Conclusions: The analyses of cause of death associations exclusively from death certificates should be performed after an assessment of the magnitude of both collider and reporting biases. If they cannot be corrected, results from these analyses should not be extrapolated to the general population.
When weighting goes wrong: The implications of M-bias for analyzing survey data

Alexander Murray-Watters

GESIS – Leibniz Institute for the Social Sciences, Survey Design and Methodology

Weighting and complex survey designs are common techniques for improving the accuracy and efficiency of survey data. As both methods rely on auxiliary information, their joint use in, say, a regression, can result in bias due to conditioning on a collider (i.e., m-bias). M-bias in survey data is particularly nefarious, as there may not be a way to correct for the bias, especially if a relevant collider was conditioned on as part of the survey gathering process (e.g., the collider is a strata in a stratified sample). Worse, knowing that a variable is a collider for a particular regression requires some knowledge of the underlying causal structure, which for many research questions is unknown, or is part of the reason the survey data is being gathered in the first place. This paper discusses several exploratory simulations of the collider bias problem for survey weighting and their implications for analyzing survey data. It explores methods for partially ameliorating the m-bias problem for survey weighting via the use of causal search algorithms. Taking the underlying causal structure into account when gathering and using survey data reduces or eliminates an unnecessary source of bias, increasing the accuracy of effect estimates while not increasing the cost of gathering data. The paper also attempts to bridge the gap between the causality and survey literatures, corrects some common misconceptions about weighting, and suggests ways to bring researchers from both fields together.
Simulation-based sensitivity analysis for interference in observational studies with unmeasured links

Laura Forastiere¹, Alessandra Mattei², Fabrizia Mealli²

¹Yale Institute for Network Science, Yale University
²Department of Statistics, University of Florence

Oftentimes estimates on the treatment effect are based on the no-interference assumption, i.e., a subject’s outcome only depends on the treatment he received and not on other subjects’ treatments. However, interference is common in many social settings. When data are wrongly analyzed under the “no-interference assumption,” very misleading inferences can result. The bias depends on the level of interference but also on the association between a unit’s treatment and the treatment received by his neighbors. In observational studies, information on links between units is usually unavailable and interference cannot be taken into account. We develop a Monte Carlo sensitivity analysis to the violation of the “no-interference assumption”. A Monte Carlo sensitivity analysis is a simulation-based approach, which repeatedly i) draws a set of sensitivity parameters, ii) simulates potential links, and iii) re-estimates the effect of interest after adjusting for interference. We propose a model to generate the unmeasured links, which carries our belief on the level of interference and on the level of association between the individual and the neighborhood treatments. If we assume interference to operate only through a function of the vector of neighbors’ treatments, after a network is drawn we can compute such function and estimate the direct effect of the treatment taking interference into account. Different functions can be used. This approach has the additional advantage of adjusting for neighborhood and network covariates.
Disentangling spillover effects inside the principal strata in the presence of network data, with application to a field experiment on teens’ museum attendance

