
BOOK OF ABSTRACTS

EUROPEAN CAUSAL INFERENCE MEETING

EUROCIM 2018

Florence, 11-13 April 2018

Posters

Contents

Effect of socio-economic status on access to a high volume hospital in a cohort of breast cancer patients: a causal inference approach (<i>Anita Andreano, Antonio Giampiero Russo, Maria Grazia Valsecchi</i>)	1
Estimating causal effects in complex systems (<i>KF Arnold ; WJ Harrison ; MS Gilthorpe ; A Heppenstalla</i>)	1
Integrating experts knowledge constraint in causal graph structure learning: extension to repeated measurement of immunological markers (<i>VahĀl Asvatourian, Philippe Leray, Stefan Michiels, Emilie Lanoy</i>)	2
Stable ipw estimation for longitudinal studies (<i>Vahe Avagyan , Stijn Vansteelandt</i>)	3
Assessing causal effects of retirement on mental health across european countries (<i>Marina Ballerini, Laura Forastiere, Alessandra Mattei</i>)	3
An extension of the causal tree framework in the presence of irregular assignment mechanisms (<i>Falco J. Bargagli Stoffi, Giorgio Gnecco</i>)	4
Assessing the impact of non-randomised drop-out on treatment switching adjustment in the relapsing-remitting multiple sclerosis clarity trial and the clarity extension study (<i>Helen Bell Gorrod , Nicholas R. Latimer , Doris Damian , Gerard Harty , Robert Hettle , Schiffo L. Wong</i>)	4
Mathematical coupling of proportions with common denominators: revisiting pearson, neyman, and fisher with causal graphs (<i>Laurie Berrie, Peter WG Tennant, Paul D Norman, Paul D Baxter, Mark S Gilthorpe</i>)	5
Education and life-expectancy and how it is mediated: evidence from the uk using principal strata (<i>Govert E. Bijwaard and Andrew M. Jones</i>)	6
Matching, differencing on repeat propensity score matching and difference-in-differences with repeated crosssectional data: methodological guidance and an empirical application in education (<i>Michele Binci, Madhumitha Hebbar, Paul Jasper and Georgina Rawle</i>)	6
The Causal Impact of Anthropogenic Emissions on Global Warming (<i>Valentina Angius and Chiara Binelli</i>)	7
Randomization-based Inference for Bernoulli Trial Experiments and Implications for Observational Studies (<i>Zach Branson and Marie-Abele Bind</i>)	7
On the Support of Matching Algorithms (<i>Massimo Cannas and Gavino Puggioni</i>)	8
The Effect of Unilateral Divorce on Prostitution: Evidence from Divorce Laws in U.S. States (<i>Riccardo Ciacci</i>)	8
How to give support to fundamental conditions in a cause of effect problem? (<i>Fabio Corradi and Monica Musio</i>)	8

Maternal Stress and Pregnancy Outcomes Evidence from the 2004 Madrid Bombing (<i>Marco Cozzani, Moris Triventi and Fabrizio Bernardi</i>)	9
Decision theoretic foundations for statistical causality (<i>A. P. Dawid</i>)	10
Impact of spasticity on functioning in spinal cord injury: an application of graphical modelling (<i>Cristina Ehrmann, Seyed Mahdi Mahmoudi, Kerstin Hug, Birgit Prodinge, Carlote Kiekens, Per Ertzgaard</i>)	10
The Effect of Unemployment on the Smoking Behavior in Couples (<i>Jakob Everding and Jan Marcus</i>)	11
A systematic map of within-study comparisons: examining bias in non-randomised studies (<i>Paul FENTON VILLAR and Hugh WADDINGTON</i>)	12
Propensity score diagnostics: the challenges we face when providing evidence that a propensity based estimate is unbiased (<i>Emily Granger, Dr. Jamie C Sergeant, Dr. Mark Lunt</i>)	12
Causal inference models comparing the effectiveness of clinical procedures in babies born with complex cardiac abnormalities (<i>Wendy J Harrison, Ngoni K Zava, James R Bentham</i>)	13
Causal inference from treatment-control studies having a pseudo-factor with unknown assignment mechanism (<i>Nicole Pashley and Kristen Hunter (presenters) with Katy McKeough and Tirthankar Dasgupta</i>)	14
Estimating treatment effects from observational data using instrumental variables (<i>John ER, Sheehan NA, Abrams KR, Brightling CE</i>)	15
Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach (<i>Michael C. Knaus, Michael Lechner and Anthony Strittmatter</i>)	15
Decomposing zero: the (non-) effect of health insurance on infant mortality in indonesia, explained by mediation analysis (<i>Noemi Kreif, Kala Diaz-Ordaz, Andrew Mirelman, Marc Suhrcke</i>)	16
Comparison of causal inference methods for exposome research (<i>Jessica E. Laine, Giovanni Fiorito, Paolo Vineis, and Exposomics Consortium</i>)	17
A bayesian approach to surrogate endpoints assessment through principal stratification with application to a clinical trial for postoperative pain control (<i>Andrea Lazzerini</i>)	17
Applying causal mediation analyses to randomised controlled trials in health-care. two case examples (<i>Hopin Lee, PhD, Sallie E Lamb, DPhil, Robert D Herbert, PhD, James H McAuley, PhD</i>)	18
Statistical assessment of strict population overlap in observational studies (<i>Lihua Lei, Alexander D'Amour, Peng Ding, Avi Feller and Jasjeet Sekhon</i>)	19
Process evaluations of complex interventions using causal inference (<i>Ruijie Li, Kimberley Goldsmith, Sabine Landau</i>)	19
Bayesian principal stratification analysis in Regression Discontinuity Designs with multiple outcomes: An application to students' academic performances (<i>F. Licari and A. Mattei</i>)	20
Causal mediation analysis for randomised studies with longitudinal data using structural equation modelling (<i>Wen Wei Loh, Beatrijs Moerkerke, Tom Loey, Stijn Vansteelandt</i>)	21
Performance assessment as an application of causal analysis (<i>Nicholas T. Longford</i>)	21

The problem of patient localization in emergency care (<i>Elena Lucchese</i>)	22
A dynamic definition of linear Markovian structural causal models (<i>Alessandro Magrini</i>)	22
The effects of a new tramway line on the commercial vitality of peripheral neighborhoods (<i>Giulio Grossi, Patrizia Lattarulo, Marco Mariani, Alessandra Mattei, ÁÚzge ÁŰner</i>)	23
Misclassification of exposures in sufficient cause interaction analysis (<i>Murthy N Mittinty</i>)	23
Detecting plasma proteins causally related to ms: a mendelian randomization approach (<i>Morani G., Fazia T., Bernardinelli L.</i>)	24
The causal effect of education on diseases in old age (<i>Gianmaria Niccodemi, Govert E. Bijwaard</i>)	25
Impact of bolsa familia program on new case detection of leprosy in brazil (<i>Julia Moreira Pescarini, Laura Forastiere, Joilda Silva Nery, Elizabeth B. Brickley, Mauricio Lima Barreto, Laura C. Rodrigues, Gerson Oliveira Penna</i>) .	25
Goodbye Smokers Corner: Health Effects of School Smoking Bans (<i>Gregor Pfeifer, Mirjam Reutter, Kristina Strohmaier</i>)	26
Impact evaluation for small areas (<i>Setareh Ranjbar, Nicola Salvati</i>)	26
Matching with text data: an experimental evaluation of methods for matching documents and of measuring match quality (<i>Reagan Mozer (Rose), Luke Miratrix, Aaron Russell Kaufman, and L. Jason Anastasopoulos</i>)	27
The Morning After: Prescription-Free Access to Emergency Contraceptive Pills (<i>Gregor Pfeifer, Mirjam Reutter</i>)	28
Data Based Covariate Selection for High Dimension Low Sample Size Data (<i>Lu Rui and Bryan Keller</i>)	28
Continuous-time msms: weight estimation and comparison to IPTW (<i>Pal Christie Ryalen, Mats Julius Stensrud and Kjetil Rysland</i>)	29
Pseval: an r package for evaluating principle surrogates (<i>Michael C Sachs, Erin E Gabriel</i>)	29
Help, I am Disabled and Looking for a Regular Job! Who will Help me After January 1st 2015? (<i>Klarita Sadiraj, Stella Hoff, Maroesjka Versantvoort</i>) .	30
Individual heterogeneity and time-varying cardiovascular risk factors contribute to explain the obesity paradox in mortality: results from a hierarchical joint modelling approach (<i>Juan G. Sanchez-Plaza, Paloma Saez-Diaz, Ana Arcon-Fernandez, Gernot Pichler, Jose Bermudeza, Josep Redona, Maria Tellez-Plazab.</i>)	30
Differences in conditional versus marginal estimates in collapsible estimators (<i>Ian Shrier, Annabelle Redelmeier, Russell Steele, Mireille Schnitzer</i>)	31
The estimation of neighbourhood effect: comparing logistic regression and multitreatment inverse propensity score weighting (<i>Silan Margherita, Arpino Bruno, Boccuzzo Giovanna</i>)	32
From average treatment effects to the identification of treatment responders (<i>Dr Joel Smith, Prof. John Forbes, Dr Anna OÁŠDonnell, Prof. Alastair Gray, Prof. Peter Sandercock on behalf of the IST-3 Collaborative Group</i>) .	32
A causal model of cell transformation assays in multi-laboratory experiments (<i>Federico M. Stefanini</i>)	33

Causal inference in the relative survival framework (<i>Elisavet Syriopoulou, Mark J Rutherford, Paul C Lambert</i>)	34
Impact of declining exposure to secondhand tobacco smoke in public places to decreasing smoking-related cancer mortality in the US population. (<i>Esther Garcia-Esquinas, Angelica Jimenez, Roberto Pastor-Barriuso, Maria Grau-Perez, Miranda R. Jones, Beatriz Perez-Gomez, Ana Navas-Acien, Maria Tellez-Plaza</i>)	34
Simplification in causal inference (<i>Santtu Tikka and Juha Karvanen</i>)	35
Effects of maternal education on early non-monetary investments in child development (<i>Emine Kubra Usta</i>)	36
Causal time series segmentation (<i>Aleksander Wieczorek and Volker Roth</i>)	36
Evaluating evidence of mechanisms in medicine: principles and procedures (<i>Jon Williamson</i>)	37
Matching on generalized propensity scores with continuous treatments (<i>Xiao Wu, Fabrizia Mealli, Danielle Braun, Francesca Dominici</i>)	37

Effect of socio-economic status on access to a high volume hospital in a cohort of breast cancer patients: a causal inference approach

Anita Andreano^{1,2}, Antonio Giampiero Russo², Maria Grazia Valsecchi

¹ Center of Biostatistics for Clinical epidemiology, School of Medicine and Surgery, University of Milano Bicocca

² Epidemiology Unit, Agency for Health Protection of Milan

Aim: To investigate the effect of social status (SS) on receiving surgery at a low volume hospital (LVH) in a cohort of breast cancer patients of the Milan province (2007-2012). **Methods** Exclusion criteria: no surgery, stage IV and having had a previous tumor. Hospital oncologic breast surgical volume was dichotomized in low (<135 interventions/year) and high (=135). Marital, educational, and occupational status were derived linking administrative databases and combined into one ten-category variable. Because 34% of cases had at least one missing SS variable, multivariate imputation was performed (MICE, m=30). A directed acyclic graph was used to analyze confounding. Through standardization, we estimated the relative risk (RR) of being operated in a LVH for each SS category compared to Married-High School or higher-Employed. 95%CI were obtained with bootstrap (n=100). **Results** After applying exclusion criteria, n= 6,333 were analyzed. 40% of women were treated in a LVH. The lowest proportion of patients treated in a LVH was in the Married-High School or higher-Employed category (25.4%), the highest in the Unmarried-Middle school or lower-Retired one (58.4%). All 3 categories including High School or higher degree had a non-significantly different risk of being treated in a LVH compared to the reference. All 6 categories with Middle School or lower degree had a RR significantly greater than 1. Within unmarried women with lower education, retirement increased the risk compared to unemployment (RR=2.0, 95%CI=1.72-2.33 vs. RR=1.4, 95%CI=1.1-1.8). **Conclusions** Receiving breast surgery at a LVH was related to lower education level and, in less educated women, to occupational status.

