University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2005 *Paper* 191

Population Intervention Models in Causal Inference

Alan E. Hubbard* Mark J. van der Laan†

*Division of Biostatistics, School of Public Health, University of California, Berkeley, hubbard@stat.berkeley.edu

†Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

Population Intervention Models in Causal Inference

Alan E. Hubbard and Mark J. van der Laan

Abstract

Marginal structural models (MSM) provide a powerful tool for estimating the causal effect of a] treatment variable or risk variable on the distribution of a disease in a population. These models, as originally introduced by Robins (e.g., Robins (2000a), Robins (2000b), van der Laan and Robins (2002)), model the marginal distributions of treatment-specific counterfactual outcomes, possibly conditional on a subset of the baseline covariates, and its dependence on treatment. Marginal structural models are particularly useful in the context of longitudinal data structures, in which each subject's treatment and covariate history are measured over time, and an outcome is recorded at a final time point. In addition to the simpler, weighted regression approaches (inverse probability of treatment weighted estimators), more general (and robust) estimators have been developed and studied in detail for standard MSM (Robins (2000b), Neugebauer and van der Laan (2004), Yu and van der Laan (2003), van der Laan and Robins (2002)). In this paper we argue that in many applications one is interested in modeling the difference between a treatment-specific counterfactual population distribution and the actual population distribution of the target population of interest. Relevant parameters describe the effect of a hypothetical intervention on such a population, and therefore we refer to these models as intervention models. We focus on intervention models estimating the effect on an intervention in terms of a difference of means, ratio in means (e.g., relative risk if the outcome is binary), a so called switch relative risk for binary outcomes, and difference in entire distributions as measured by the quantile-quantile function. In addition, we provide a class of inverse probability of treatment weighed estimators, and double robust estimators of the causal parameters in these models. We illustrate the finite sample performance of these new estimators in a simulation study.

1 Introduction

Originally, Robins (2000a) proposed estimates for a new class of models that relate the distribution of counterfactuals in a population to the universal application (or exposure) to a treatment (or risk factor). For example, Robins proposed estimates for parameters such as,$E[Y_a|V]$, which is interpreted as the mean of the outcome of interest in the target population when every subject receives the same treatment (or exposure), a, among strata defined by a subset of the baseline covariates, V (e.g., gender). Several estimators have been proposed for this parameter, including the likelihood-based G-computation estimator, and those based on an estimating equation approach (the inverse-probability-treatment-weighted, IPTW, and the doubly-robust extension; van der Laan and Robins (2002)). Given a specific model, such as $E[Y_a|V] = m(a, V|\beta) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$, these estimators require estimators of 1) an estimate of the conditional distribution of the outcome given the treatment and confounders, 2) an estimator of the treatment assignment distribution or 3) both, respectively. Here, we discuss similar approaches to a new class of regression models that are particularly relevant to population-based studies of risk factors. We term these models, population intervention models (PIM).

The estimating equation approach provides a beautiful and almost formulaic mathematical method for deriving a class of estimators for a wide variety of causal inference problems. However, before applying the machinery (van der Laan and Robins (2002)), one must first define the parameter of interest. We suggest that $E[Y_a|V]$ is often not the parameter of interest in observational studies of risk factors. The MSM's compare, for instance, the prevalence of disease in a population in where every one smokes 3 packs of cigarettes versus everyone smokes 2 packs versus ... Of more public health interest might be a model which relates the prevalence of disease in the current target population (with its distribution of smokers) to a population where no one smokes. That is, perhaps a more compelling regression might be one that returns estimates of the impact of intervention in a population, relative to the current distribution of a risk factor, for instance, $E[Y] - E[Y_a]$, or $E[Y]/E[Y_a]$ where a is a target level of the risk factor. Epidemiologist might recognize this parameter as something akin to attributable risk. In this paper, we will propose a new class of estimators for parameters comparing the distribution of an outcome in the current target population relative to that in the same population when all individuals have uniform treatment/exposure.

1.1 Organization of Paper

The major goal of the paper is to introduce a new set of estimating functions that estimate the intervention parameters of interest as a potentially smooth function of both the target levels of a risk factor in a population, a, and the strata, V. However, we will start by introducing formally the data, model and possible parameters of interest (section 2). In section 3, we discuss a set of possible approaches when only one target intervention is relevant, for instance smoking (comparing the current risk of a population to that if everyone smoke one pack/day is not a relevant parameter, at least with regards to public health). Specifically, we present estimators $m(V \mid \beta) = E[Y_{a^*} \mid V] - E[Y \mid V]$ for a single a^*. For these parameters, we first present estimators that require, among other assumptions, correct specification of the model for $E[Y \mid V]$ (section 3.1) and then follow-up with a new class of estimators that are consistent even if the model for $E[Y \mid V]$ is inconsistent (section 3.2). These sections serve as a prelude to section 4, where we present estimators for intervention parameters comparing the distribution of $Y_a \mid V$ and $Y \mid V$ as a function of both V and a; we do so for estimating risk differences, risk ratios and a more general quantile-quantile function. In Section 5 we present a simulation study and conclude this paper with a discussion in Section 6.

2 Data, Model and Parameters of Interest

Consider a study in which we observe (chronologically ordered) on each randomly sampled subject baseline covariates W, a treatment/exposure variable A, and a final outcome Y. The observed data structure $O = (W, A, Y)$ represents a random variable with a certain population distribution P_0, and the sample $O_1, \ldots, O_n$ represents n i.i.d. observations of O. In this article we are concerned with estimation and testing of the effect of an intervention, A, on the distribution of Y relative to the actual population distribution of Y, possibly within strata defined by a subset of baseline covariates $V \subset W$.