Silvia Noirjean

University of Florence – Dept. of Statistics, Informatics, Applications

The principal stratification approach is usually used to explore the causal pathways that may be triggered by the treatment among latent subgroups of subjects. More in detail, principal stratification enables to recover stratum-specific causal effects that may pass (associative effects) or not pass (dissociative effects) through the treatment uptake. However, when one allows for interference, such stratum-specific effects can be mixtures of the treatment assigned and the spillover effect for those who did not receive the treatment, and of treatment assigned, the treatment uptake and the spillover effect for those who received the treatment: this makes these estimates not easy to interpret. We show how network data can be used to disentangle spillover effects inside the principal strata: specifically we use friends’ behavioral variables, under different treatment levels, as possible “mediators” of spillover effects. To illustrate our original methodological contributions, we revisit results from a field experiment, conducted in Florence, Italy, to study the effects of different incentives to motivate high school teens to visit art museums. In the experiment, based on a clustered encouragement design, classes of students were assigned at random to one of three incremental encouragements and were offered a free visit to a main museum in the city (the treatment), in the hope to increase their future museum attendance (the outcome). Data on friendship networks were collected prior to encouragement delivery. To face identification issues, estimation is performed with Bayesian inferential methods.
This document builds on recent developments in causal inference in the presence of interference and uses a coarsened exact matching algorithm to determine the effect of being exposed to treatment in a network setting. The algorithm is applied to a simulation designed to replicate an observational study where there is no control over selection into treatment. In the simulation different kinds of treatment are considered taking into account the position of an individual in the network, the interference is carried out through this mechanism. The same algorithm is then applied to the Sexual Transmitted infections And Sexual Health (STASH) dataset. In this study, peer supporter students were nominated and given instructions to pass on knowledge about sexual health to their peers in different ways. The algorithm compares what happens to students who were connected to a peer supporter to ones that were not. Results suggest that the intervention had a positive impact on several variables related to attitudes to sexual health, and that variables have a different way of spreading through the network.
Data mining and machine learning techniques such as classification and regression trees (CART) represent a promising alternative to conventional logistic regression for propensity score estimation. Whereas incomplete data preclude the fitting of a logistic regression on all subjects, CART is appealing in part because some implementations allow for incomplete records to be incorporated in the tree fitting and provide propensity score estimates for all subjects. Based on theoretical considerations, we argue that the automatic handling of missing data by CART may however not be appropriate. Using a series of simulation experiments, we examined the performance of different approaches to handling missing covariate data; (i) applying the CART algorithm directly to the (partially) incomplete data, (ii) complete case analysis, and (iii) multiple imputation. Performance was assessed in terms of bias in estimating exposure-outcome effects among the exposed, standard error, mean squared error and coverage. Applying the CART algorithm directly to incomplete data resulted in bias, even in scenarios where data were missing completely at random. Overall, multiple imputation followed by CART resulted in the best performance. Our study showed that automatic handling of missing data in CART can cause serious bias and does not outperform multiple imputation as a means to account for missing data.
How to handle missing data in propensity score analyses: a simulation study

Jungyeon Choi¹, Olaf M. Dekkers¹,², Saskia le Cessie¹,³

¹Department of Clinical Epidemiology, Leiden University Medical Center
²Department of Endocrinology and Metabolism, Leiden University Medical Center
³Department of Biomedical Data sciences, Leiden University Medical Center

A challenge in propensity methods is missing values in confounders. Several strategies for handling missing values exist: 1) complete case analysis, 2) missing indicator method, 3) multiple imputation and 4) combining multiple imputation and missing indicator method.

In this simulation study, we compared these four strategies of handling missing covariate values in propensity matching and propensity weighting. Concurrently, we aimed to provide guidance in choosing the optimal strategy. Simulated scenarios varied regarding missing mechanism, presence of effect modification or unmeasured confounding. Additionally, we demonstrated how missingness graphs help clarifying the missing structure.

When no effect modification existed, complete case analysis yielded valid causal treatment effects even when data were missing not at random. In some situations, complete case analysis was also able to partially correct for unmeasured confounding. Multiple imputation worked well if the data were missing (completely) at random, and if the imputation model was correctly specified. In the presence of effect modification, more complex imputation models than default options of commonly used statistical software were required. Multiple imputation may fail when data are missing not at random. Here, combining multiple imputation and the missing indicator method reduced the bias as the missing indicator variable can be a proxy for unobserved confounding.

The optimal way to handle missing values in covariates of propensity score models depends on the missing data structure and the presence of effect modification. When effect modification is present, default settings of imputation methods may yield biased results even if data are missing at random.
A comparison of outcome-related diagnostics for propensity scores

Emily Granger¹, Dr. Jamie C Sergeant¹², Dr. Mark Lunt¹

¹Arthritis Research UK Centre for Epidemiology, Division of Musculoskeletal and Dermatological Sciences, Manchester Academic Health Science Centre, The University of Manchester
²Centre for Biostatistics, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester

Whilst propensity scores are commonly used to balance covariate distributions between exposure groups, it is not known how best to assess their performance. Outcome-related diagnostics which prioritise balance on covariates strongly associated with outcome would be useful, since these covariates contribute most towards confounding bias. Such diagnostics have been proposed¹³, however there is limited research comparing their performances.