Estimating causal effects in complex systems

KF Arnold^{1,2}; WJ Harrison^{1,2}; MS Gilthorpe^{1,2}; A Heppenstalla³

¹Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

²School of Medicine University of Leeds, Leeds, UK

³School of Geography, University of Leeds, Leeds, UK

Population-level health patterns emerge from a complex, dynamic, multi-layered system, characterised by individual heterogeneity and autonomy, interdependence, adaptivity and evolution, feedback loops, and threshold effects. Altering such patterns (ideally beneficially) requires knowledge about the causal relationships that are important, and at which spatial and temporal scales they operate. Directed acyclic graphs (DAGs) have facilitated a revolution in data analysis, providing a framework for estimating causal effects as counterfactual contrasts. Nevertheless, there exists growing recognition that there are many situations for which DAG-based analyses are not well-suited, including evaluating causal effects in complex systems. This recognition has led to calls for supplementary approaches to estimating causal effects, one being individual-based simulation modelling (which includes both microsimulation modelling and agent-based modelling),

as it may also be used to evaluate counterfactual contrasts. Nevertheless, uptake has been slow due to limited cross-over in the knowledge about or skills necessary for implementing them, in addition to more fundamental concerns regarding the robustness and potential transferability of any causal conclusions beyond the simulations themselves. There exist significant methodological and philosophical differences between DAG-based statistical methods and individual-based simulation methods; these include their relative reliance on theory versus data, the timescales on which they are run, their relative focus on fixed versus random effects, and the ways in which they assess the effects of interventions. These differences provide scope for using the methods synergistically, though additional methodological work is required before such potential can be fully realised.

Integrating experts knowledge constraint in causal graph structure learning: extension to repeated measurement of immunological markers

Vahã Asvatourian^{1,2}, Philippe Leray³, Stefan Michiels^{1,2}, Emilie Lanoy^{1,2}

¹Universitã Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France

²Service de Biostatistique et dãEpidãmiologie, Gustave-Roussy, Villejuif, France

³Knowledge and Decision Team, Laboratoire dãInformatique de Nantes Atlantique (LINA), Nantes, France

In the context of longitudinal immunologic markers measured under anticancer immunotherapy, we have proposed an extension of the causal graph learning PC-algorithm to estimate the completed partially DAG (CPDAG) for chronologic ordered measurements (1). In biological settings, integrating a priori expert knowledge could help to improve the structure learning algorithm. However, the methods including this a priori expert knowledge have been limited to the case of time-fixed covariates. We then aim to model and use the a priori expert knowledge in a structure learning method for time-varying covariates. The experts opinions for each pair of variables have been resumed as a set of probabilities of presence and absence of a given arrow; then we converted it into a set of constraints used during the learning process. We run a bench of simulations to compare algorithms using a priori knowledge and those who do not. We simulated several levels of experts knowledge of the true DAG such as weak, medium or perfect and an a priori input of 5% and 40%. The performance of each final graph derived from simulated data was evaluated in terms of capacity of recovering the true edges (sensitivity) and the true absences of edges (specificity). The structural hamming distance (SHD) (2) was also estimated. Best sensitivities were observed when integrating 5% of 40% of a priori input knowledge from medium or perfect experts. Lowest SHD was observed when using at least 5% a priori perfect expert knowledge.

References

1. Asvatourian V, Coutzac C, Chaput N, Robert C, Michiels S, Lanoy E. Estimating causal effects of repeated multi-dimensional biomarkers in an observational setting. under Review;
2. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach Learn.* 2006;65(1):3178.

Stable ipw estimation for longitudinal studies

Vahe Avagyan¹, Stijn Vansteelandt^{1,2}

¹ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

² Department of Medical Statistics, London School of Hygiene and Tropical Medicine, UK

In this talk, we focus on estimating the parameters of the marginal structural model (MSM) (Robins, 1997). We consider the estimation of the average exposure effect using Inverse Probability Weighting (IPW) in a longitudinal observational study design, assuming that there is no unmeasured confounding of the exposure - outcome association at each time point, given the history of exposures and confounders. Despite its popularity in causal inference and incomplete data methodology, in practice the performance of IPW estimation may degrade due to instability of the estimated weights. We propose an estimating equation-based strategy for the nuisance parameters indexing the propensity scores at each time point, aimed at preventing highly volatile weights and ensuring the stability of IPW estimation. Extensive simulation studies demonstrate a desirable performance of the proposed approach.

Assessing causal effects of retirement on mental health across european countries

Marina Ballerini, Laura Forastiere, Alessandra Mattei
University of Florence

According to previous studies, quitting work could be harmful or beneficial for health. The literature reported mixed results positive or negative on both objective and subjective health measures, cognitive functions and mortality. However, most of the early studies did not consider health status before retirement, thus they can only infer correlation, not causality. Our goal is to assess the causal impact that the transition from work to retirement could produce on individual mental health in various European countries. We focus on three measures of mental health: self-rated health, mental health, depression. We used data coming from SHARE, the Survey of Health, Ageing and Retirement in Europe, a longitudinal ongoing study, started in 2004. SHARE is an excellent source of information regarding various key areas of life: work, health, socio-economic status and family networks. We used the potential outcome approach to causal inference under the assumption of Selection on Observables. We applied a propensity score matching to estimate causal effects of retirement. Results suggest that, considering all countries together, retirement seems to improve self-perceived health and depression but worsens mental health. We also conducted the analyses separately for groups of countries, characterized by homogeneous welfare system. Finally we made a longitudinal analysis since the effect of retirement on mental health could be cumulative and depend also on the years into retirement. We used Marginal Structural Models estimated by the Inverse Probability of Treatment Weighting under the assumption of Strong Ignorability. For each additional year passed into retirement mental health worsens.

An extension of the causal tree framework in the presence of irregular assignment mechanisms

Falco J. Bargagli Stoffi, Giorgio Gnecco
IMT School for Advanced Studies Lucca

This paper provides a linkage between machine learning techniques and causal inference in observational studies where the reception of the treatment is not randomized but the assignment to the treatment can be assumed to be randomized (irregular assignment mechanism). Specifically, the Classification and Regression Trees (CART) algorithm (Breiman et al., 1984) is reworked, following the seminal approach by Athey and Imbens (Athey and Imbens, 2015), to fit inferential goals within the Rubin’s potential outcome framework for causality (Rubin, 1974, 1978, 1980, 1989, 1990). The main strength of this algorithm and its variations applied to causal inference, as compared to other methods, is that they can deal efficiently with the heterogeneity of causal effects. This paper contributes to the growing applied machine learning literature by proposing two modified versions of the Causal Trees (Athey and Imbens, 2015) to draw inference from irregular assignment mechanisms: Causal Trees with Instrumental Variable and Tailored Trees. These methods are developed by merging the Causal Tree approach with the Instrumental Variables approach to causal inference (Angrist et al., 1996). The methods reviewed and proposed in the paper are then used to evaluate a public policy implemented by the Tuscan Regional Administration during the biennium 2003/2005 which aimed at easing the access to credit for small firms. While focusing on the heterogeneity of the treatment effects for the different firms, results are compared to those obtained through propensity score-based matching methods (Mattei and Mauro, 2007) as a robustness check on the overall treatment effect.

Assessing the impact of non-randomised drop-out on treatment switching adjustment in the relapsing-remitting multiple sclerosis clarity trial and the clarity extension study

Helen Bell Gorrod¹, Nicholas R. Latimer¹, Doris Damian², Gerard Harty², Robert Hettle³, Schiffon L. Wong²

¹ School of Health and Related Research (ScHARR), University of Sheffield, UK

²Merck EMD Serono,

³ AstraZeneca

Objectives: Treatment switching adjustment methods are commonly applied to oncology trials, where control arm patients switch onto the intervention treatment following disease progression. Alternative switching mechanisms or trial designs may require different adjustment methods. We applied a recognised method and a novel method to adjust for treatment switching in the context of the CLARITY and CLARITY Extension studies to compare low-dose cladribine tablets with placebo in relapse-remitting multiple sclerosis. **Methods:** We applied the rank preserving structural failure time model (RPSFTM) which relies on common treatment effect (CTE) and randomisation assumptions. To test the sensitivity of the results, we applied a propensity score matching (PSM) method combined with inverse probability of censoring weights (IPCW). These methods

do not rely on CTE and randomisation assumptions, but require measurement of all relevant confounders. Analysis end points were first qualifying relapse (FQR), 3-month, and 6-month confirmed disability progression (3mCDP, 6mCDP). Results: The RPSFTM resulted in hazard ratios (HRs) of 0.48 (95% CI 0.46-0.84) for 3mCDP, and 0.62 (95% CI 0.44-0.88) for 6mCDP. The PSM+IPCW resulted in HRs of 0.47 (95% CI 0.38-0.63) for FQR, 0.61 (95% CI 0.43-0.86) for 3mCDP and 0.63 (95% CI 0.40-0.87) for 6mCDP. Conclusions: The PSM+IPCW HRs were consistent with HRs from the RPSFTM, indicating that the results were not substantially biased by informative drop-out, assuming that all relevant confounders were controlled for in the PSM+IPCW model. The results show no statistical evidence of a waning cladribine treatment effect during the Extension study.

Mathematical coupling of proportions with common denominators: revisiting pearson, neyman, and fisher with causal graphs

Laurie Berrie^{1,2}, Peter WG Tennant¹, Paul D Norman³, Paul D Baxter^{1,2}, Mark S Gilthorpe^{1,2}

¹Leeds Institute for Data Analytics, University of Leeds

²School of Medicine, University of Leeds

³School of Geography, University of Leeds

Mathematical coupling of pairs of proportions that share a common denominator has long been acknowledged as creating artefactual bias in correlation and regression analyses. Independently, Pearson, Neyman and Fisher each proposed conditioning on the common denominator by including it as a separate variable in a regression model; however, potential issues have not been discussed within a causal framework. Using Directed Acyclic Graphs (DAGs) the relationships between proportions with a common denominator are illustrated under three possible scenarios: (i) the common denominator causes both proportion numerators and there is no causal relationship between the numerators; (ii) the common denominator causes both proportion numerators while one proportion numerator causes the other; and (iii) one proportion numerator causes the common denominator and both cause the other numerator. Artefactual associations that result from the mathematical coupling of proportions are highlighted and we confirm that adjusting for the common denominator as though it is a confounder of the exposure-outcome relationship avoids these artefactual associations. We further illustrate that adjusting for the common denominator fails to deliver meaningful causal inference if the common denominator lies on the causal path from one numerator to the other numerators. We provide real-world examples of each of these scenarios to demonstrate how causal diagrams may be used to understand a problem first identified over a century ago and illustrate how historical solutions to this problem work, but also show under which circumstances they are invalid. We discuss the implications for future research.

Education and life-expectancy and how it is mediated: evidence from the uk using principal strata

Govert E. Bijwaard and Andrew M. Jones
NIDI, University of York

We investigate the educational gain in life-expectancy using data from the UK Health and Lifestyles Survey and how it is mediated through health behaviour (smoking, doing exercise). The association between education and mortality may be confounded by factors that influence both education choice and health. For causal identification of the educational gain we employ a Regression Discontinuity Design implied by the increase in the UK minimum school leaving age in 1947 (from 14 to 15) together with a principal stratification method of the mortality hazard rate. The principal stratification defines complier types for both the educational attainment, depending on the schooling reform, and the level of the mediators, depending on the education level. This method allows us to derive the direct and indirect (through one or more mediators) effect of education on the implied life-expectancy. The educational gain based upon basic maximum likelihood estimation of a standard Gompertz hazard model for the mortality rate suggests that staying at school for more than 15 years significantly increases life-expectancy with 13 years, with a direct effect of education of 4 years and an indirect effect of health behaviour of 8 years. In contrast, estimates from the principal strata method indicate that an (significant) educational gain does not exist for those who remain at school longer. The direct effect of education is even negative (but insignificant). We do find a positive indirect effect of education running through health behaviour changes.

