Education for Statistics in Practice

Assessing direct and indirect effects - from structural equation models to causal mediation analysis

Presenters: Vanessa Didelez[1] and Johan Steen[2]

1: Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany
didelez@leibniz-bips.de

2: Ghent University Hospital, Intensive Care Medicine, Belgium
johan.steen@ugent.be

Extended abstract

This short course will cover concepts and inference for direct and indirect effects, as relevant in situations where we are interested in whether and how much the effect of an exposure/treatment is mediated via other factors to affect a final outcome. This is often relevant for gaining an understanding of the underlying causal mechanisms and hence planning of potential future interventions. For instance, we may want to know whether changes in children's well-being affect cardio-metabolic markers via changes in life-style. Traditionally this has been modeled with linear structural equation models (SEMs) but the limitations of this methodology have lead to alternative approaches and generalisations known as *causal mediation analysis*. The latter is based on potential outcomes and causal graphs. In this course we will show how (in)direct effects can be defined non-parametrically, i.e. without presuming any particular parametric model. Recent approaches to causal mediation modeling and inference will then be addressed, such as the mediational g-formula or natural effect models, as well as their practical application with R packages such as medflex.

Participants are expected to have a fair knowledge of regression analysis and generalised linear models.

Structural Equation Models (SEMs)

Structural equation models, especially linear SEMs, have long been a popular tool in the social sciences. They are increasingly used in medical applications and epidemiology, e.g. to investigate the mechanisms linking childhood obesity, physical activity, sleep patterns and well-being to health outcomes. SEMs cover a wide range of models, especially involving latent variables, such as factor models. In this course we will focus on their role in modelling complex direct and indirect relations linking sets of variables. In this context, the key property of SEMs is that they allow one variable to be the response in one regression equation but a predictor variable in another regression equation. The famous Baron and Kenny (1986) result for linear SEMs is the decomposition of the total effect into the direct and indirect effect in terms of regression coefficients. This approach, however, has been much criticised (De Stavola et al, 2015; VanderWeele, 2016) due to its strong assumptions / oversimplification and lack of generalisability.

Causal mediation analysis

In their seminal papers, Robins and Greenland (1992) and Pearl (2001) proposed general approaches to defining direct and indirect effects in terms of counterfactuals that are *not* confined to particular parametric models. As will be discussed in the course, the key innovation here, is to make it explicit in terms of interventions what we actually mean by direct and indirect effects, i.e. what these mean in terms of useful real-world quantities, and what fundamental assumptions are required to identify them from observational or experimental data (Didelez, 2018). Moreover, at a more technical level, the linear regressions of the linear SEMs can now be replaced by any suitable and flexible regression models. This has lead to a flurry of research activities on computational issues and applications of these new concepts in the biomedical and epidemiological literature. We will consider how, in certain cases, explicit formulas can still be obtained for the (in)direct effects, but also how Monte Carlo methods can be used for fitting more complex non-linear models (Imai et al, 2010; see R package mediation).

A further development is to model and parameterise direct and indirect causal effects within one model explicitly, instead of using a set of regressions-type models that need to be combined; this alternative is known as natural effect models and can be implemented with the R package medflex (Steen et al, 2017). We will demonstrate with examples how these models are easier to interpret and how they can easily be fitted using inverse-probability weighting and other methods.

Practical implementation

Throughout the course will focus on basic concepts and practical interpretation more than on technical details. Methods discussed during the lectures will be illustrated by examples demonstrating the use of the R package medflex, with brief comparisons to the packages sem and mediation, to implement the analyses.

About the presenters

Vanessa Didelez is Professor of Statistics with Focus on Causal Inference at the University of Bremen and the Leibniz Institute for Prevention Research and Epidemiology – BIPS. Her research interests include graphical models / causal diagrams, methods for time-varying confounding and general time-to-event data, in addition to causal mediation analysis.

Johan Steen is a postdoctoral researcher / statistician at the Intensive Care department of the Ghent University Hospital. He completed his PhD on causal mediation analysis at Ghent University in 2016 and is author of the medflex R package for flexible mediation analysis. His current research focuses on drawing and improving causal inferences from routinely collected hospital data to better inform clinical decision making.

References

Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51(6): 1173–1182.

De Stavola, B.L. and R.M. Daniel, G.B. Ploubidis, N. Micali. 2015. Mediation analysis with intermediate confounding: Structural Equation Modeling viewed through the causal inference lens. American Journal of Epidemiology 181(1): 64–80.

Didelez V. 2018. Defining causal meditation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis. (Epub 2018 Sep 14).

Imai, K., L. Keele, and D. Tingley. 2010. A general approach to causal mediation analysis. Psychological Methods 15(4): 309–334.

Pearl, J. 2001. Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, ed. M. Kaufmann, 411–420. San Francisco, CA.

Robins, J.M., and S. Greenland. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3(2): 143-155.

Steen, J. and T. Loeys and B. Moerkerke and S. Vansteelandt. 2017. medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. Journal of Statistical Software, 76(11): 1 – 46.

VanderWeele, T.J. 2016. Mediation Analysis: A Practitioner's Guide. Annual Review of Public Health, 37(1): 17-32.