A simulation study was conducted to compare seven outcome-related diagnostics in terms of their correlation with bias. Diagnostics included the standardised difference in prognostic scores¹ between exposure groups and six weighted averages of balance across covariates. Weighted averages were defined by the combinations of three balance metrics (standardised difference, Kolmogorov-Smirnov statistic and overlapping coefficient) and two weighting methods. For covariate 𝑥𝑖, Weights 1 (2) were defined as the 𝑖𝑡ℎ coefficient obtained after regressing the outcome on 𝑥𝑖 (all covariates), multiplied by the standard deviation of 𝑥𝑖.

Using the Kolmogorov-Smirnov statistic or overlapping coefficient led to negligible correlation with bias in all scenarios. For scenarios with independent covariates and linear outcomes, prognostic scores and weighted standardised differences (using either weights) performed well. When covariates were correlated, Weights 1 performed poorly. For nonlinear outcomes, only correctly estimated prognostic scores performed well (weighted averages and linear prognostic scores obtained weak correlations).

A correctly estimated prognostic score would be a useful diagnostic for assessing how well propensity scores have removed confounding bias. In real data however, the true prognostic scores are unknown and it is unclear what the best approximation to these values would be.

Ordered matching for incomplete matching problems: a gender gap study

Massimo Cannas

University of Cagliari (Italy)

Matching algorithms are commonly used for balancing covariates in observational studies but they may give unsatisfactory results when a complete matching is not possible. While greedy matching algorithms can be customized to account for matching priority optimal matching algorithms find a minimum cost matching of maximum size assuming matching priority is the same for all units. We recall a classic combinatorial result showing the existence of a maximum size ordered matching coinciding with the classic one in the equal priority case. We describe an algorithm for finding this matching, analyze its properties and discuss statistical applications. Software implementation is described through a case study concerning gender gap of ranked women and men executives.
Risk of acute kidney injury with SGLT2 inhibitors compared to sulfonylureas in patients with type 2 diabetes

Ghadeer Dawwas, Steve M. Smith

University of Florida, College of Pharmacy

Background
Recent reports submitted to the Food and Drug Administration suggested a potential association between sodium glucose cotransporter-2 (SGLT2) inhibitors and the risk of acute kidney injury (AKI) but the association remains unclear. We aimed to assess the risk of AKI with SGLT2 inhibitors vs. sulfonylureas in patients with type 2 diabetes.

Methods
A retrospective cohort analysis using Truven Commercial and Medicare supplemental files was conducted among patients with type 2 diabetes. Patients who initiated SGLT2 inhibitors or sulfonylureas between 2013-2015 who had a diagnosis of type 2 diabetes within 12 months of treatment initiation were included. The risk of AKI was compared between users of SGLT2 inhibitors and sulfonylureas. Cox-proportional hazard model after propensity score matching (PSM) was used to obtain the hazard ratio (HR) and 95% confidence interval (CI).

Results
After PSM, a total of 127,440 patients were included in the analysis (n= 63,720 for SGLT2 and n= 63,720 for sulfonylureas). The incidence of AKI was 2.3 and 4.5 per 1000 person-years in the SGLT2 inhibitors and sulfonylureas groups, respectively. In Cox proportional hazard model, the use of SGLT2 inhibitors was associated with a decreased risk of AKI compared to sulfonylureas (HR, 0.50, 95% CI, 0.41; 0.61).

Conclusion
In this population-based cohort of patients with type 2 diabetes, the use of SGLT2 inhibitors was associated with lower risk of AKI compared to sulfonylureas.
Causal inference for spatiotemporal epidemiological data

Rohan Arambepola, Ewan Cameron, Peter Gething

Malaria Atlas Project, Big Data Institute, University of Oxford

Understanding causal relationships between environmental and socio-economic variables, public health interventions and disease risk is an essential starting point for answering many health policy questions. However, applying causal inference algorithms to real life epidemiological data is challenging. Common independence and conditional independence tests may be inappropriate for data that is not Gaussian (such as case counts or categorical variables) and where there exists non-linear relationships between variables. Furthermore, spatiotemporal autocorrelation in the data can lead to spurious inferred relationships. To address these challenges, our methods implemented the PC algorithm using spatiotemporal pre-whitening and scalable non-parametric kernel-based independence and conditional independence tests.