Matching, differencing on repeat propensity score matching and difference-in-differences with repeated crosssectional data: methodological guidance and an empirical application in education

Michele Binci, Madhumitha Hebbar, Paul Jasper and Georgina Rawle
Oxford Policy Management

We present an innovative approach to combining Propensity Score Matching (PSM) and Difference-in-Differences (DID) when only cross-sections of data are available. We illustrate the methodology in the impact evaluation led by Oxford Policy Management (OPM) of EQUIP-T, an education intervention in Tanzania funded by the UK Department for International Development (DFID). EQUIP-T is a four-year programme focused on improving teacher performance, increasing the quality of primary education and improving pupil learning outcomes. This study is likely to represent the first practical application of a PSM with DID procedure for a repeated cross-section in an education evaluation. We will review the implementation of the methodology in the context of the EQUIP-T programme and will discuss strengths, appropriate contexts, and caveats to the approach. Two separate analytical strategies are presented: in the first strategy, the Average Treatment Effect on the Treated (ATT) is compared across time. In the second, PSM is used to match treatment units (pupils and teachers in EQUIP-T schools) over time to construct a pseudo panel from repeated cross-sections to estimate overall ATT. In the absence of panel data, the conventional PSM approach of matching individuals at

baseline and then calculating impact at endline is not possible. The innovative pseudo panel approach addresses this. Impact estimates on pupil tests are presented. The PSM-DID analysis finds strong evidence that EQUIP-T has reduced the proportion of pupils in the bottom performance band for Swahili in programme schools. These results remain strong and highly significant across both our PSM with DID strategies.

The Causal Impact of Anthropogenic Emissions on Global Warming

Valentina Angius¹ and Chiara Binelli²

¹University of Milan-Bicocca

²University of Milan-Bicocca, Center for Research and Social Progress, and European University Institute

This paper uses machine learning methods to estimate the causal impact of anthropogenic emissions on global warming. We find that anthropogenic emissions are responsible for a substantial increase in average global temperature of about 0.7 degrees Celsius in the last forty years. The results show the important contribution of machine learning to address casual research questions when using a randomized control trial is not viable since the treatment affects all subjects simultaneously (Varian 2016).

Key Words: Global Warming; Machine Learning; Causality.

Randomization-based Inference for Bernoulli Trial Experiments and Implications for Observational Studies

Zach Branson and Marie-Abele Bind

Harvard University

We present a randomization-based inferential framework for experiments characterized by a strongly ignorable assignment mechanism where units have independent probabilities of receiving treatment. Previous works on randomization tests often assume these probabilities are equal within blocks of units. We consider the general case where they differ across units and show how to perform randomization tests and obtain point estimates and confidence intervals. Furthermore, we develop rejection-sampling and importance-sampling approaches for conducting randomization-based inference conditional on any statistic of interest, such as the number of treated units or forms of covariate balance. We establish that our randomization tests are valid tests, and through simulation we demonstrate how the rejection-sampling and importance-sampling approaches can yield powerful randomization tests and thus precise inference. Our work also has implications for observational studies, which commonly assume a strongly ignorable assignment mechanism. Most methodologies for observational studies make additional modeling or asymptotic assumptions, while our framework only assumes the strongly ignorable assignment mechanism, and thus can be considered a minimal-assumption approach.

On the Support of Matching Algorithms

Massimo Cannas¹ and Gavino Puggioni²

¹University of Cagliari

²University of Rhode Island, Kingston, USA

In causal inference a matching algorithm can be used to assign a subset of control units to each treated unit. Using combinatorial techniques we explore the support of matching algorithms to provide counting results, also investigating how support size is affected by the number of covariates.

The Effect of Unilateral Divorce on Prostitution: Evidence from Divorce Laws in U.S. States

Riccardo Ciacci

European University Institute

Do more flexible divorce regimes affect prostitution? Differences in the timing of entry into force of unilateral divorce laws across U.S. states provide a quasi-experimental setting allowing to estimate the effect of these laws on female prostitution (proxied by female prostitutes arrests). Using a difference-in-difference estimation approach, it is found that unilateral divorce reduces prostitution by about 11%, which amounts to a decline of 6,540 arrests per year. This result remains robust to changes in the specification of the estimated regression model. Combining various datasets, I explore several mechanisms that could rationalize this negative relationship. These include a general decrease in arrests of all sorts of crimes, and demand and supply-driven channels suggested by theoretical models of prostitution. The mechanism that fits best the evidence is one where unilateral divorce improves the option value of getting married by increasing wives bargaining power. As a result, the opportunity cost of becoming a female prostitute increases, and the supply of prostitution declines.

How to give support to fundamental conditions in a cause of effect problem?

Fabio Corradi¹ and Monica Musio²

¹Dipartimento di Statistica, Informatica, Applicazioni, Università di Firenze

¹Dipartimento di Matematica e Informatica, Università di Cagliari

In causal inference a problem of interest is how to use statistical evidence in a Cause of Effect problem (CoE). The question is: how might one use epidemiological data to assign a Probability of causation (PC) in an individual case? A simplified CoE question is: Ann had a headache and decided to take aspirin. Her headache went away. Was that caused by the aspirin? To deal with the problem often potential outcomes are introduced and the outcome is described assuming Ann receives the treatment, having decided to get it, otherwise assuming she was not exposed. (Dawid et al. (2016)) detailed three conditions that must be satisfied to estimate bounds for PC. 1. exogeneity: the potential

outcomes have the same joint distribution, among both treated and untreated study subjects. 2. sufficiency: Anns intention to take or not to take the treatment does not affect the distribution of the potential responses. 3. exchangeability: Anns potential responses are exchangeable with those of the treated trial subjects. The question we address is $\hat{\text{A}}\hat{\text{S}}$ What is the subgroup best predicting the Ann headache disappearance if there is exchangeability between the group and Ann ? At the same time for the untreated group what is the best group which makes the response independent of the Anns choice to assume the treatment? Can these two goals be achieved simultaneously? We set the question up as a MCMC Bayesian model selection problem, where each model specifies a particular choice for the background information. The models will be evaluated considering the marginal likelihood and a EBDIC prior (Chen (2008)).

References

- Chen, J. and Chen, Z., Extended Bayesian Information Criteria for Model Selection with large Model Spaces. *Biometrika*, 95, 3, 759-771, 2008.
- Dawid, A. P., Musio, M. and Fienberg, S. E., From Statistical Evidence to Evidence of Causality. *Bayesian Analysis*, 11, 725-752, 2016.

Maternal Stress and Pregnancy Outcomes Evidence from the 2004 Madrid Bombing

Marco Cozzani, Moris Triventi and Fabrizio Bernardi

The period a child spends in the womb is a crucial developmental stage, which influences children's outcomes both at birth and later in life. Although recent research has extensively investigated this developmental stage, we contribute by exploring the effect of prenatal maternal stress on pregnancy outcomes (low birth weight and prematurity). Most importantly, we examine its heterogeneity across different socioeconomic status (SES). We study the 2004 Madrid Bombing as an exogenous source of prenatal maternal stress with a difference-in-difference approach. The analyses are carried out on the population of singleton births from January 2000 to December 2005 in Spain, which includes 2,495,011 children. We find that the exposure to the shock increases the probability of low birth weight deliveries, but we find inconclusive results for prematurity. More interestingly, we find that this effect is concentrated among low-SES offspring. We hypothesized two possible mechanisms to explain these findings, which have broader implication for the transmission of status across generations.

Decision theoretic foundations for statistical causality

A. P. Dawid

University of Cambridge

Statistical Decision Theory aims to help me choose between available treatment interventions (causes), with the aim of bringing about a desired outcome (effect). To this end I need to assess the distribution, P_x , of the outcome variable, Y , for each treatment x (a value of the treatment variable X) that might be applied. These assessments might be improved when I have access to external data on X and Y (and other relevant variables); but when those data are observational their appropriate use may not be straightforward. The Decision-Theoretic Approach to Statistical Causality addresses this difficulty by identifying conditions and methods by means of which the data can be used to help me solve my decision problem. In this work we set out the basic framework and assumptions underlying this approach. Specific features include: making a clear distinction between stochastic variables and decision variables; appropriate assumptions of exchangeability; and consideration of the hypothetical distribution of the response when applying a treatment different from that intended. This last distributional reworking of the idea of a potential response variable allows us to usefully reinterpret and apply such concepts as ignorable treatment assignment and stable unit-treatment value assumption, all without invoking counterfactuals, which are not only metaphysically problematic but pragmatically an unnecessary complication. We further develop connexions with graphical representations, including influence diagrams, causal DAGs, and SWIGs.

Impact of spasticity on functioning in spinal cord injury: an application of graphical modelling

Cristina Ehrmann^{1,2}, Seyed Mahdi Mahmoudi³, Kerstin Hug⁴, Birgit Proding^{1,2},
Charlotte Kiekens⁵, Per Ertzgaard⁶

¹Swiss Paraplegic Research (SPF), Nottwil, Switzerland

² Department of Health Sciences and Health Policy, University of Lucerne and SPF, Nottwil, Switzerland

³ Faculty of Mathematics, Statistics and Computer Science, Semnan University, Semnan, Iran

⁴ REHAB Basel, Basel, Switzerland

⁵ Department of Physical and Rehabilitation Medicine, University Hospitals Leuven, Belgium

⁶ Department of Rehabilitation Medicine and Department of Medicine and Health Sciences, Linköping, Linköping University, Sweden

Background: People living with spinal cord injury (SCI) experience spasticity, a complication that interferes with their overall functioning. Using graphical modelling approach, the objective of this study was to describe the impact of moderate to severe spasticity on functioning by 1) identifying and visualizing the functioning domains associated with spasticity and ii) identifying the potential intervention targets for improving spasticity. Methods: Cross-sectional data from the Swiss Spinal Cord Injury Cohort Study were used. It included individuals (N=1436) with a permanent residence in Switzerland, 16 years or older and with a history of traumatic or non-traumatic SCI. Spasticity and other forty-seven categories of functioning were operationalized using reliable and valid instruments (SCI Secondary Conditions Scale, the SF-36, the SCI Independence

Measure Self-Report and the Utrecht Scale for Evaluation Rehabilitation-Participation). The missing data were imputed using random forest imputation technique. The PC algorithm was employed to identify and visualize the functioning categories associated with spasticity. Pearl's do-calculus of intervention implemented by Mahmoudi et al. was employed to identify the potential intervention targets for improving spasticity. Results: 51% reported moderate to complete problems in spasticity. Chronic pain, contractures, sexual dysfunctions, feeling tired, doing housework, and respiratory problems were associated with spasticity and in the top ten potential intervention targets for improving the experience of Spasticity. Contractures and chronic pain were associated with spasticity in both tetraplegic and paraplegic SCI sub-samples. Conclusions: Combining graphical models with other statistical concepts in this way should improve our knowledge about spasticity in people living with SCI.

The Effect of Unemployment on the Smoking Behavior in Couples

Jakob Everding and Jan Marcus

Hamburg Center for Health Economics, Universität Hamburg, Hamburg, Germany

Involuntary job loss with subsequent unemployment is associated with various negative effects. While unemployment results in long-lasting earnings losses, for instance, this is likely to impact the spouses of the unemployed as well. Moreover, unemployment can lead to social isolation, lower mental health, and changed health behaviors—all of which might influence the spouses of the unemployed, too. However, there is surprisingly little research examining spousal spillover effects of unemployment. We contribute to filling this gap by studying whether unemployment affects also the smoking behavior of spouses.

