Detecting New Risk Factors via Causal Discovery in Aortic Surgery

Mirthe van Diepen

Understanding the causal relationships between demographic information and biomarkers can be extremely useful to get a better understanding of causal risk factors in healthcare. It can motivate future studies to search for an intervention that lowers the risk or the search for possible treatment alternatives that can improve quality of life expectations. Using random controlled trials (RCTs) we can try to infer specific causal relationships However, it is not always possible to directly intervene on (proxy) variables due to ethical reasons or it is just impossible in practice. Causal discovery algorithms try to address this problem, by searching for the causal structure between variables in an observational data set instead of using interventions on the variables. However, currently, in medical journals, the methods to analyze data are usually not based on causal discovery methods due to the assumptions made which are difficult to test for, and the non-intuitive definitions that are required for this field. Here we show how to handle these using a specific case study that exhibits many of these challenges. This study is motivated by a data set containing subjects who had aortic surgery at the St. Antonius Hospital in Nieuwegein. We use this data set to demonstrate what important steps are needed for the analysis. Challenges of this aortic surgery data set are (1) small sample size, (2) consisting of a complex combination of very different variables, both discrete and continuous, (3) unknown causal structure (there might be unknown confounders or cycles in the causal structure), (4) context variables and time-dependent variables (variables from the different phases in the perioperative period), and (5) missing values. We will show what to consider when choosing a causal discovery method and the impact of different choices for the hyperparameters for it. Moreover, we suggest how one can combine the outputs of a causal discovery method with bootstrapping to make it more robust for small data sets, how to deal with context variables, and how to deal with mixed data.