Inferring causal pathways from data: challenges and some solutions

Stijn Vansteelandt
Ghent University, Gent, Belgium, and the London School of Hygiene & Tropical Medicine, London, United Kingdom

The question what is the mechanism whereby an exposure affects an outcome is key to many fields of science, most notably medicine, epidemiology, psychology and sociology. A proliferation of so-called mediation analysis techniques has therefore developed since the early 1980’s with the aim to infer mechanistic pathways from data, using the linear structural equation model as a backbone. Here, each pathway describes how exposure affects an intermediate variable or mediator, which then in turn affects the outcome. Developments in causal inference, pioneered by Robins, Greenland and Pearl, have revealed major limitations of these techniques. In this talk, I will discuss these limitations and provide an introduction to causal mediation analysis, which has mostly developed over the past 10 years. The majority of these techniques has focused on simple applications involving a single mediator, in view of methodological difficulties of handling multiple mediators, which I will discuss. However, the scientific question of interest often makes reference to multiple pathways. Moreover, even when a single pathway is of interest, recognizing multiple mediators is usually necessary for a valid analysis because of confounding induced by other mediators. I will address this concern by discussing recent developments on multiple and longitudinal mediators. I will illustrate the developments along the lines of various case studies in genetics, public health and HIV prevention, as well as in a more extensive re-analysis of the LEADER trial to explain the important reductions in the hazard of a first major cardiovascular event that have been observed in patients with Type II diabetes on liraglutide (as opposed to placebo).