In order to formulate such a parameter of interest and corresponding model we will use the counterfactual framework in which the observed data structure O is viewed as a missing data structure on a collection of treatment specific data structures, the so called counterfactuals. First, we assume that each subject has a set of associated treatment specific outcomes $(Y_a : a \in \mathcal{A})$, so that $X = ((Y_a : a \in \mathcal{A}), W)$ represents a full data random variable with a population distribution F_{X0}. Here $\mathcal{A}$ denotes the set of possible treatments. Secondly, one assumes that the observed data, $O = (W, A, Y = Y_A)$, corresponds with observing the A-specific component of the collection of treatment specific data structures $X = (X_a \equiv (y_a, W) : a \in \mathcal{A})$ corresponding with the treatment A the subject actually received, that is, treatment assignment serves as a censoring variable (A is short-hand for a vector of censoring indicators, where the 0's and a single 1 correspond with unobserved counterfactuals,X_a, $a \neq A$, and the observed counterfactual, X_A, respectively). This assumption is referred to as the consistency assumption (CA). Since O is now a missing data structure on the full data structure X with missingness variable A, the distribution of O is a function of the distribution F_{X0} of X and the conditional distribution g_0 of treatment A, given X: $O \sim P_{F_{X0},g_0}$, where $g_0(a \mid X) \equiv P(A = a \mid X)$ is called the treatment mechanism. In order to have the an identifiable causal parameter we will also assume the so called randomization assumption on g_0:

$$g_0(a \mid X) = g_0(a \mid W)$$

or equivalently, A is conditionally independent of X, given W, an assumption often referred to as *no unmeasured confounding*.

Under these assumptions the density of the observed data structure w.r.t. an appropriate dominating measure factorizes as follows:

$$p_0(O) = f_W(W) f_{Y|A,W}(Y \mid A, W) g_0(A \mid X).$$

If $g_0(a \mid W) > 0$ F_W-a.e., then it also follows that the first two factors of the observed data density identify the counterfactual distribution of (Y_a, W):

$$p_{Y_a,W}(y, w) \equiv f_{Y|A,W}(y \mid A = a, w) f_W(w), \tag{1}$$

where (1) was named the G-computation formula by Robins (2000a).

This counterfactual framework allows us to define our parameter of interest as some difference between the conditional distribution $F_{Y_a|V}$ and the population distribution $F_{Y|V}$, where this difference can be parameterized in several ways. Formally, our parameter of interest is

$$\psi_0(a, V) = \Psi(F_{X0}, g_0)(a, V) \equiv \Phi(F_{Y_a|V}, F_{Y|V})$$

for some known functional Φ. In words, ψ_0 measures the effect of setting $A = a$ for everybody in our population on the distribution of Y, within strata $V = v$. A Φ-specific intervention model is now defined as a model on this parameter ψ_0:

$$\psi_0(a, V) = m(a, V \mid \beta_0)$$

for some Euclidean parametrization $\beta \to m(a, V \mid \beta)$. The following sections present estimators for three classes of parameters: the additive risk, relative risk, and quantile-quantile function:

$$\begin{aligned} \psi_{0,AR}(a, V) &= E(Y_a \mid V) - E(Y \mid V) = m(a, V \mid \beta_0) \qquad (2) \\ \psi_{0,RR}(a, V) &= \frac{E(Y_a \mid V)}{E(Y \mid V)} = m(a, V \mid \beta_0) \\ \psi_{0,QQ}(a, V)(q) &= F^{-1}_{Y_a|V} F_{Y|V}(q) = m(q \mid a, v, \beta_0). \end{aligned}$$

Note, that the structural nested mean models of Robins (Robins (1989) and Robins (1994) provide parameters, which compare the mean of an outcome, within strata, under the "natural" population distribution of treatment versus that under an intervention when $A = 0$ for all subjects. Also, as discussed in the appendix of van der Laan et al. (2005), one can also derive the marginal or stratified causal parameters of interest from these structural nested mean models.

3 Estimating Additive Risk for a Single Target Intervention

This section presents estimators when there is only one suitable choice for the target intervention, a^*, for instance smoking cigarettes ($a^* = 0$ in this case). Estimators of interest or those for parameters that are functions of the conditional mean of Y and Y_{a^*} given V, or $\psi_0(V) = \phi(E(Y_{a^*} \mid V), E(Y \mid V))$, and approximated by some finite parameter model $m(V \mid \beta)$, where β are regression coefficients. The first subsection will present estimates of the parameter of interest, $\psi_0(V)$ as a regression of the difference of estimators of $E(Y_{a^*} \mid V)$ and $E(Y \mid V)$ on some function of V; these are referred to as *substitution* estimators. In section 3.2, a direct estimation equation approach for $\psi_0(V)$ and estimators are introduced that are robust to misspecification of the model for $E(Y \mid V)$. Note, that table 1 lists the specific assumptions necessary for the consistency of each of the estimators presented in this section (note that the IPCW estimators only need that $g(a^* \mid W) > 0, F_{W_0} - a.e.$ for the single a^* of interest).

3.1 Substitution Estimators

Estimators presented in this section rely on the consistency of an estimate of $\eta(V) \equiv E(Y \mid V)$ and use previously proposed estimators for the conditional counterfactual regression $E(Y_{a^*} \mid V)$, for which we discuss several existing approaches. First, Robins (2000a) formalized a G-computation approach that is easy to implement in this point treatment setting. Specifically, the G-computation regression estimator in this case is:

$$E[Y_a|V] = E(E[Y_a|W]|V) = E(E[Y|A = a, W]|V).$$

Thus, one simply needs a regression estimator for $E[Y|A = a, W]$, for example linear regression if Y is continuous and logistic if Y is binary. Then, $E[Y_a|V]$ is estimated as a regression of $\hat{E}[Y|A = a^*, W_i]$ on V_i; we propose estimating the additive PIM as a function of V by regressing $\hat{\eta}(V_i) - \hat{E}[Y_a|V_i]$ on V_i assuming the model $m(V \mid \beta)$, for instance, $m(V \mid \beta) = \beta_0 + \beta_1 V$.