Our empirical strategy exploits plausibly exogenous entries into unemployment using data from the German Socio-Economic Panel (SOEP). To identify causal effects, we use a combined matching and difference-in-differences approach and employ the post-double-selection method, a supervised machine learning tool for the selection of control variables. We analyze the effects of unemployment on the smoking status and the smoking intensity of directly and indirectly affected spouses.

We find that the unemployment of one spouse significantly increases the smoking probability and intensity of both spouses to a similar extent. In specific, both spouses are approximately 4 percentage points more likely to smoke and increase their daily number of cigarettes smoked by 10 percent, on average. Treatment effect heterogeneity analyses reveal that the effects are stronger if the male enters unemployment and if the spouses are smokers.

The results highlight that among couples, both spouses respond with significant health-threatening behavior to unemployment. The findings are of high policy relevance as they emphasize the importance of taking spillover effects into consideration when estimating the social cost of unemployment.

A systematic map of within-study comparisons: examining bias in non-randomised studies

Paul FENTON VILLAR¹ and Hugh WADDINGTON²

¹ University of East Anglia (UEA),

² International Initiative for Impact Evaluation (3ie)

Can non-randomised estimates of the causal effects of social and economic programs approximate estimates from well-designed and well-executed randomised experiments? What factors determine bias amongst non-randomised estimators? How accurately do we predict bias arising from non-randomised studies? These are all examples of questions that within-study comparisons may help us to examine. A within-study comparison compares a non-randomised comparison group (or estimator) with an unbiased causal benchmark control group (or estimator). They have been conducted in the social sciences since the 1980s, following the internal replication of a randomised evaluation of the National Supported Work (NSW) Demonstration programme in the USA (Lalonde, 1986). This poster presentation synthesises the findings of a systematic map funded by the Campbell Collaboration, the American Institutes for Research (AIR), and the International Initiative for Impact Evaluation (3ie) on within-study comparisons. The map summarises the largest known collection of within-study comparisons, describing the characteristics of more than 95 studies on programs concerning education, labour markets and employment, social welfare, political science, and the environment. The findings provide direction for further developing this evidence base through primary and replication studies and will lead to a systematic appraisal of studies assessing biases in non-randomised estimates.

Propensity score diagnostics: the challenges we face when providing evidence that a propensity based estimate is unbiased

Emily Granger¹, Dr. Jamie C Sergeant^{1,2}, 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

Recent years have seen a marked increase in the use of propensity scores to balance covariate distributions between exposure groups in observational data. However, there is scepticism about results obtained using propensity scores¹. One concern is that an inadequately estimated propensity score may not effectively balance covariate distributions, which could lead to biased results. There are a variety of diagnostics being used to assess covariate balance after propensity-adjustment, however no consensus on the best way to do this^{2,3}. A simulation study was conducted to compare diagnostics in terms of their ability to identify different types of propensity score misspecification. Diagnostics included are categorised as follows: 1) mean-based, 2) distribution-based and 3) prevalence-based. Categories 1 and 2 respectively include diagnostics which compare covariate means and distributions between treatment groups. Category 3 diagnostics have only recently been proposed. They involve comparing the number of exposed subjects at

each covariate value to that predicted by the propensity score. Results indicated that mean-based diagnostics can fail to identify when a nonlinear term is incorrectly omitted from the propensity score model and distribution-based diagnostics are unreliable at identifying omission of interaction terms. However, the prevalence-based diagnostics performed well at identifying all types of model misspecification. To help overcome scepticism about propensity scores, being able to reliably assess covariate balance obtained after propensity-adjustment is essential. Unfortunately, our results demonstrate that some of the most widely used (category 1) or readily available (categories 1 and 2) diagnostics can be misleading. Further development of prevalencebased diagnostics is recommended.

1. Nuttall GA, Houle TT. Liars, Damn Liars, and Propensity Scores. *Anesthesiology* 2008; 108: 34.
2. Ali MS, Groenwold RHH, Belitser S V., et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: A systematic review. *J Clin Epidemiol* 2015; 68: 112121.
3. Yao XI, Wang X, Speicher PJ, et al. Reporting and Guidelines in Propensity Score Analysis: A Systematic Review of Cancer and Cancer Surgical Studies. *JNCI J Natl Cancer Inst* 2017; 109: 19.

Causal inference models comparing the effectiveness of clinical procedures in babies born with complex cardiac abnormalities

Wendy J Harrison¹, Ngoni K Zava², James R Bentham²

¹Leeds Institute for Data Analytics, University of Leeds, UK

²Yorkshire Heart Centre, Leeds General Infirmary, Leeds, UK

While the assessment of procedural effects is usually the domain of clinical trials, observational data can offer a rich resource, although with methods challenges due to data complexity. Research questions relating to causal factors (upon which one might intervene) necessitate a causal inference perspective, not a predictive modelling approach. We demonstrate the utility of causal inference modelling on data from a UK multi-centre cohort study to assess cardiac surgical and intervention outcomes (Bentham et al., 2017). A directed acyclic graph (DAG) was constructed to depict assumed causal relationships amongst covariates, assessed for consistency with the dataset using the R package dagitty (Textor et al., 2016). Covariates were identified as confounders, mediators or competing exposures; a propensity score was calculated including all confounding variables together with the fixed effect of cardiac centre. Initial analyses consider the propensity score in single-level logistic regression models (primary and secondary outcomes) and Cox proportional hazards models (longer-term survival outcomes). Further analyses consider multivariable models that explicitly adjust for each confounder using an appropriate minimally sufficient adjustment set (MSAS) identified from the DAG. Findings from both approaches were contrasted with and without the consideration of additionally accommodating competing exposures. Results were consistent for primary and secondary outcomes, and the inclusion of competing exposures in calculating propensity scores showed little improvement in the precision of estimates of effect. The survival analysis of re-intervention with the MSAS differs marginally from that

using the propensity score. These findings offer confidence in demonstrating tangible advantages for a novel clinical procedure.

References:

Bentham, J.R., Zava, N.K., Harrison, W.J., Shauq, A., Kalantre, A., Derrick, G., Chen, R.H., Dhillon, R., Taliotis, D., Kang, S-L., Crossland, D., Adesokan, A., Hermuzi, A, Kudumula, V., Yong, S., Noonan, P., Hayes, N., Stumper, O., Thomson, J.D.R. 2017. Duct stenting versus modified Blalock Taussig shunt in neonates with duct-dependent pulmonary blood flow. Associations with clinical outcomes in a multicenter national study. *Circulation*. 2018;137:581-588. <https://doi.org/10.1161/CIRCULATIONAHA.117.028972>
Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M and Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*. 2016;45:1887-1894.

Causal inference from treatment-control studies having a pseudo-factor with unknown assignment mechanism

Nicole Pashley and Kristen Hunter (presenters) with Katy McKeough and Tirthankar Dasgupta

Department of Statistics, Harvard University

In standard clinical trials a treatment factor of primary interest is randomly assigned to experimental units. However, there are situations in which a treatment factor of secondary interest is applied simultaneously, but its assignment mechanism is either not known to or not controlled by the experimenter. Such a secondary treatment factor can be considered a pseudo-experimental factor, and the set-up of the experiment can be regarded as a semiobservational factorial design. Motivated by a complex real-life clinical trial, we study this problem using the potential outcomes framework for factorial experiments. We discuss possible estimators for the main effect of the primary treatment factor, and theoretically examine how the assignment to secondary treatment impacts the bias and variance of these estimators. The assignment mechanism of the secondary factor is assumed to be probabilistic but unknown. We consider two cases: one in which the analyst does not even observe the assignment of the secondary factor, and another in which the assignment is observed, although the underlying mechanism is unknown. We show how such assignment information in the latter case can be incorporated into the analysis to improve inference, and quantify the gains through theoretical derivations and simulations.

Estimating treatment effects from observational data using instrumental variables

John ER, Sheehan NA, Abrams KR, Brightling CE

University of Leicester

Instrumental variable (IV) approaches have been advocated to estimate causal effects in the presence of unmeasured confounding. IV methods are commonly used in settings when randomisation breaks down and typically deliver an estimate of a local causal effect e.g. effect of treatment on those who comply with their assigned treatment regime. The increasing availability of large cohort data, such as CPRD (Clinical Practice Research Datalink), has heightened interest in using such data for treatment effect estimation. In this setting we are typically more interested in estimating a population causal parameter such as the average treatment effect across all patients. One of the issues with conducting an IV analysis is choosing an appropriate IV. In the Mendelian randomisation setting, the availability of genetic IVs with known (or partially understood) biological function enables successful targeting of population causal parameters in certain settings. However, non-genetic IVs are harder to find and to justify. Commonly used IVs include physicians prescribing preference, distance to treatment facilities and regional variation in treatment rates. While for some particular effects they may be reasonable, it is likely that they are directly linked to the outcome. Moreover, issues of generalisation of the findings are analogous to those of trials, especially when there is heterogeneity in treatment effects. Using a small cohort of COPD patients I have explored the issues that arise for IV analyses of treatment effects. A simulation study was subsequently performed, based on the COPD data, to obtain broader results about the effect of violating IV assumptions.

Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach

Michael C. Knaus¹, Michael Lechner^{1,2} and Anthony Strittmatter

¹IZA, Bonn

²CEPR, London, CESifo, Munich, IAB, Nuremberg

We systematically investigate the effect heterogeneity of job search programmes for unemployed workers. To investigate possibly heterogeneous employment effects, we combine nonexperimental causal empirical models with Lasso-type estimators. The empirical analyses are based on rich administrative data from Swiss social security records. We find considerable heterogeneities only during the first six months after the start of training. Consistent with previous results in the literature, unemployed persons with fewer employment opportunities profit more from participating in these programmes. Finally, we show the potential of easy-to-implement programme participation rules for improving average employment effects of these active labour market programmes. Financial support from the Swiss National Science Foundation (SNSF) is gratefully acknowledged. The study is part of the project Causal Analysis with Big Data which has grant number SNSF 407540_166999 and is included in the Swiss National Research Programme Big Data (NRP 75). We presented a previous version of this paper

at the Humboldt- University Berlin, Cal Poly in San Luis Obispo, University of Maas-tricht, Workshop on Unemployment and Labour Market Policies in Barcelona, the IZA Summer School in Labor Economics in Ammersee, Institute for Employment Research (IAB) in Nuremberg, and Computational Social Science Workshop in Konstanz. We thank participants, and in particular we thank Hugo Bodory, Bruno Crepon, Carlos Flores, Bernd Fitzenberger, Chris Hansen, Jeff Smith, and Martin Spindler, for helpful comments and suggestions. The usual disclaimer applies.

Decomposing zero: the (non-) effect of health insurance on infant mortality in indonesia, explained by mediation analysis

Noemi Kreif¹, Kala Diaz-Ordaz², Andrew Mirelman¹, Marc Suhrcke¹

¹ University of York, Centre for Health Economics

² London School of Hygiene and Tropical Medicine, Department of Medical Statistics

While health policy makers expect that publicly provided health insurance will improve health, rigorous empirical evidence of the causal impact is scarce. In an ongoing evaluation in Indonesia, we found no significant effect on infant mortality (IMR) for two major health insurance schemes: free health insurance for the poor and contributory insurance for formal sector employees. We use mediation analysis to understand this finding, hypothesising that health insurance should improve health outcomes via increased access to healthcare. We decompose the total effect of insurance into direct and indirect effects mediated by the birth being attended by healthcare professionals. We follow the literature on these topics to select the common causes of insurance enrolment and IMR, and those of the mediator and outcome. Given the large number of potential confounders in household survey data, we also screen for predictors of infant mortality using machine learning. Conditional on the selected variables, we fit a natural effects model allowing for treatment-mediator interaction using an imputation based approach. We find that for the free health insurance, the indirect effect through decreased access to assisted birth significantly increased IMR. In contrast, for the contributory insurance, we find a significant reduction in IMR through increased access to assisted birth. This is in line with previous suggestive evidence on limited (and low quality) health care supply available for those with subsidised health insurance. In conclusion, mediation analysis can help understand negative findings in policy evaluation studies, providing information on why a programme may (not) work as expected.