The first application of these methods aims to identify important causal covariates for modelling malaria incidence at a high temporal resolution. We applied the PC algorithm to monthly health-facility case data from Madagascar and environmental and socio-economic covariates at various time lags to identify the drivers of malaria incidence in the country and the timescales on which they act. We plan to compare the predictive performance of spatiotemporal models using these covariates with more common covariate selection techniques, particularly the forward predictive accuracy which may have applications to malaria early warning systems.

Our second application is to estimate the efficacy of interventions. We have applied the PC algorithm to datasets that include malaria interventions, such as insecticide treated bednets, in order to identify potential confounders. We plan to use non-parametric regression to control for these confounders in order to assess the effect of these interventions on malaria risk.
We investigate the finite sample performance of causal machine learning estimators for heterogeneous causal effects at different aggregation levels. We employ an Empirical Monte Carlo Study that relies on arguably realistic data generating processes (DGPs) based on actual economic data. We consider 24 different DGPs, eleven different causal machine learning estimators, and three aggregation levels of the estimated effects. In the main DGPs, we allow for selection into treatment based on a rich set of observable covariates. We provide evidence that the estimators can be categorized into three groups. The first group performs consistently well across all DGPs and aggregation levels. These estimators have multiple steps to account for the selection into the treatment and the outcome process. The second group shows competitive performance only for particular DGPs. The third group is clearly outperformed by the other estimators.
Using machine learning methods to improve propensity score estimation in observational studies when the treatment is rare

Ian Nason, Rebecca Gourevitch, Noémie Sportiche

All authors are PhD candidates in the Department of Health Care Policy at Harvard Medical School in Boston, Massachusetts

The propensity score is central to methods for causal inference in observational studies. However, little research has addressed performance of these methods when the sample sizes of the treated and control groups are imbalanced, which can lead to biased or inefficient casual effect estimation. Resampling methods developed in machine learning can ameliorate these issues. We bridge these two literatures and apply three resampling methods to class imbalanced data, using simulated scenarios and applied data, prior to estimating average treatment effects (ATE) using inverse propensity score weighting (IPW). We find that resampling methods reduce bias and improve efficiency of our ATE estimates. Among the resampling methods, synthetic minority oversampling technique (SMOTE) yields the greatest improvements in bias and efficiency, and under-sampling yields the best balance on covariates as measured by the average difference in means. These results are largely robust to violations of ignorability and to misspecification of the propensity score. Future work will examine the robustness of these results to additional datasets, estimation techniques, and resampling methods.
Different approaches to minimise confounding when emulating a surgical randomised clinical trial: an application to partial vs total knee replacement

Albert Prats-Uribe, Victoria Y Strauss, Klara Berencsi, Daniel Prieto-Alhambra

Musculoskeletal Pharmaco- and device epidemiology Group, Centre for Statistics in Medicine, NDORMS, University of Oxford

Real world evidence has been proposed as an efficient alternative to costly and lengthy randomised clinical trials (RCT), particularly for surgical interventions. However, lack of randomisation in such observational studies makes them more vulnerable to confounding. We aimed to test the performance of several analytical approaches to emulate an ongoing surgical RCT comparing partial to total knee replacement, the Total or Partial Knee Arthroplasty Trial (TOPKAT).

TOPKAT randomly allocated 264 patients into each arm, and primary outcome was one-year post-operative patient reported outcomes (PROMs) Oxford Knee Score. In the observational analysis, we aimed to replicate TOPKAT using participants in the UK National Joint Registry (NJR) linked to Hospital records and national PROMs databases. NJR participants eligible for TOPKAT were included. Propensity score (PS) matching, stratification, adjustment and inverse probability weighting with linear mixed models were used to evaluate the treatment effect for comparable subjects according to available confounders. Instrumental variables (IV) were also evaluated including surgeon preference, surgical experience, surgical volume, geographical location, and calendar time. Only IVs fulfilling the underlying testable assumptions were used in two-stage IV regression.