Another approach for estimating $E[Y_{a^*}|V]$ is called the inverse probability of censoring weighted (IPCW) estimator (Robins and Rotnitzky (1995)), which does not relies on correct specification of the regression of Y on A and W as the G-computation estimator, but instead relies on estimating the treatment mechanism or what the latter has been referred to as the propensity score: $P(A = a^* \mid X) = g(a^* \mid X)$. Specifically, these estimators are the solution corresponding with estimating functions:

$$D_h(0|g, \alpha) = \frac{h(V)I(A = a^*)}{g(a^* \mid W)}(Y - m(V \mid \alpha)),$$

where a sensible choice of h is:

$$h(V) = \frac{d}{d\beta}m(V \mid \alpha)g(a^* \mid V)E^{-1}(\epsilon^2(\alpha) \mid V).$$

Here, $\epsilon^2(\alpha)$ is the residual $Y - m(V \mid \alpha)$. This estimator can be conveniently implemented using existing regression techniques by simply supplying weights,

$$wt = \frac{I(A = a^*)}{g(a^* \mid W)}$$

to regressions of Y on V. One can estimate $g(a^* \mid W)$ with logistic regression of $I(A = a^*)$ on W.

Table 1: Assumptions underlying candidate estimators for $E[Y_{a^*} \mid V]$

Assumption	G-comp.	IPTW	IPCW	DR-IPTW	DR-IPTC
CA	X	X	X	X	X
RA	X	X	X	X	X
ETA		X	X	X	X
g correct	X	X	X		
Q correct	X				
g or Q correct			X	X	

Whereas the consistency of IPCW estimator relies on the consistency of the estimate of the treatment mechanism, g, there is an extension to this estimator that is consistent if either g *or* $Q(A, W) = E(Y \mid A, W)$ is consistently estimated, referred to as doubly-robust estimators (van der Laan and Robins (2002)).

In summary, the substitution estimators we propose use a two-step procedure 1) first use independent estimates of $E[Y_{a^*} \mid V]$ and $E[Y \mid V]$ and 2) regress the difference of these on V using the proposed intervention model. Of course, it is trivial to extend this general procedure to estimate other parameters (such as the ratio as opposed to the difference).

3.2 Direct Estimating Equation Approach

Note, if $m(V \mid \beta) = E(Y_{a^*} \mid V) - E(Y \mid V)$ then $E(Y_{a^*} \mid V) = m(V \mid \beta) + E(Y \mid V)$. Thus, we can use the estimating functions provided above for models $E(Y_{a^*} \mid V) = m(V \mid \beta) + E(Y \mid V)$. For instance, the IPCW-estimating function for $E(Y_{a^*} \mid V)$ is:

$$\begin{aligned} D_{h,IPCW}(0|g,\eta,\beta) &= \frac{h(V) * I(A = a^*)}{g(a^* \mid W)}(Y - E(Y_{a^*} \mid V)) \\ &= \frac{h(V) * I(A = a^*)}{g(a^* \mid W)}(Y - \eta(V) - m(V \mid \beta)). \qquad (3) \end{aligned}$$

The double robust estimating function extension to this estimator is given by (see van der Laan and Robins (2002), page 35):

$$D_{h,DR}(0|g,\eta,Q,\beta) = \frac{h(V)*I(A=a^*)}{g(a^* \mid W)}(Y-\eta(V)-m(V\mid\beta)) \\ - \frac{I(A=a^*)-g(a^*\mid W)}{g(a^*\mid W)}(h(V)(Q(W,A)-\eta(V)-m(V\mid\beta))),$$

where the estimator of β is consistent if either $Q(W,A) = E(Y \mid W,A)$ or $g(A \mid W)$ is consistently estimated.

These estimators rely on consistent estimation of $\eta(V)$, which can be problematic if V is high-dimensional and nonparametric estimation is infeasible. However, a small modification to both the $D_{h,IPCW}$ and $D_{h,DR}$ estimators yield estimates consistent *even when* $\eta(V)$ *is misspecified.* Specifically, we propose estimators based on estimating functions:

$$D_{h,\eta_1,IPCW}(0|g,\beta) = D_{h,IPCW}(0|g,\eta_1,\beta) + C_h(O \mid \eta_1,\beta) \quad (4)$$

$$D_{h,\eta_1,DR}(0|g,\beta) = D_{h,DR}(0|g,\eta_1,Q,\beta) + C_h(O \mid \eta_1,\beta) \quad (5)$$

where $C_h(O \mid \eta_1,\beta) = -Y * E(h(V)) + E(h(V)\eta_1(V))$ and η_1 may or may not be a consistent estimator of η_0, where the η_0 is based on the true data-generating distribution.

4 Estimating Intervention Effects over All Possible Interventions

The main thrust of the this paper is to propose a new class of estimating functions for the intervention models: the additive risk difference (AR), the relative risk (RR) and the quantile-quantile (QQ) function (2. For the quantile-quantile function, we focus on continuous outcome, Y, but discuss a model mapping the quantiles of $F_{Y|V}$ into $F_{Y_a|V}$ for binary Y in the appendix). In order to deal with the curse of dimensionality it will also be necessary to either assume a model on one of the factors $g_0(A \mid W)$ or $f_{Y|A,W}$ of the density of the observed data structure O. These parameters will appear as nuisance parameters in our proposed class of estimating equations for β_0. Specifically, if one uses our IPTW-type estimating functions, then one only needs to model g_0, while if one uses our DR-IPTW type estimating functions, then one needs to estimate g_0 and a functional of $f_{Y|A,W}$.

Given a model $\mathcal{G}$ for g_0 one can estimate it with a maximum likelihood estimator:

$$g_n \equiv \arg\max_{g \in \mathcal{G}} \prod_{i=1}^{n} g(A_i \mid W_i),$$

or versions of this involving regularization and/or model selection. Our estimating functions only depend on $f_{Y|A,W}$ through its mean $E(Y \mid A, W)$. Thus, in these cases one can use direct regression methods to estimate this nuisance parameter. The consistency of our proposed double robust estimators only rely on consistent estimation of *either* g_0 or $Q_0 \equiv E_0(Y \mid A, W)$ where the 0 subscript indicates it is based on the true data-generating distribution.