Comparison of causal inference methods for exposome research

Jessica E. Laine, Giovanni Fiorito, Paolo Vineis, and Exposomics Consortium
Imperial College London

The exposome, in its original form, was proposed to encompass the totality of exposures across the life course from conception onwards. This paradigm provides a framework within epidemiological research to identify new environmental risk factors, elucidate associations between multiple exposures and diseases, determine putative mechanisms that act as intermediates in pathways between exposures and diseases, and extrapolate subsequent causal effects of diseases for disease treatment and prevention. Measuring mediating mechanisms, in particular through omics (e.g. transcriptomics, epigenomics, metabolomics and proteomics) are a strength of the field of the exposome; however, there has been limited applications of causal inference within the exposome paradigm. One approach to address causality is the meet in the middle (MITM) approach employed in the Exposomics study. Through omics a mechanistic chain linking air pollution to coronary heart disease (CHD) risk was identified, where results (among non-smokers) identified inflammatory and oxidative pathways from DNA methylation markers that were associated with both air pollution measures and CHD. Following this finding, we compare this approach to counterfactual models to contrast methods of causal inference within Exposomics. We further discuss the applicability of the counterfactual model and other causal inference methods for exposome research.

A bayesian approach to surrogate endpoints assessment through principal stratification with application to a clinical trial for postoperative pain control

Andrea Lazzerini
Alma Mater Studiorum - Università di Bologna

In clinical research, the most significant endpoint regarding the effectiveness of a therapy is often not easy to measure. In these cases, we must rely on alternative or surrogate endpoints. Since 1989, many surrogate-evaluation methods, based on or revising Prentices definition (Prentice, 1989), have been designed to check if the surrogate and clinical endpoints are correlated. Prior to principal stratification (Frangakis and Rubin, 2002), all the approaches to surrogate assessment were based only on the association between the observed values of the outcome and the observed values of the surrogate, not providing a causal effect. Consequently Frangakis and Rubin (2002) offer a new criterion for surrogacy using principal stratification and principal causal effects. Gilbert and Hudgens (2008), for the purpose of quantifying these effects, introduced the concept of causal effect predictiveness (CEP) surface. Unlike using the first developed sequential estimation methods for the CEP surface, which provide bias, we extend previously Bayesian approaches that model simultaneously the imputation stage and the outcome model stage (Belin and Zigler, 2012). In our research, we analyze data on a double-blind randomized controlled trial, designed to investigate the effects of preoperative oral administration of morphine sulphate on patients postoperative pain (Baccini et al., 2015; Borracci et al., 2013). We propose to extend the Belin and Ziglens approach, when both the surrogate

and the primary outcome are continuous, to evaluate the pain intensity at 4 hours as a surrogate endpoint for pain intensity at 24 hours.

Applying causal mediation analyses to randomised controlled trials in healthcare. two case examples

Hopin Lee, PhD^{1,2,3}, Sallie E Lamb, DPhil¹, Robert D Herbert, PhD^{3,4}, James H McAuley, PhD^{3,5}

¹Centre for Statistics in Medicine, Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

²School of Medicine and Public Health, University of Newcastle, NSW, Australia

³Neuroscience Research Australia, Sydney, NSW, Australia

⁴School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia

⁵ Prince of Wales Clinical School, University of New South Wales, Sydney, NSW, Australia

Since the recent inception of patient-centred care, the medical profession has emphasised the importance of patient-reported outcomes such as function and quality of life. Accordingly, clinical trials in medicine often designate patient-reported outcomes as primary endpoints, while impairment level outcomes are considered secondary. In such trials, it is often assumed that intervening on a set of impairments should yield improvement in outcomes that are relevant and perceptible to everyday functioning. However, this assumption is often not evaluated in randomised controlled trials. The application of causal mediation analysis to randomised controlled trials presents an opportunity to extend beyond analyses of efficacy and to investigate the mechanisms of intervention. In this poster presentation, we will present the findings of causal mediation analyses applied to two randomised controlled trials in rehabilitation healthcare. We will demonstrate why a rehabilitation program for patients with ankle fracture (n=214) was not effective in improving patient function by estimating causal pathways through a targeted mechanism physical activity. Then we will demonstrate how an exercise program for people with rheumatoid arthritis (n=490) led to a positive effect on hand function by estimating causal pathways through the targeted mechanism hand strength. We will discuss that appropriate interpretation of these mechanistic analyses can provide clinical and policy implications that extend beyond those provided by standard efficacy analyses of randomised controlled trials.

Statistical assessment of strict population overlap in observational studies

Lihua Lei, Alexander D'Amour, Peng Ding, Avi Feller and Jasjeet Sekhon

Department of Statistics, UC Berkeley

In causal inference, strict population overlap, also known as positivity or common support, is a central assumption for identification and efficient estimation of causal effects. This assumption asserts that the propensity score is bounded away from 0 and 1. Strict overlap is often treated as axiomatic; however, in this paper, we show that the strict overlap assumption has empirical implications that can be observed in finite samples. In particular, we specify several estimators for the overlap slack, or the constant that bounds the propensity score away from 0 and 1. Our procedures provide upper confidence bounds, which can be interpreted as optimistic estimates of the overlap slack in the study population. We specify two exact, finite-sample confidence bounds: one based on the Dvoretzky-Kiefer-Wolfowitz inequality, and one based on an empirical Bernstein-type inequality. We also specify an approximate, asymptotically justified confidence bound based on the bootstrap. These confidence bounds require no conditions on the covariates, and can be made more powerful by incorporating modern machine learning or ensemble classification methods. Our estimators can be used as a diagnostic in observational studies, where a small estimated overlap slack would suggest that causal effect estimators will perform poorly. These estimators are particularly useful in studies with high-dimensional covariates, where poor overlap can be difficult to discern by the conventional practice of examining fitted values from a propensity score model. We demonstrate our methodology on simulated and real data.

Process evaluations of complex interventions using causal inference

Ruijie Li, Kimberley Goldsmith, Sabine Landau

Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London

Introduction

Complex interventions in healthcare are interventions that have multiple components and hence multiple plausible causal pathways on the outcome of interest. Increasingly, these are being evaluated as part of process evaluations but there is much ambiguity over what constitutes process evaluation and how it can be used to infer causality. My work will clarify the scope of process evaluation and provide examples of how methods from the causal inference field can inform quantitative process evaluations.

Methods

A systematic review was conducted to clarify the scope of process evaluation and its methods. Methodological gaps within the current practice of process evaluation are identified and potential solutions are proposed.

Results

The systematic review identified three definitions of process evaluation. These definitions consistently imply an element of causality, however, most studies identified primarily used qualitative methods and do not make use of methods of causal inference in

the analysis. Multiple sequential mediation was identified as one of the potential methods for use in process evaluation and is illustrated using data from a clinical trial of caregiver education for caregivers of people with eating disorders. Possible extensions to such a model and its advantages are also discussed.

Conclusion

Quantitative process evaluation has much potential to provide a better understanding of how interventions work provided studies are designed appropriately and causal estimands are targeted by principled methods.

Bayesian principal stratification analysis in Regression Discontinuity Designs with multiple outcomes: An application to students' academic performances

F. Licari and A. Mattei

University of Florence

In Italian Universities, Schools of Engineering require freshmen to take a pre-test and enrollment depends on the results of this pre-test. Specifically, students with a test score less or equal to a prefixed threshold must comply with additional educational requirements, which we refer to as OFA (Obblighi Formativi Aggiuntivi), to enroll at the school. The assignment rule underlying the OFA defines a Sharp Regression Discontinuity (SRD) design. Following Li, Mattei, Mealli (2015)¹, we adopt a probabilistic formulation of the assignment mechanism underlying RD designs and we select suitable subpopulations around the threshold on the basis of observed covariates using a Bayesian model-based approach.

Our primary aim is to evaluate causal effects of OFA on students' academic performances, defined by two variables: the number of exams passed at the end of the first year after the enrollment and the related grades. We do this accounting for the fact that after the test some students may decide not to enroll in the School of Engineering, so their performances are not defined: they are *truncated by death*.

In order to assess causal effects on students's performances, we use the principal stratification framework within the Rubin Causal Model. Specifically, we focus on assessing local causal effects on academic performances for the principal stratum of students who would always enroll irrespective of their OFA status considering the two academic performance measures. We use a Bayesian model-based approach to inference, modeling the joint distribution of the two academic performance measures. The presence of multivariate outcomes may also help improving inference.

¹Li, F., Mattei, A., Mealli, F.: Evaluating the effect of university grants on student dropout: Evidence from a regression discontinuity design using bayesian principal stratifications Analysis, *The Annals of Applied Statistics*. 9(4), 1906-1931 (2015).

Causal mediation analysis for randomised studies with longitudinal data using structural equation modelling

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

¹ Department of Data Analysis, Ghent University, Gent, Belgium

² Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium

In a randomised study with longitudinal data on the mediator and outcome, the direct effect of the treatment on the outcome at a particular time includes all pathways that avoid earlier instances of the mediator. Estimation of the direct effect thus requires adjusting for confounders between the outcome and earlier instances of the mediator. But when the set of confounders are themselves affected by treatment, standard regression adjustment is prone to possibly severe bias. Under a certain class of linear models, traditional path analysis methods provide unbiased estimates of the controlled direct effect, which are obtained by combining the estimated path coefficients for the constituent paths. We describe how the path analysis approach can be embedded within the structural equation modelling framework and propose extensions to settings with latent mediator and outcome. However, when time-varying exposure-confounder and mediator-confounder interactions are present, the path analysis approach can produce biased estimates. We propose a G-estimation approach that can incorporate nonlinearities and interactions in the presence of posttreatment confounding and yields unbiased estimates of the controlled direct effects. The G-estimation approach adapts existing methods for time-varying treatments and generalizes the proposed path analysis approach to settings with noncontinuous mediators and confounders. The approaches are illustrated using simulation studies and applied to a randomised study on the effects of a mobile health intervention on behaviour change, via attitude to health.

Performance assessment as an application of causal analysis

Nicholas T. Longford

Department of Medicine, Imperial College London

An example of case-mix adjustment in a national audit of neonatal care (broncho-pulmonary dysplasia in the UK National Neonatal Audit Programme) is presented. The solution is based on comparing neonatal units and their networks on the hypothetical (potential) outcomes of one or several synthetically constructed sets of infants, defined by their background variables. Such reference sets may be defined upfront and submitted for inspection or consultation by all the stakeholders, and can be retained for several years, supporting inter-year comparisons. A default choice for a reference set is a (national) random sample from the infants in the recent past, although samples drawn from certain subpopulations, or with stratification on them, can be used for assessing the performance (with a focus) on particular subgroups of infants. Failure to match some infants is a valuable diagnostic, indicating that the audited institutions have different distributions (profiles) of backgrounds, and are therefore not directly comparable. In this case, a league table is inappropriate. The advantages of this method over regression, which appeal to the clinicians involved, are the clear statement of the statistical task

(the counterfactual issue and the task of matching), transparency of the method and the potential, and welcome, engagement of all stakeholders in the definition of the reference set, motivated as a standard for the comparison(s).

The problem of patient localization in emergency care

Elena Lucchese

University of Bologna

The problem of patient localization is often mentioned on public documents and by anecdotal evidence. This paper quantifies for the first time the magnitude of the patient location problem in emergency care. To do so, I adopt a difference-in-differences identification strategy on non-deferrable ambulance missions in Liguria in 2013-2014. I find that the location problem causes a delay of about 5 minutes, one third of the average ambulance driving time to go. The effect is constant, regardless of the distance travelled by the ambulance. It drops to 3 minutes for non-urgent missions, probably because the caller communicates more clearly her location.

