Causal modeling in epidemiological practice

van der Wal, W.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Summary

This thesis addresses the application and implementation of modern causal modeling techniques, in particular marginal structural models (MSMs), with a focus on epidemiological practice.

In the general introduction (chapter 1), we give a literature overview. MSMs model the population distribution of potential outcomes. These are the outcomes that could have been observed, perhaps contrary to the fact, when a certain exposure level would have been given, for all possible exposure levels. Marginal causal effects as estimated using MSMs, are contrasts applied to the population distribution of potential outcomes. Such a causal effect can be estimated either from clinical trial data or from observational data. When using observational data in which confounding is present, standard conditional methods to correct for confounding can have important drawbacks. These include 1) adjusting away the effect, 2) non-collapsibility and 3) inducing bias through collider-stratification. An easily implemented alternative to conditioning to correct for confounding, that does not have these drawbacks, is inverse probability weighting (IPW). When using IPW, observations are weighted by the inverse of the probability of the observed exposure level, given the observed value of confounders. It is straightforward to combine different IPW corrections, and to use IPW to correct for forms of selection bias such as informative censoring. Other methods to fit MSMs include G-computation, doubly robust (DR) estimation and targeted maximum likelihood estimation (TMLE).

We have developed the R package ipw for estimating inverse probability weights, which we describe in chapter 2. We illustrate how to use the package to fit MSMs through IPW. Our package can be used with data from a point treatment situation as well as with a time-varying exposure and time-varying confounders. It can be used with binomial, categorical, ordinal and continuous exposure variables. Inference can be performed using robust standard error estimators such as a sandwich estimator, or by bootstrapping.

IPW to fit a MSM is applied throughout this thesis. In the study described in chapter 3, we have compared the causal effect of peritoneal dialysis (PD) and hemodialysis (HD) treatment on mortality in end-stage renal disease patients, using observational data. We fitted a MSM to estimate the causal effect of dialysis
type, adjusted for baseline as well as time-varying confounders, and adjusted for informative censoring. We used the MSM to compare the hazard of death as well as cumulative survival between the potential treatment trajectories “always PD” and “always HD” over time since the start of dialysis treatment, conditional on age and diabetes mellitus status. For all ages, with or without diabetes, our results indicate equal survival during an initial period of 24 to 36 months, after which survival for PD gradually becomes worse than for HD, although not statistically significantly in this study.

In chapter 4, we describe a study in which we have estimated the causal effect of glomerular filtration rate (GFR) on mortality in end-stage renal disease patients on dialysis. GFR is a continuous variable with zero-inflation. We describe a MSM to estimate the causal effects both of the value of GFR when it is not completely lost, and of the full loss of GFR on mortality. Effect estimates were adjusted for possible baseline and time-varying confounders using IPW. From the results, we conclude that preventing or delaying the full loss of GFR can improve survival in dialysis patients.

In chapter 5 we explore the relationship between AIDS defining conditions (ADCs), use of combination antiretroviral therapy (cART), calendar time and mortality. Using observational data, we have fitted a MSM through IPW to estimate the causal effect of initial ADCs and cART use on mortality, including the interaction, and allowing for effect-modification by calendar time. Our results show that 1) the effects of ADCs on mortality have decreased over calendar time, 2) the ADCs differ in their impact on mortality and 3) cART use reduces the mortality risk considerably for most ADCs, except for active tuberculosis, herpes simplex and cytomegalovirus for which cART does not change the mortality risk, and mycobacterial infections other than tuberculosis, for which cART therapy increases the mortality risk. We have contrasted our approach with an alternative using a history-adjusted MSM (HA-MSM), and describe some disadvantages of the latter. HA-MSMs can suffer from bias due to misclassification, can require substantially more computer memory than MSMs, and produce effect estimates that are not as straightforward to interpret as those obtained from MSMs.

In chapter 6 we examine the development of health related quality of life (HRQoL) and functional status over time in end-stage chronic obstructive pulmonary disease (COPD) patients. We used standard conditioning both to adjust for informative censoring and to examine subgroups in a longitudinal study of quality of life measurements. In this study it was not necessary to use IPW to correct for informative censoring, since the drawbacks of standard conditioning did not apply.

Because IPW can suffer from small sample bias (see chapter 8), it is useful to study alternative techniques to fit MSMs that do not suffer from this drawback. In chapter 7 we describe an easily implemented algorithm that can be used to fit a MSM through G-computation, when estimating the effect of a secondary illness with a marker for primary disease progression that is both a confounder and intermediary for the effect of the secondary illness. Simulation confirms that the
algorithm can produce unbiased causal effect estimates. We used the algorithm to estimate the total causal effect of active tuberculosis on AIDS-related mortality in HIV-infected individuals, corrected for confounding by time-varying CD4 count, and found a hazard ratio of 3.5 (95% confidence interval 1.2-9.1).

In chapter 8 we describe another easily implemented alternative to IPW. We propose an algorithm that performs IPW iteratively (iterative inverse probability weighting - IIPW), combined with truncating the weights at each iteration. We have examined the properties of the resulting causal effect estimators in simulation studies with different point treatment situations. We have found that such IIPW estimators 1) suffer substantially less from small sample bias than IPW estimators, 2) are less variable than standard IPW estimators, and 3) can be robust against model misspecification. We give heuristics why this method improves upon standard IPW.

We summarize the main findings of our study in the general discussion (chapter 9), these are all but one repeated above in this summary. What is not yet noted in this summary is the conclusion that it is important to distinguish between conditioning to correct for confounding (or for selection bias such as informative censoring), and conditioning to estimate conditional effects. We also give some important suggestions for future research, including 1) to study the small sample behavior of causal effect estimators, 2) to study the use of different inference methods when using IPW, such as robust variance estimators, bootstrapping or the delta method, 3) to further implement causal modeling methods, 4) to further disseminate causal modeling methods, 5) to study the interpretation of causal effect estimates for exposures other than medical treatments such as the initiation of a disease and 6) to tailor causal modeling methods to specific practical settings such as cost benefit-analysis or longitudinal public health research.