4.1 Estimation and Inference.

In each of the models we will apply the following general strategy for deriving a class of estimating functions. In order to illustrate this general approach we will use as example the additive risk intervention model.

- Given the population parameter η_0 the intervention model is using for a baseline comparison, the intervention model implies a marginal structural model for the conditional distribution of Y_a, given V. For example, for a given $\eta_0(V) = E(Y \mid V)$, the additive risk intervention model implies the marginal structural model $E(Y_a \mid V) = \eta_0(V) + m(a, V \mid \beta_0)$.

- Given η_0, this marginal structural model implies a class of IPTW and DR-IPTW estimating functions as established in previous literature. For example, the class of DR-IPTW estimating functions for $E(Y_a \mid V) = \eta_0(V) + m(a, V \mid \beta_0)$ treating η_0 as known are given by:

$$\begin{aligned} D_{h_1}(O \mid g, Q, \eta_0, \beta) \equiv & \frac{h_1(A,V)}{g(A \mid W)}(Y - \eta_0(V) - m(A, V \mid \beta)) \\ & - \frac{h_1(A,V)}{g(A \mid W)}(Q(A,W) - \eta_0(V) - m(A, V \mid \beta)) \\ & + \sum_{a \in \mathcal{A}} h_1(a,V)(Q(a,W) - \eta_0(V) - m(a, V \mid \beta)). \quad (6) \end{aligned}$$

The index h_1 can be an arbitrary function of A, V, and the nuisance parameters of the estimating functions are the treatment mechanism g_0 and $Q_0(A, W)$.

The first term of this estimating function represents the IPTW estimating function, and the last two terms are its projection onto the nuisance tangent space $T_{RA} = \{\phi(A, W) : E(\phi(A, W) \mid W) = 0\}$ of the treatment mechanism only assuming the randomization assumption.

These estimating functions are double robust w.r.t. misspecification of g_0 and Q_0 in the sense that if $\max_a \frac{h_1(a,V)}{g_1(a|W)} < \infty$ F_{W0}-a.e, and either $g_1 = g_0$ or $Q_1 = Q_0$, then

$$E_0 D_{h_1}(O \mid g_1, Q_1, \eta_0, \beta_0) = 0.$$

- Determine the influence curve of the estimator β_n solving $0 = \sum_i D_{h_1}(O_i \mid g_0, Q_0, \eta_n, \beta)$ for an appropriate estimator η_n of η_0 under a nonparametric model for η_0. Now, we treat this class of influence curves (ignoring standardizing constants/matrices) as a new class of estimating functions. This results in a corrected class of estimating functions

  $$D_{h_1}(O \mid g_0, Q_0, \eta_0, \beta_0) + C_{h_1}(O \mid \eta_0, \beta_0).$$

  The latter term actually corresponds with the influence curve of the "estimator" $E_0 D_{h_1}(O \mid g_0, Q_0, \eta_n, \beta_0)$ of the parameter $E_0 D_{h_1}(O \mid g_0, Q_0, \eta_0, \beta_0)$, treating the parameters other than η_0 as known. For example, in the additive risk intervention model we have

  $$C_{h_1}(O \mid \eta, \beta) = -\sum_{a \in \mathcal{A}} h_1(a, V) Y + E\left(\sum_{a \in \mathcal{A}} h_1(a, V) \eta(V)\right).$$

- Finally, we note that in each of our three intervention models we have that the corrected estimating functions remain unbiased if η_0 is misspecified. This allows us to treat η_0 as another index in the class of estimating functions, which results in the final class of proposed estimating functions indexed by functions h_1, h_2:

  $$D_{h_1,h_2}(O \mid g, Q, \beta) \equiv D_{h_1}(O \mid g, Q, h_2, \beta) + C_{h_1}(O \mid h_2, \beta).$$

The members in our class of estimating functions $\{D_{h_1,h_2} : h_1, h_2\}$ are double robust against misspecification of its two nuisance parameters g_0, Q_0 in the following sense: if $\max_a h_1(a, V)/g_1(a \mid W) < \infty$ F_{W0}-a.e, and either $g_1 = g_0$ or $Q_1 = Q_0$, then

$$E_0 D_{h_1,h_2}(O \mid g_1, Q_1, \beta_0) = 0.$$

For each of our models we propose a particular choice (h_1^*, h_2^*) depending on the true data generating distribution, which can be easily data adaptively estimated. Misspecification of this choice does not affect the unbiasedness of the estimating function, but only results in using a different (sub-optimal) unbiased estimating function. As a consequence, the misspecification of h_1^*, h_2^* does not affect the consistency and asymptotic linearity of the resulting estimator of β_0.

4.2 Class of estimating functions for additive risk intervention models

For a given $\eta_0(V) = E(Y \mid V)$, the additive risk intervention model implies the marginal structural model $E(Y_a \mid V) = \eta_0(V) + m(a, V \mid \beta_0)$. Our general strategy results now in the following class of estimating functions for β_0 indexed by arbitrary functions $h_1(A, V)$ and $h_2(V)$, and depending on two nuisance parameters g_0 and $Q_0(A, W)$:

$$\begin{aligned} D_{h_1,h_2}(O \mid g, Q, \beta) \equiv\ & \frac{h_1(A,V)}{g(A \mid W)}(Y - h_2(V) - m(A, V \mid \beta)) \\ & -\frac{h_1(A,V)}{g(A \mid W)}(Q(A, W) - h_2(V) - m(A, V \mid \beta)) \\ & +\sum_{a \in \mathcal{A}} h_1(a, V)(Q(a, W) - h_2(V) - m(a, V \mid \beta)) \\ - & \sum_{a \in \mathcal{A}} h_1(a, V) Y + E\left(\sum_{a \in \mathcal{A}} h_1(a, V) h_2(V)\right). \end{aligned} \quad (7)$$