JEL classification: C26, I12, I18, R41

Keywords: location, driving time, response time, emergency, EMS, efficiency, performance

A dynamic definition of linear Markovian structural causal models

Alessandro Magrini

University of Florence

Linear regression with lagged covariates (distributed-lag linear regression) is a standard methodology for dynamic modelling in biometry and in econometrics, but it is limited to a single outcome of interest and cannot provide insights on the relationships holding between the covariates. Here, distributed-lag linear regression is applied to Markovian structural causal models to overcome these limitations. Several rules are provided to compute dynamic causal effects from regression coefficients at different time lags. The proposed methodology is illustrated through a real-world application.

The effects of a new tramway line on the commercial vitality of peripheral neighborhoods

Giulio Grossi¹, Patrizia Lattarulo¹, Marco Mariani¹, Alessandra Mattei^{1,2}, Ãzge Ãner^{1,3}

¹ IRPET Ã Regional Institute for Economic Planning of Tuscany

² University of Florence

³ University of Cambridge

The creation of urban rail transport infrastructure may have general implications on the economic landscape of the peripheral city areas that are served. Our motivating application focuses on the causal effects of a tramway line built in Florence (Italy) on a number of outcomes related to the local commercial vitality and on the economic performance of the area retailers. It considers a time horizon that encompasses both the construction and the later period of operation of the infrastructure. In theory, the streets where the tramway is built might suffer from harder accessibility and poor attractiveness during construction works, whereas alternative streets in the same neighborhood where traffic is diverted might benefit from the ongoing works. More ambiguously, when the infrastructure is ready, the new tramway line might: i) have beneficial effects on both the streets it serves and the neighboring ones due to improved area accessibility and attractiveness; ii) have beneficial effects on the streets it serves and harmful (crowd out) effects on alternative streets in the neighborhood; iii) crowd out shops in the neighborhood in favor of those located in distant areas that are made more accessible by the tramway. This kind of observational setting poses interesting challenges for causal inference, as it clearly requires to allow for possible interference between streets. To conduct our analysis at the street level, we adopt the potential outcomes framework and provide an original extension of the synthetic control group approach to a setting where between-street interference is possible. Exploiting information from multiple control streets that, because of their distance from the tramway, may, or may not, interfere with the streets under treatment, we disentangle the direct and the spillover-mediated effects of the tramway in a nonparametric fashion.

Misclassification of exposures in sufficient cause interaction analysis

Murthy N Mittinty

School of Public Health, the University of Adelaide, Australia.

For the case when the outcome, and exposures are binary VanderWeele¹ derived the required conditions which allow interpret statistical interactions as sufficient cause interactions, using the empirical data. However, when the exposures are misclassified the tests of interactions may not be valid unless one makes certain assumptions on the bounds of the probability of misclassification. When the exposures are subject to forms of independent differential misclassification, VanderWeele², showed that the probability of misclassification is no larger than 0.5 and 0.25, based on whether there is misclassification in the distribution of negatives or positives. In this paper, we extend these techniques to the case where one exposure is binary and other exposure is a categorical, and both the exposures are misclassified. Similar to the results of binary exposures the

probability of misclassification is no different to this case. In this considered case, we derive the bounds under the weak and definitive sufficient cause interactions. We also extend these techniques to ordinal variables where once again the bounds on the probability of misclassification are no larger than 0.5 and 0.25.

Reference

1 VanderWeele, T. J. Sufficient cause interactions and statistical interactions. *Epidemiology* 20, 6-13 (2009).

2 Vanderweele, T. J. Inference for additive interaction under exposure misclassification. *Biometrika* 99, 502-508 (2012).

Detecting plasma proteins causally related to ms: a mendelian randomization approach

Morani G., Fazia T., Bernardinelli L.

Department of Brain and Behavioral Science, University of Pavia, Pavia, Italy

The pathogenesis of Multiple Sclerosis (MS) is far from being completely understood. The identification of novel causal pathways could lead to development of new therapeutic strategies. We explored the existence of causal MS pathways among a subset of candidate plasma proteins, selected on the basis of heritability estimates and literature review, by analyzing 20 multiplex families from the founder and genetically homogeneous population of the Nuoro province, Sardinia (Italy). To overcome the limitations of observational studies we invoked the concept of Mendelian Randomization (MR), using genetic variants as instrumental variable to assess the possible causal effect of interest. We applied different MR methods: Inverse-Variance Weighted as the main causal analysis together with Weighted Median Estimator and Egger regression as two sensitivity analysis. On the statistically significant causal proteins, we performed a bidirectional MR analysis to investigate potential reverse causation. Seven out of 20 candidate proteins showed statistically significant causal association with MS. The most interesting is IL12A, causally associated with MS both in the main and sensitivity analysis. The use of Inverse-Variance Weighted method in combination with the two sensitivity analysis has helped to assess robustness of causal findings.

The causal effect of education on diseases in old age

Gianmaria Niccodemi, Govert E. Bijwaard

University of Groningen, Netherlands Interdisciplinary Demographic Institute

Several studies have found a positive association between education and health. It is commonly assumed that a large part of this association derives from the causal effect of education on health outcomes. Confounding factors that affect both education choices and health, such as intelligence and parental background, play an important role in shaping this association. We use a structural model to estimate the effect of education on the occurrence of diseases in old age, controlling for this endogeneity. The analysis is based on data of all men who were born in The Netherlands in 1944-1947 and examined at age 18 for national conscription; a sample of 35,000 individuals is linked to medication use over the period 2006-2014. We investigate different medications that are linked to diseases in old age. We estimate a four-part structural model, consisting of (i) an ordered probit model for the educational attainment (ii) a Gompertz mortality model for surviving up to old age (iii) a probit model for medication use in old age (iv) a measurement system using IQ-tests to identify the latent intelligence distribution. Educational attainment, surviving up to old age and medication use all depend on observed confounders and latent intelligence. From the estimation results we derive the causal impact of education on diseases in old age. We have found a strong effect of education on physical diseases (hyperlipidemia, diabetes, hypertension, cardiac diseases, COPD and asthma) but low or no effect on mental disorders (depression and anxiety).

Impact of bolsa familia program on new case detection of leprosy in brazil

Julia Moreira Pescarini¹, Laura Forastiere², Joilda Silva Nery³, Elizabeth B. Brickley⁴, Mauricio Lima Barreto¹, Laura C. Rodrigues⁴, Gerson Oliveira Penna⁵

¹ Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs), Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil.

² Department of Statistics, Computer Science, Applications, University of Florence, Italy.

³ Universidade Federal do Vale do São Francisco (UNIVASF), Paulo Afonso, Brazil.

⁴ Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom.

⁵ Centro de Medicina Tropical, Universidade de Brasília (UNB), Brasília, Brazil

Conditional cash transfer (CCT) programs have been implemented in more than 60 countries and provide cash payments to poor families on their meeting certain conditions such as preventive healthcare use or children's school enrollment. The impact of such programs on health, especially on infectious diseases, is still an open question. Here we focus on the impact of the Brazilian CCT - Bolsa Familia Program (BFP) on leprosy, a neglected infectious disease, concentrated among poor communities, with over 30,000 new cases detected yearly in Brazil. The evaluation of CCTs poses several challenges, given the specific design based on selection criteria and conditionalities. Given the assignment of BFP, based on a poverty indicator (per capita income), the evaluation of BFP reflect the typical regression discontinuity design. Moreover, several mechanisms

embedded in the application and roll out process lead to fuzzy designs, where eligibility does not coincide with cash transfers receipt. We analysed a random sample of 1.2 million newly registered families in Brazil social registry (CadUnico) from 2007 to 2014 (including baseline sociodemographic characteristics, per capita income and BFP receipt) linked with national data on leprosy in the same period (864 leprosy cases). We used a Fuzzy RD analysis to estimate the effect of a family start receiving BFP transfers within a given time after applying to CadUnico on future leprosy cases in that family. We explored the heterogeneity of the effect using two eligibility criteria for the program (poor and extremely poor families), as well as the effect of the program duration. My participation in the conference is supported by CONFAP/ESRC/MRC/BBSRC/FAP-DF 2015 Neglected Diseases. Project Effect of social determinants and impact of social programs on leprosy in the '100 million Brazilians' cohort – Process number 193.000.008/2016.

Goodbye Smokers Corner: Health Effects of School Smoking Bans

Gregor Pfeifer, Mirjam Reutter, Kristina Strohmaier

We study the impact of school smoking bans on individual health behavior in Germany. Using a multiple difference-in-differences approach combined with randomization inference, we find that the propensity towards smoking reduces by 14-21 percent, while the number of smoked cigarettes per day decreases by 7-25 percent. After elaborating on treatment effect heterogeneity and intensity, we evaluate spillovers to other health behavior of the treated individual and to smoking behavior of non-treated persons living in the same household.

Impact evaluation for small areas

Setareh Ranjbar, Nicola Salvati

University of Geneva, GSEM, Università Di Pis, DEM

When doing impact evaluation in many cases, it is important to acknowledge the heterogeneity of the treatment effects for different domains. Where certain geographic, socio-demographic, or socio-economic domains may benefit from a program/ policy intervention, others may be worse off. If the domain for which we are interested in the impact, is small with regards to its sample size (or even zero in some cases), then the evaluator has entered the small area estimation dilemma. We propose new methods that allows one to estimate the area specific average treatment effects for such domains. The techniques are based on the modification of Inverse Propensity Weighting and the (robust) small area estimators. In case of using the robust techniques a bias calibration method is provided. These methods enable us to estimate the impact even for those small domains where all units belong to treated or control group. This is through borrowing strength from other domains that can form the control or treated group for this specific small domain. The results of our simulations that are based on realistic scenarios show a clear gain of the proposed techniques to the existing methods. This is the case where the average treatment effects are homogenous or heterogenous across all domains. By

means of these methods we can also provide a map of policy impacts, that can help to better understand the outcome of an intervention and also can help to better target the treatment group(s).

Matching with text data: an experimental evaluation of methods for matching documents and of measuring match quality

Reagan Mozer (Rose), Luke Miratrix, Aaron Russell Kaufman, and L. Jason Anastasopoulos

Matching for causal inference is a well-studied problem, but standard methods fail when the units to match are text documents: the high dimensional and rich nature of the data makes exact or coarsened exact matching infeasible, causes propensity scores to produce incomparable matches, and renders assessing balance and other forms of match quality nearly impossible. In this paper, we develop a framework for characterizing text matching methods, investigate what types of text matching within this framework perform best as measured by human validation, and use our best methods to examine media bias in newspaper articles. In our framework, we first decompose the matching process for text data into (1) a text representation, and (2) a distance metric. Text representations describe how to operationalize unstructured text documents as covariate vectors; we consider n-gram vectors, Structural Topic Model loadings, propensity scores, and document embeddings. Distance metrics capture the high-dimensional proximity of two covariate vectors; we explore Euclidean, Mahalanobis, coarsened exact, and cosine distances, each with a variety of calipers. Altogether we evaluate 84 unique text matching methods. For evaluation of match quality we exploit a useful property of text matching: in many cases human coders can accurately gauge the similarity of a proposed matched document pair when working with text, allowing us to avoid assessing balance metrics in a high dimensional space. To assess the different matching methods, we conduct a systematic multifactor evaluation experiment using human subjects to identify which methods dominate in terms of producing the largest number of matches without sacrificing match quality. We enhance the precision of this approach by developing a predictive model to predict the match quality of pairs of text documents as a function of our 84 distance scores. This model, which we find successfully mimics human judgment, also allows for approximate evaluation of the match quality of new procedures without requiring further human evaluation. Our experimental results identify matching procedures that produce matches with higher subjective match quality than current state-of-the-art techniques. Finally, we apply our framework to a substantive debate in the study of media bias using a data set of front-page news articles from thirteen news sources. Using the optimal matching method identified in our experiments to control for topic selection, we can separate overall media bias into topic selection bias and presentation bias. In our case study we find that both components contribute significantly to media bias, though some news sources have more than one component than the other.