We finally compared TOPKAT findings with those obtained from each of the proposed analyses. A Chi-square test for heterogeneity was used to test for significant differences between TOPKAT and each of the proposed methods.
The heart of the scientific enterprise is a rational effort to understand the causes behind the phenomena we observe. In disciplines dealing with complex dynamical systems, such as the Earth system, replicated real experiments are rarely feasible. However, a rapidly increasing amount of observational and simulated data opens up the use of novel data-driven causal inference methods beyond the commonly adopted correlation techniques. The key idea shared by several approaches is that, while the truism “correlation does not imply causation” holds, causal relations among variables can be estimated from their joint probability distribution given some assumptions. Causal inference is indeed a rapidly growing field with enormous potential to help answer long-standing scientific questions. Unfortunately, many methods are still little known and therefore rarely adopted in Earth system sciences. In this talk I will present an upcoming Perspective Paper in Nature Communications which identifies key tasks and major challenges where causal methods have the potential to greatly advance the state-of-the-art in Earth system sciences. I will also present a novel causal inference benchmark platform that aims to assess the performance of causal inference methods and to help practitioners choose the right method for a particular problem. Several methods will be illustrated by “success” examples where causal methods have already led to novel insights in Earth sciences and I will close with an outlook of this exiting field.

A Trump Effect on the EU’s popularity? The U.S. presidential election as a natural experiment

Lara Minkus¹, Emanuel Deutschmann², Jan Delhey³

¹University of Bremen
²European University Institute
³Otto-von-Guericke University Magdeburg

Did the election of Donald Trump affect the popularity of the European Union (EU) in Europe? Theoretically, both a positive rally effect (due to a perceived external threat) and a negative domino effect (due to resignation among Europhiles and/or reinforcement among eurosceptic nationalists) are plausible. We treat Trump’s unexpected victory as an external shock and use a Eurobarometer survey that was conducted in all EU-28 member states four days prior to (control group) and six days after the election (treatment group) as source material for a natural experiment. The analysis reveals that the election of Trump caused a significant increase in the EU’s popularity in Europe immediately after the election. This “Trump effect” is considerable in size, roughly equivalent to three years of education. Gains in popularity were particularly high among respondents who perceived their country as economically struggling and, surprisingly, among the political right, suggesting that Trump’s victory broadened and ideologically diversified the EU’s base of support.
A call for counterfactual reasoning when predicting patient prognosis

Romin Pajouheshnia¹, Noah A Schuster², Rolf HH Groenwold³, Linda M Peelen⁴

1. Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University
2. Division of Epidemiology and Biostatistics, VU University Medical Center
3. Department of Clinical Epidemiology, Leiden University Medical Centre
4. Julius Center, University Medical Center Utrecht

Causal reasoning is typically not considered in models for making clinical predictions. Predictive (or prognostic) models can be used to estimate the probability of future health outcomes for patients (e.g. mortality risk or survival time) to guide interventions or treatments. The implicit purpose of these models is to predict a patient’s health outcome if they were or were not to receive a certain intervention [1]. Conventional approaches to develop a prediction model do not address this counterfactual scenario and can produce models with poor predictive accuracy [2].

We demonstrate the need for counterfactual reasoning when developing a prediction model that will be used to guide (treatment) decisions. As an example, we developed a model to predict outcomes in the absence of beta-blocker therapy, for patients with chronic obstructive pulmonary disorder, using electronic health record data. First, we defined our estimand as the probability \( P(Y_{T=0} = 1 | X) \), where \( Y_{T=0} \) is the outcome if individuals were to remain untreated and \( X \) is a vector of predictors. Using DAGs, we assessed the validity of four candidate estimators: 1) a model ignoring treatment, 2) a model censoring treatment users, 3) a model conditioning on treatment use, and 4) an inverse probability of treatment weighted (IPTW) estimator. We found that the IPTW estimator was the most suitable estimator and we will present its application.

Our findings provide motivation for further consideration of counterfactual reasoning and methods for causal inference in prediction modelling.