Our proposed choice for (h_1, h_2) is given by:

$$\begin{aligned} h_1^*(A, V) &\equiv \frac{d}{d\beta} m(A, V \mid \beta) g_0(A \mid V) \\ h_2^*(V) &\equiv E(Y \mid V). \end{aligned}$$

4.3 Class of estimating functions for relative risk intervention models

For a given $\eta_0(V) = E(Y \mid V)$, the intervention relative risk model implies the marginal structural model $E(Y_a \mid V) = \eta_0(V)m(A, V \mid \beta_0)$. Our general strategy now results in the following class of estimating functions for β_0 indexed by arbitrary functions $h_1(A, V)$ and $h_2(V)$, and depending on two nuisance parameters g_0 and $Q_0(A, W) \equiv E_0(Y \mid A, W)$:

$$
\begin{aligned}
D_{h_1,h_2}(O \mid g, Q, \beta) \;\equiv\; & \frac{h_1(A,V)}{g(A \mid W)}(Y - h_2(V)m(A,V \mid \beta)) \\
& -\frac{h_1(A,V)}{g(A \mid W)}(Q(A,W) - h_2(V)m(A,V \mid \beta)) \\
& +\sum_{a \in \mathcal{A}} h_1(a,V)(Q(a,W) - h_2(V)m(a,V \mid \beta)) \\
& -\sum_{a \in \mathcal{A}} h_1(a,V)m(a,V \mid \beta)Y + E\left(\sum_{a \in \mathcal{A}} h_1(a,V)m(a,V \mid \beta)h_2(V)\right).
\end{aligned}
$$

Our proposed choice for (h_1, h_2) is given by:

$$
\begin{aligned}
h_1^*(A,V) &\equiv h_2^*(V)\frac{d}{d\beta}m(A,V \mid \beta)g_0(A \mid V) \\
h_2^*(V) &\equiv E(Y \mid V).
\end{aligned}
$$

4.4 General intervention mean models

We now consider a general mean intervention model corresponding with a general choice of $\Phi(E(Y_a \mid V), E(Y \mid V)) = m(a, V \mid \beta)$. Given $\eta_0(V) = E(Y \mid V)$, this model results in a marginal structural model $E(Y_a \mid V) = f(\eta_0(V), m(a, V \mid \beta))$ for some bivariate real valued function f. Our general strategy now results in the following class of estimating functions for β_0:

$$\begin{aligned}
D_{h_1}(O \mid g, Q, \eta, \beta) \equiv & \frac{h_1(A,V)}{g(A \mid W)}(Y - f(\eta(V), m(A,V \mid \beta)) \\
& - \frac{h_1(A,V)}{g(A \mid W)}(Q(A,W) - f(\eta(V), m(A,V \mid \beta)) \\
& + \sum_{a \in \mathcal{A}} h_1(a,V)(Q(a,W) - f(\eta(V), m(a,V \mid \beta)) \\
& - \sum_{a \in \mathcal{A}} h_1(a,V) f^{10}(\eta(V), m(a,V \mid \beta))Y \\
& + E\left(\sum_{a \in \mathcal{A}} h_1(a,V) f^{10}(\eta(V), m(a,V \mid \beta)\eta(V)\right),
\end{aligned}$$

where $f^{10}(x,y) \equiv d/dx f(x,y)$.

However, it follows that $E_0 D_{h_1}(O \mid g_0, Q_0, \eta_1, \beta_0)$ only equals zero at a misspecified η_1 if $x \rightarrow f(x,y)$ is linear, which precisely corresponds with the additive and relative risk models. For example, if one models the log odds in the binary outcome case, then this class of estimating functions does rely on consistent estimation of the nuisance parameter $\eta_0 = E(Y \mid V)$, beyond consistent estimation of either g_0 or Q_0.

4.5 Class of estimating functions for quantile-quantile function

Let $\eta_0 = F_{Y|V}$ be the cumulative distribution function of Y, given V. Given η_0, the quantile-quantile intervention model implies a marginal structural model $F_{Y_a|V} = \eta_{0,Y|V} m^{-1}(\cdot \mid a, V, \beta)$, where $m^{-1}(q \mid a, V, \beta)$ is the inverse of the function $q \rightarrow m(q, a, V \mid \beta)$. Thus $m^{-1}(q \mid a, V, \beta_0) = F^{-1}_{Y|V} F_{Y_a|V}(q)$ is the inverse of the modeled quantile-quantile function. Denote this model for the conditional distribution $F_{Y_a|V}$ with $F_{a,\beta,\eta}(\cdot \mid V)$.

The orthogonal complement of the nuisance tangent space in the full data model for this marginal structural model is given by:

$$T^{F,\perp}_{nuis} = \left\{ \sum_a \Phi(a, Y_a, V) - \int \Phi(a, y, V) dF_{a,\beta_0,\eta_0}(y \mid V) : \Phi(a, \cdot, \cdot) \in L^2_0(F_{Y_a,V}) \right\}.$$

We now note that by the transformation rule for integration

$$\int_y h(a, y, V) dFa, \beta, \eta(y \mid V) = \int_y h(a, m(y \mid a, V, \beta), V) d\eta(y \mid V).$$

Consequently, the class of IPTW-estimating functions for the marginal structural model with η_0 known are given by

$$\left\{ \frac{\Phi(A, Y, V) - \int_y \Phi(A, m(y \mid A, V, \beta), V) d\eta_0(y \mid V)}{g(A \mid W)} : \Phi \right\}. \quad (8)$$

Thus, the class of DR-IPTW-estimating functions are given by

$$\begin{aligned} D_h(O \mid g, Q, \beta, \eta) \quad &= \quad \frac{h(A, Y, V) - \int h(A, m(y \mid A, V, \beta), V) d\eta(y \mid V)}{g(A \mid W)} \\ -\frac{E_Q(h(A, Y, V) \mid A, W) - \int_y h(A, m(y \mid A, V, \beta), V) d\eta(y \mid V)}{g(A \mid W)} \\ +\sum_{a \in \mathcal{A}} E_Q(h(A, Y, V) \mid A = a, W) - \int_y h(a, m(y \mid a, V, \beta), V) d\eta(y \mid V), \quad (9) \end{aligned}$$

where the nuisance parameter Q_0 denotes the true conditional distribution $F_{Y|A,W}$.