The Morning After: Prescription-Free Access to Emergency Contraceptive Pills

Gregor Pfeifer, Mirjam Reutter

We provide a Europe-wide analysis regarding the prescription-free availability of the so-called morning-after pill, which aims to prevent unintended pregnancy up to five days after unprotected sex. Exploiting country-year variation in the switch from prescription only availability to over-the-counter availability while accounting for the clustered data structure by randomization inference, we measure a sharp increase in sales (107%) and revenue (126%). Most importantly, we find a significant decline in abortions of about 20% and a slight increase in fertility of 3%. This is not accompanied by an increase in sexually transmitted infections, which is in contrast to the recent literature. However, we also find slightly negative effects on some labor market outcomes for women.

Data Based Covariate Selection for High Dimension Low Sample Size Data

Lu Rui and Bryan Keller

Teachers College, Columbia University

With few exceptions, the propensity score literature has focused on estimating causal effects with moderate to large sample sizes. In the social and medical sciences, however, non-equivalent comparison group designs with small sample sizes are not atypical. Conditioning on many covariates in an attempt to satisfy the ignorability assumption may lead to $p > n$ estimation problems or inefficiency. In such cases, data-driven algorithms for selecting minimum covariate subsets may be useful. The primary aim of this study is to investigate the properties of three data-driven covariate selection techniques when used with small sample sizes under varying conditions. Stepwise logistic regression, Bayesian networks and random forests, are studied in a Monte Carlo simulation. In each scenario, we simulate small samples ranging from 50 to 500 with 90 noise covariates and 10 target covariates that have some association with either the propensity score or the potential outcomes. We generate data from several DAGs and implement de Luna, Waernbaum & Richardson (2011) algorithms for covariate selection. Rosenbaum and Rubins (1984; 2009) stepwise logistic regression approach is used as benchmark for comparison. The simulation results indicate random forest and Bayesian networks (using mutual information) outperform stepwise logistic regression by successfully reducing the dimension of the data set and including appropriate covariates suggested by the backdoor path criteria. Propensity scores based on selected covariate sets are used to assess bias and mean square error for each method. Results and implications for covariate selection with small samples are discussed.

Reference

- de Luna, X., Waernbaum, I., and Richardson, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika*, 98, 861-875.
- Guo, S., & Fraser, M. W. (2009). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage. 133-134.
- Rosenbaum, P.R., & Rubin, D.B. (1984). Reducing bias in observational studies using sub

classification on the propensity score. *Journal of the American Statistical Association*, 79, 516-524.

Continuous-time msms: weight estimation and comparison to IPTW

Pal Christie Ryalen, Mats Julius Stensrud and Kjetil Rysland

Department of Biostatistics, University of Oslo

Marginal structural models (MSMs) allow for causal interpretations of longitudinal observational data. The standard MSMs is based on discrete time models. We have developed continuous time MSMs that is a conceptually appealing alternative for survival analysis. It is often assumed that the treatment weights are known, but usually these weights are fundamentally unknown and must be estimated from data. We have proved that continuous-time MSMs are consistent, even when the weights are estimated. We describe an estimation approach based on additive hazards models, and show that it is favorable to the discrete weights estimated with logistic regression in a survival analysis setting. To make our approach viable for applied researchers we have developed the R-packages ‘ahw’ and ‘transform.hazards’, and we give a simple demonstration on how they can be used for causal survival analysis.

Pseval: an r package for evaluating principle surrogates

Michael C Sachs, Erin E Gabriel

Karolinska Institutet, Unit of Biostatistics

We describe an R package that contains the core methods for the evaluation of principal surrogates in a single clinical trial. It provides a flexible interface for defining models for the risk of outcome conditional on treatment and the surrogate, the models for integration over the missing counterfactual surrogate responses, and the estimation methods. Estimated maximum likelihood and pseudo-score methods are implemented for estimation, while bootstrap for inference. Our package uses the + sign to combine function calls into a single object. This is called overloading the + operator and is most well known from the ggplot2 package. This functionality allows users to build up analysis objects starting from the design, and ending with the estimation. The distinct analysis concepts of the design, risk model specification, integration model, and estimation/bootstrap approaches are separated into distinct function calls, each with a limited number of parameters. This makes it easier for users to keep track of their models, makes it easier to understand the methods involved, and allows for the specification of a wide variety of models by mixing and matching the function calls. A variety of post-estimation summary methods are also provided, including print, summary, plot, and testing. We summarize the new features, and the main statistical methods that are implemented in the package and illustrate its use from the perspective of a novice R user.

Help, I am Disabled and Looking for a Regular Job! Who will Help me After January 1st 2015?

Klarita Sadiraj, Stella Hoff, Maroesjka Versantvoort
The Netherlands Institute for Social Research (SCP)

The Dutch Social Work Welfare Act (WSW) is meant for the disabled people who cannot work under normal circumstances. People with a WSW indication can work in or through a social employment service (SW). Since 1 January 2015, new disabled entrants with ability to work are no longer allowed to work in SW. They are expected to find a regular job. This study quantifies the employment effects of this new Participation law. Using administrative Dutch employment data, we assess whether changes in the employment rates and other job characteristics, such as duration of the job or the sort of working contract, can be contributed to the implementation of the new law. We employ a difference-in-differences analysis in combination with propensity score matching and regression discontinuity analysis by exploiting the variation of the outcome before and after the Participation law. The outcome in our study is changes in the employment share two years later for the group that lost the employment protection rights (people with WSW indication on the waiting lists on December 2014) compared to a similar group who did not enjoy these rights prior to the reform. We follow the groups on the waiting lists and those receiving assistance on Decembers of 2010 ÅÚ December 2014. We control for background characteristics and for the development of unemployment rates in different regions. The results of this study show a negative effect of the reform on the employment rates and other job characteristics for the group of people who lost their employment protection rights due to the Participation law.

Individual heterogeneity and time-varying cardiovascular risk factors contribute to explain the obesity paradox in mortality: results from a hierarchical joint modelling approach

Juan G. Sanchez-Plaza^{1,2}, Paloma Saez-Diaz², Ana Arcon-Fernandez², Gernot Pichlera², Jose Bermudeza, Josep Redona,b, Maria Tellez-Plazab.

¹University of Valencia

² Institute for Biomedical Research Hospital Clinic de Valencia (INCLIVA). Funded by the Strategic Action in Health Science, Carlos the third Institute, Madrid Spain (PI15/00071) and co-funded with FEDER funds

Background: The obesity paradox in all-cause mortality, which is the fact that in epidemiologic studies participants with overweight and obesity have consistently been at decreased risk of death as compared to participants with normal weight, remains an enigma. Whole population-based studies with follow-up data on time-varying risk factors and mortality, however, are scarce. **Methods:** We evaluated the multi-adjusted prospective association of baseline body mass index and all-cause mortality before and after considering time-varying cardiovascular risk factors (diabetes, hypertension, dyslipidemia and associated treatments) and body mass index trajectories including random effects in 72,925 cardiovascular disease-free patients at high cardiovascular risk who were users of the universal public health system of Valencian Community (Spain) in 2008-2012. Hierarchical joint models for longitudinal and time-to-event data were implemented with

the JoineR package. Results: The prevalence of obesity, hypertension, dyslipidemia and diabetes at baseline was 47, 83, 88 and 64 %, respectively. The autocorrelation function of body mass index decreased from 0.99 at baseline to 0.89 at the end of follow-up. We observed 1,458 deaths during an average follow-up of 3.3 person-year. The dose-response of body mass index and mortality changed from a U shape to a mostly linear, positive and monotonic shape before and after considering random effects and time-varying cardiovascular risk factors. Conclusions: Individual heterogeneity and time-varying risk factors contributed to explain the obesity paradox in all-cause mortality.

Differences in conditional versus marginal estimates in collapsible estimators

Ian Shrier, Annabelle Redelmeier, Russell Steele, Mireille Schnitzer
McGill University

A study reports the logistic regression risk ratio (RR) for a treatment A on outcome Y is 2.00 after adjusting for confounders. Many readers would misinterpret the RR=2 as a marginal effect. However, regression produces conditional estimates, which break the association between the confounder and the outcome. Alternatively, marginal structural models create marginal estimates, which break the association between the confounder and treatment. It is well-known that conditional and marginal estimates differ with a non-collapsible estimator (odds ratio) when the outcome is common. The purpose of this study is to show when marginal and conditional estimates differ for collapsible estimators (e.g. RR). We first simulated a marginal effect of binary treatment A on binary outcome Y (RR=2), with a binary confounder C causing A (RR=0.7) and Y (RR=0.3) for 1M subjects. Results were unbiased in large samples for both regression (1.99) and marginal structural model (1.99) estimators. Next, we simulated data with the same causal effects as before, but added an interaction A*C on Y (RR=1.5). Conditional estimates yielded RR=2.00 for A, 2.00 for C, and 1.49 for A*C. The marginal structural model estimate was 2.66 for the effect of A on Y. Although combining the conditional estimates for A and A*C provided the correct marginal effect (2.67 as expected), it is rarely done. Further, because most studies are not powered for statistical significance of interaction term, interaction terms are often dropped from models. When marginal estimates are the objective, marginal estimators should be used even for collapsible estimators.

The estimation of neighbourhood effect: comparing logistic regression and multitreatment inverse propensity score weighting

Silan Margherita¹, Arpino Bruno², Boccuzzo Giovanna¹

¹Department of Statistical Sciences, University of Padua ²Department of Political and Social Sciences and The Research and Expertise Centre for Survey Methodology (RECSM), Pompeu Fabra University

Neighbourhood effect is defined as the independent causal effect of a neighbourhood on any health or social outcomes (Oakes, 2004). The main goal in this work is to estimate the effect of neighbourhoods on hospitalized fractures among elderly using data from the Longitudinal Study of Turin, a city in northern Italy with 900.000 inhabitants. For instance, living in green and pedestrian areas decreases the probability to fall and to have serious injuries. Using data from observational studies, individuals are not randomly assigned to each neighbourhood, indeed the geographic distribution of people and the composition of the population in each neighbourhood depends on individuals' characteristics. In order to overcome this problem, different neighborhoods (considered as treatments) will be made comparable using propensity score inverse weighting; in this way it is possible to estimate the neighborhood effect without bias related to the neighborhoods' different observable composition. We compared the performance of two methods to estimate the neighbourhoods effect considering ten areas through a simulation study: logistic regression model that includes neighbourhoods' dummies and inverse weighting with a propensity score computed through Generalized Boosted Models (McCaffrey, 2013). This second method has been implemented with different levels of weights trimming in order to obtain lower relative bias and to reduce variance.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine*, 32(19), 3388-3414.

Oakes, J. M. (2004). The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine*, 58(10), 1929-1952.

From average treatment effects to the identification of treatment responders

Dr Joel Smith¹, Prof. John Forbes², Dr Anna O'Connell², Prof. Alastair Gray¹, Prof. Peter Sandercock³ on behalf of the IST-3 Collaborative Group

¹Nuffield Department of Population Health, University of Oxford

²Graduate Entry Medical School, University of Limerick

³Centre for Clinical Brain Sciences, University of Edinburgh

Experimental designs in clinical research commonly focus upon demonstrating that a particular treatment is better on average than an alternative treatment and reporting an average treatment effect is typical of this literature. Observed patient level characteristics, such as age, sex and severity of disease, may all represent important determinants of treatment response. Identifying baseline covariates associated with treatment responses that may differ markedly from the population average treatment effect is of central interest to policy makers and reimbursement agencies and may also inform

clinical decision making in an age of precision medicine. Treatment-covariate interactions are frequently used to test for the homogeneity of treatment effects, but the lack of statistical power to detect subgroup differences increases the probability of Type-II errors. Modern methods using data driven approaches are also more informative for learning about treatment effect heterogeneity rather than the identification of patients with positive treatment effects. We examine how an alternative approach using Bayesian hierarchical linear models with a sceptical prior can be used to identify individuals who benefit most from a particular treatment while accounting for multiplicity and statistical power. The use of sceptical priors within a novel approach called Bayesian Credible Subgroups ensures evidence of large heterogeneous effects are unlikely but also does not rule out the possibility that they may exist through the posterior distribution. We use data from a large multicentre randomised controlled trial in stroke (Third International Stroke Trial) to derive more credible subgroups based on a decision rule of three possible class memberships.