The correction term one must add to (9) to orthogonalize these estimating functions w.r.t. to η_0 is given by

$$-\sum_a h(a, m(Y \mid a, V, \beta), V) + E\left(\sum_a \int_y h(a, m(y \mid a, V, \beta), V) d\eta(y \mid V) \right).$$

Finally, by noting that with this correction the unbiasedness of the estimating functions are fully protected against misspecification of η_0, we obtain the following class of estimating functions for our quantile-quantile intervention model:

$$
\begin{aligned}
D_{h_1,h_2}(O \mid g, Q, \beta) &\equiv \frac{h_1(A, Y, V) - \int_y h_1(A, m(y \mid A, V, \beta), V) dh_2(y \mid V)}{g(A \mid W)} \\
&- \frac{E_Q(h_1(A, Y, V) \mid A, W) - \int_y h_1(A, m(y \mid A, V, \beta), V) dh_2(y \mid V)}{g(A \mid W)} \\
&+ \sum_{a \in \mathcal{A}} E_Q(h_1(A, Y, V) \mid A = a, W) - \int_y h_1(a, m(y \mid a, V, \beta), V) dh_2(y \mid V) \\
&- \sum_a h_1(a, m(Y \mid a, V, \beta), V) + E\left(\sum_a \int_y h_1(a, m(y \mid a, V, \beta), V) dh_2(y \mid V)\right).
\end{aligned} \tag{10}
$$

4.6 Estimation and Asymptotic Inference

Given estimators h_{1n}, h_{2n}, g_n, Q_n of h_1, h_2, g_0, Q_0, let β_n be the solution of the estimating equation

$$0 = \sum_{i=1}^{n} D_{h_{1n},h_{2n}}(O_i \mid g_n, Q_n, \beta).$$

One can use the Newton-Raphson algorithm for solving this estimating equation, and, if one has a good initial estimator available, then this estimator β_n will be asymptotically equivalent with the one-step estimator as obtained in the first step of the Newton-Raphson algorithm:

$$\beta_n^1 = \beta_n^0 - c_n^{-1} \frac{1}{n} \sum_{i=1}^{n} D_{h_{1n},h_{2n}}(O_i \mid g_n, Q_n, \beta_n^0),$$

where

$$c_n = \frac{d}{d\beta_n^0} \frac{1}{n} \sum_{i=1}^{n} D_{h_{1n},h_{2n}}(O_i \mid g_n, Q_n, \beta_n^0).$$

Statistical inference for β_0 can now be based on $\beta_n \sim N(\beta_0, \Sigma_n)$, where

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^{n} \hat{IC}(O_i) \hat{IC}(O_i)^\top$$

and $\hat{IC}(O)$ is the estimate of the influence curve $IC(O \mid P_0)$ of our estimator (the derivation of the influence curve is provided in the appendix). In general, one can avoid deriving the influence curve of the estimator by using the bootstrap to estimate the distribution of $\sqrt{n}(\beta_n - \beta_0)$.

5 Simulation Study

To examine the finite sample properties of our estimator and the impact of the misspecification of $E(Y|V)$, we performed a simulation study of the additive risk difference. Using the syntax introduced in section 2, the observed data is $O = (W = (V, Z), A, Y)$ and we are interested in estimating $m(a, V \mid \beta) = E[Y|V] - E[Y_a|V]$. In this simulation, V and $Z \sim N(0, \sigma = 0.5)$, $A \in \{0,1\}$, $logit(P(A = 1 \mid W)) = Z - V + 0.5ZV$ and finally, $Y \in \{0,1\})$, $logit(P(Y = 1 \mid W, A)) = -2 + Z + V^2 + 0.5A + 0.25AV^2$, which results in $m(a, V \mid \beta) = 0.051 - 0.091a - 0.032V - 0.0026V^2 + 0.0027aV$. For all these simulations, the functional form of $m(a, V \mid \beta)$ is correct, i.e., $m(a, V \mid \beta) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 V^2 + \beta_4 aV$. The relative performance of two estimators are examined: 1) an IPTW estimator without the correction factor $(C_{h_1}(O \mid \eta_0, \beta_0))$, which is the solution to the estimating equation based upon the estimating function defined as the first line of (6), and 2) the DR-IPTW estimator defined by the solution of estimating equations based on (7). For the both the IPTW and the DR-IPTW, we correctly estimate the treatment mechanism g (using the correct logistic form) and for the DR-IPTW estimator we also estimate $Q(A, W) = E(Y \mid A, W)$ consistently. We misspecifiy the functional form of $E[Y \mid V]$ for both the IPTW and DR-IPTW estimators. One would expect the IPTW estimator to be biased, due to misspecification of $\eta_0(V)$, whereas the DR-IPTW will remain consistent. We compare the two estimators at sample sizes of both $n = 100$ and $n = 1000$ and report our results with regards to bias, variance and mean- square error as well as the relative efficienty with respect to the IPTW-estimator (e.g., MSE(IPTW)/MSE(DR-IPTW)).

The results (table 2) suggests a modest gain in the relative efficiency for a sample size of 100, but as the sample size increases (and the bias of the IPTW-estimator does not decrease), we can see the virtue of using the double-robust estimators, both for reductions in bias (robustness to misspecification of $\eta(V)$) and variance. In the appendix, R code (Ihaka and Gentleman (1996)) is provided to calculate the objective function for the DR-IPTW

Table 2: Squared bias, Variance (normalized by n) and relative mean-squared error (RMSE) of IPTW and DR-IPTW estimators for $n = 100$ and $n = 1000$.

Est.		β_0	β_1	β_2	β_3	β_4
n=100						
IPTW	$Bias^2$	.22	< 0.001	< 0.001	3.57	< 0.001
	Var.	.39	.77	1.52	2.59	4.37
DR	$Bias^2$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Var.	0.27	0.79	1.45	0.94	4.57
	RMSE	2.3	0.98	1.05	6.53	0.95
n=1000						
IPTW	$Bias^2$	2.7	.0016	.0011	41.7	.0014
	Var.	.37	.71	1.39	2.62	3.99
DR	$Bias^2$	.0013	< 0.001	.0030	< 0.001	.0017
	Var.	.25	.71	1.22	.81	4.01
	RMSE	12.32	1.00	1.14	54.4	1.00

estimator, which can be used in a general multi-dimensional minimization routine, such as *optim* to estimate the intervention model of interest.

We also examined the simple IPTW version of the quantile-quantile estimator (8). For this simulation, the parameter of interest was a model, m, of the quantile-quantile function $F_{Y_a}^{-1}F_Y(q) = m(q \mid a, \beta_0)$ (not stratifying by baseline covariate, V). The data distribution used can be described in the following manner for continuous W, Y and binary $A = (0, 1)$:

$$\begin{aligned} W &\sim N(0, \sigma = 0.5) \\ logit(P(A = 1 \mid W)) &= 3 * W \\ Z &\sim exponential(\lambda(A, W)) \\ log(\lambda(A, W)) &= -2 + W + 3 * A, \end{aligned}$$

where the outcome of interest was $Y = log(Z)$. As opposed to determining the true quantile-quantile function analytically, we generated trios of Y_0, Y_1, Y and modeled the difference of the quantiles between Y_a and Y as a function of Y and a. That is, the simulations examined the performance of the simple IPTW quantile-quantile estimator of a model relative to the asymptotic fit based on a full data model where all counterfactuals as well as the treatment assignment are observed. In this case, although not the true model, assuming a simple linear model,

$$F_{Y_a}^{-1} F_Y(q) = m(q \mid a, \beta_0) = \beta_0 + \beta_1 q + \beta_2 a + \beta_3 q a \tag{11}$$

results in a reasonably good prediction of the quantiles of Y_a from Y. The IPTW-estimator was the solution to the following estimating function:

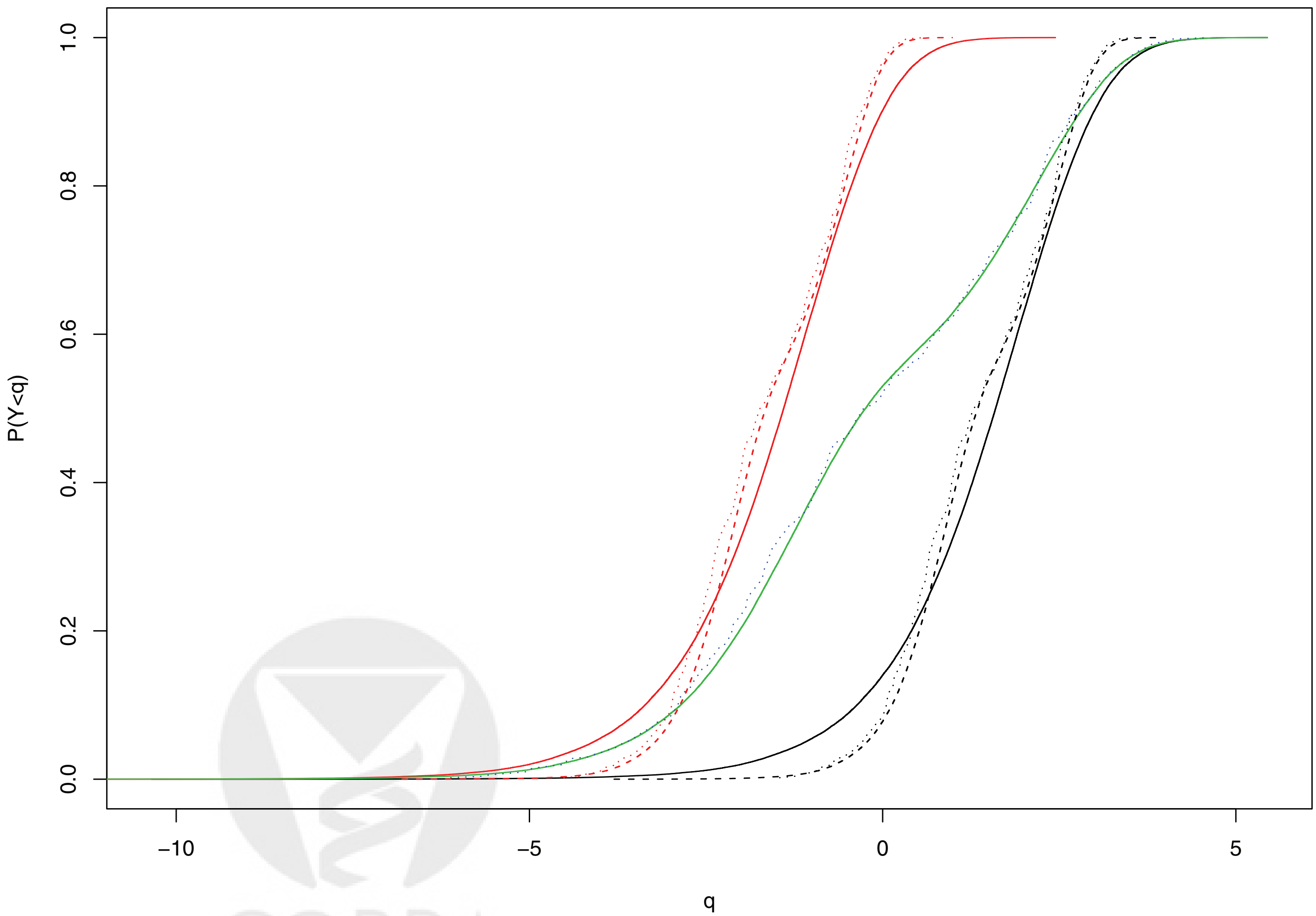
$$\frac{\Phi(A, Y) - \int_q \Phi(A, m(a \mid A, \beta)) d\eta_0(q)}{g(A \mid W)},$$