A causal model of cell transformation assays in multi-laboratory experiments

Federico M. Stefanini

Department of Statistics, Computer Science, Applications, University of Florence

Cell Transformation Assays (CTAs) are in-vitro methods increasingly used to screen chemical substances for carcinogenicity without resorting to animal-based testing. Multi-laboratory replicates of the same randomized experiment sometimes produce inconclusive evidence thus the whole testing procedure is rejected. In the causal perspective on CTAs adopted in this work, it is recognized that serum and cell passage (steps from the primary cell culture) may exert a relevant effect on the response, the number of fully transformed foci grown within Petri dishes. Dosage (quantification) of grow factors and other important serum-medium components is likely to be not accepted in production because it is expensive and time consuming. Similarly, the molecular basis determining differences due to cell passage is not yet well understood and challenging to assess. In this work, a Bayesian causal model for CTA is presented and an extension is proposed where indirect assessment of laboratory dependent supplies is performed using positive controls in multi-laboratory experiments. Modifications of the current CTA protocol are proposed to avoid the introduction of different versions of the intended treatment in multilaboratory experiments.

Causal inference in the relative survival framework

Elisavet Syriopoulou¹, Mark J Rutherford¹, Paul C Lambert^{1,2}

¹Biostatistics Research Group, Department of Health Sciences, University of Leicester,
Leicester, UK

²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm,
Sweden

In cancer epidemiology, the event of interest is usually death due to cancer. However, other events, so-called competing events, which can potentially impede the occurrence of the event of interest may be present. Relative survival is a commonly used measure in cancer studies that has the advantage of circumventing problems caused by the inaccuracy or nonavailability of the cause of death information. This project extends causal inference and mediation analysis methods to the relative survival framework for estimating standardised survival differences. By applying the mediation analysis setting and using flexible parametric modelling, a methodology that has a number of advantages in terms of predictions and extrapolations in comparison with other commonly used survival models such as the Cox model, we illustrate the methods and explore mechanisms that generate inequalities in prognosis of colon cancer patients across different socioeconomic groups. The analysis showed a substantial variation in survival of patients by socioeconomic status with patients from a high socioeconomic background found having higher survival. Within socioeconomic groups, survival was strongly determined by stage at diagnosis that acts as a mediator for the effect of socioeconomic status to time-to-death. We also estimated the potential gain in life years by removing the stage differences seen across socioeconomic groups. We conclude that significant improvements would be observed in survival if appropriate policy interventions were implemented. Extending causal inference to relative survival enables a better understanding of the causal structure of the variables involved in the analysis and aids targeting the most affected population groups.

Impact of declining exposure to secondhand tobacco smoke in public places to decreasing smoking-related cancer mortality in the US population.

¹Esther Garcia-Esquinas,^a Angelica Jimenez^{2,3}, Roberto Pastor-Barriuso⁴, Maria Grau-Perez^{2,3}, Miranda R. Jones⁵, Beatriz Perez-Gomez⁴, Ana Navas-Acien⁶, Maria Tellez-Plaza⁵

¹Autonomous University of Madrid

²University of Valencia

³ Institute for Biomedical Research Hospital Clinic de Valencia

⁴ National Center for Epidemiology at the Carlos the third Health Institute (Madrid)

⁵ Johns Hopkins University (Baltimore)

⁶Columbia University (New York). Funded by the Strategic Action in Health Science, Carlos the third Institute, Madrid Spain (PI15/00071) and co-funded with FEDER funds.

Background: The major decrease in exposure to second hand smoke (SHS) in public places in recent decades could have contributed to the decline in smoking-related cancer mortality observed in the US population. Methods: Prospective study among 11,856

non-smoking adults aged ≥ 40 years who participated in NHANES 1988-1994 or 1999-2004 and were followed for mortality through 2006. We estimated the amount of change in cancer mortality over time attributed to the intermediate pathway of changes in SHS exposure, after adjustment for risk factors and SHS exposure at home. Results: The adjusted smoking-related cancer mortality rate ratios (95%CI) for a two-fold increase in serum cotinine and a 1-hour increase in SHS exposure time at work were: 1.10 (1.03, 1.17) and 1.14 (1.06, 1.24), respectively, for all-cancer; and 1.13 (1.03, 1.24) and 1.14 (1.02, 1.26), respectively, for smoking-related cancer. The absolute reduction in mortality comparing 1999-2004 to 1988-94 was 75.8 (-25.5, 177.0) and 77.0 (2.6, 151.4) deaths/100,000 person-years, for all-cancer and smoking-related cancer, respectively. Among these avoided all-cancer deaths, 45.8 (2.8, 89.5) and 18.1 (-1.2, 39.6)/100,000 person-year were attributable to changes in serum cotinine and SHS exposure time at work, respectively. The corresponding numbers of smoking-related cancer avoided deaths were 36.4 (0.7, 72.8) and 9.9 (-3.8, 24.9)/100,000 person-year. Conclusions: Declines in SHS exposure were associated with reductions in all-cancer and smoking-related cancer mortality, supporting that smoking bans in public places may have reduced cancer mortality among non-smoking adults.

Simplification in causal inference

Santtu Tikka and Juha Karvanen

Department of Mathematics and Statistics
University of Jyväskylä, Finland

Identification of a causal effect of an action means deriving an expression from the observed probability distribution for the interventional distribution resulting from this action. Such an expression can be obtained for an identifiable causal effect by an identifiability algorithm or by application of do-calculus. Software implementations are available to reduce the need for manual derivations. However, in many cases an identifiability algorithm may return a complicated expression that contains variables that are in fact unnecessary. In practice this can lead to additional computational burden and increased bias or inefficiency of estimates when dealing with measurement error or missing data. We present a method to simplify expressions for causal effects. As the first step, we apply graphical criteria to detect variables that are redundant in identifying causal effects. These variables are removed or projected before an identifiability algorithm is applied. As the second step, we apply an algorithm that seeks to symbolically eliminate unnecessary variables from the expressions returned by the identifiability algorithm. The simplification algorithm takes advantage of the structure of the underlying graphical model. As a result of these two steps, the length of the expression can be sometimes reduced as much as 70%. Our method is applicable to all causal effect formulas and is readily available in the R package `causaleffect`.

References

- S. Tikka, J. Karvanen, Simplifying probabilistic expressions in causal inference, *Journal of Machine Learning Research*, 18(36), 1-30, 2017.
- S. Tikka, J. Karvanen, Enhancing identification of causal effects by pruning, Submitted manuscript, 2017.

Effects of maternal education on early non-monetary investments in child development

Emine Kubra Usta

University of Bologna (PhD candidate)

In this study, I document the effects of education on the early investments of mothers in their young children in the context of a middle income developing country. Early investments are crucial as they are very productive in child development and nonmonetary investments are helpful to compensate income gap in child development for economically disadvantaged families. I exploit the changes in parental investments induced by a compulsory schooling reform in Turkey and focus mostly on health investments. I do not find any effect on the two main variables of interest: duration of breastfeeding and taking the doses of vaccines that are in the universal program and provided for free. Then, I investigate mothers health care use behavior. I find a strong increase in the C-section births among more educated women which is associated with higher chance of obesity and diabetes in later years of child's life in the medical literature. The only positive effect of education is on the probability of child's receiving a postnatal check in a formal health institution which is in line with the finding that educated mothers started pre-natal care earlier and received more prenatal care in the formal institutions. The net impact of these prenatal and postnatal mother investments is hard to tell in terms of child's health capital. The puzzling aspect of the findings is that although mothers increased health care use for themselves during pregnancy, there is no strong spillover from their own health care use to health care use for their children. This point has important policy implications as it shows the necessity of more targeted policies to increase parental inputs in early child development.

Causal time series segmentation

Aleksander Wieczorek and Volker Roth

University of Basel, Switzerland

We propose a new method of discovering causal relationships in temporal data based on the notion of causal compression. Expressing Pearl's intervention calculus in terms of information theoretic concepts capturing the difference between interventional and observational distributions resulted in rich literature. As a result, asymmetrical information theoretic measures are used for modelling causal relationships in graphical models. In this paper, we propose a method of identifying time points responsible for the transfer of causal information in time series. We first unify existing frameworks of directed information (in time series, time series with reduced assumptions regarding time ordering and in general Pearlian graphs) and then use the directed information as an information theoretic tool for quantifying causality. We introduce chain rule for directed information and use it to motivate causal sparsity. We show an application of the proposed method: causal time series classification which selects time points capturing the causal flow between time points belonging to different signals. We prove that modelling of causality

in the adopted set-up only requires estimating the copula density of the data distribution and thus does not depend on its marginals. We evaluate the method on EEG data. Unlike cognate approaches to causality modelling in time series, we do not treat whole time series as nodes and do not model relationships between such nodes. We propose to model causal relationships between specific time points of the time series instead. We also do not have to make any assumptions concerning stationarity.

Evaluating evidence of mechanisms in medicine: principles and procedures

Jon Williamson
University of Kent

Russo & Williamson (2007) argued that causal inference in the health sciences requires making use of two sorts of evidence: evidence that the putative cause is associated with the putative effect (obtained, e.g., from clinical or epidemiological studies) and evidence of the mechanism of action (obtained, e.g., from laboratory and animal studies). Present-day evidence-based medicine (EBM) tends to focus on the former evidence to the exclusion of the latter. Partly this is because it is perceived to be difficult to assess evidence of mechanisms and to integrate this evidence with evidence of association in order to evaluate a causal claim. This talk will report on new methods for tackling this problem, developed by researchers on the EBM+ network. The methods are split into three phases. First, strategies are developed for searching for relevant mechanistic studies. Second, these mechanistic studies are evaluated according to fixed criteria and the status of a mechanism of action assessed. Third, this status is combined with the status of the claimed association in order to provide an overall assessment of the causal claim. The talk will go on to discuss how evidence of mechanisms can be used to help extrapolate a causal claim from a study population to a target population of interest. In addition, some evidence appraisal tools will be presented that facilitate the evaluation of mechanistic evidence.

Matching on generalized propensity scores with continuous treatments

Xiao Wu, Fabrizia Mealli, Danielle Braun, Francesca Dominici

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
Department of Statistics, Informatics, Applications, University of Florence, Florence, Italy

Generalized propensity scores (GPS) are used to adjust for confounding and estimate causal effects in observational studies when the treatment (or exposure) is continuous. Existing approaches have been proposed to implement GPS using inverse probability of treatment weighting or by including the estimated GPS as covariate in the outcome model. We propose a new GPS matching framework to estimate causal effects when the treatment is continuous in observational studies. Assuming non-interference, overlap and weak unconfoundedness, we show that under moderate regularity conditions, matching on GPS removes bias associated with observed pre-treatment confounders.

We describe the conditions under which our proposed matching estimators attain 1/2-consistency. We conduct simulation studies with varying degrees of confounding bias and shapes for the exposure-response function and show that GPS matching eliminates confounding bias. We also compare the proposed matching approach to existing state-of-the-art alternatives by evaluating bias reduction, mean squared error and covariate balance. We apply the proposed method to Medicare Part A data in New England, in which our interest is in estimating the causal effect of long-term exposure to fine particles (PM_{2.5}), a continuous exposure (i.e. the treatment assignment), on mortality for the period from 2000 to 2015.

Key words: Causal Inference, Continuous Treatment, Generalized Propensity Scores, Matching, Observational Study