where $\Phi(A, Y) = \frac{d}{d\beta} m(Y \mid A, \beta)$. With a sample size of 500, using the empirical to estimate F_Y and using the same model for $m(q \mid a, \beta)$ (11), the simulations demonstrate that the IPTW estimators results in a close fit to the optimal fit (the projection of the true distributions onto that implied by (11). Figure 1 has a) the true distribution of Y, Y_0, Y_1 (solid lines), b) the projection of the true Y_a onto the estimating model using (11) (dashed lines) and c) the IPTW-estimate transforming the empirical distribution of Y on Y_a using the estimates of β and mapping from the empirical distribution of Y (dotted lines). As expected, the IPTW estimates (1) are converging to the optimal fit (hard to distinguish dotted and dashed lines).

6 Discussion

The set of intervention parameters and estimators we propose continue a general estimating equation approach for parameters in censored data models, such as the missing data models one uses in causal inference, originally developed by Robins and Rotniksky (Robins and Rotnitzky (1992), Robins (2000a), Robins (2000b)) and further formalized and generalized in van der Laan and Robins (2002). The doubly-robust estimators we propose for the additive risk, relative risk and quantile-quantile function are consistent if either the treatment mechanism, $g(A \mid X)$, or $E(Y \mid A, W)$ is correctly specified; the estimators are consistent regardless of the model for $Y \mid V$. As the simulations show, the doubly robust estimators provide some bias protection as well as greater efficiency. For many contexts, the relevant parameter of interest is not comparing populations exposed to different levels of the risk factor, but the impact on the disease distribution of intervening in the population by eliminating the risk factor. For some risk factors, the target level is obvious (i.e., it would make no sense to examine the impact

Figure 1: True distributions of Y_1, Y and Y_0 (top, middle and bottom solid lines), best fit using the linear model described in the text(dashed lines) and the IPTW-estimator (dotted lines)



of having everyone in a population smoke 3 cigarettes a day, as that would require getting many people to start smoking). However for others, such as exposure to air pollution, examining the impact for different target levels makes more sense, and so we propose models for parameters, such as $m(a, V \mid \beta) = E[Y_a - Y \mid V]$. We have also provided estimators that can use existing methods, such as the likelihood-based G-computation formula, as well as the IPCW estimator.

Although not presented here, the intervention models approach can easily be extended to time-dependent treatments, where the data structure includes both treatments $A(j)$ and confounders $L(j)$ measured at times, $j = 1, \ldots, p$, and a final outcome Y. In this case, the counterfactuals of interest, $Y_{\bar{a}}$ are indexed by an entire treatment history, $\bar{a} = (a(0), a(1), \ldots, a(j))$. As outlined above, if $\eta_0(V)$ is known or can be estimated non-parametrically, we can use the existing estimating equations developed for estimating the mean of $E[Y_{\bar{a}} \mid V]$. Specifically, using the doubly-robust estimating equations presented by van der Laan and Robins (2003) for marginal structural models for time-dependent treatments, we can simply plug in $\eta_0(V) + m(\bar{a}, V \mid \beta)$ for $E[Y_{\bar{a}} \mid V]$ where $m(\bar{a}, V \mid \beta) = E[Y_{\bar{a}} \mid V] - E[Y \mid V]$. As we have done in this paper, these estimating equations can also be altered for robustness against misspecification of $\eta_0(V)$.

APPENDIX

Quantile-quantile function

Consider the following parameter of interest, which maps the quantiles of the distribution of the observed $Y \mid V$ into quantiles of a particular counterfactual (e.g., all subjects given treatment a) distribution:

$$\psi_{0,QQ}(a,V)(q) = F^{-1}_{Y_a|V} F_{Y|V}(q) = m(q \mid a, v, \beta_0). \tag{12}$$

This model requires that Y is a continuous outcome and maps any p-th quantile of $F_{Y|V}$ into the difference between this p-th quantile and the p-th quantile of $F_{Y_a|V}$. In the case that Y is binary we have shown (van der Laan et al. (2005)) that the analogue of the latter parameter is given by the so called switch relative risk $\psi_0 = (\theta_0, \mathcal{A}_0)$, where

$$\theta_0(a,V) \equiv \frac{E(Y_a \mid V)}{E(Y \mid V)} I_{\mathcal{A}_0}(a,V) + \frac{1 - E(Y_a \mid V)}{1 - E(Y \mid V)} I_{\mathcal{A}_0^c}(a,V),$$

and

$$\mathcal{A}_0 \equiv \{(a,V) : E(Y_a \mid V)/E(Y \mid V) \leq 1\}.$$

Modelling approaches for this parameter are presented in detail in van der Laan et al. (2005), and they result in a model

$$(\theta_0(a,V) = \theta(a,V \mid \beta_0), \mathcal{A}_0 = \mathcal{A}(\beta_0)),$$

where $\theta(a,V \mid \beta)$ and $\mathcal{A}(\beta)$ are parametrizations of θ_0 and $\mathcal{A}_0$ in terms of a common Euclidean parameter β. Note that this switch relative risk identifies the set $\mathcal{A}_0$ interventions and corresponding baseline covariate values v for which the intervention is protective, the relative increase in risk of Y for this set, and the relative increase in risk of $1 - Y$ for values of a, V for which intervention is harmful. Off course, if intervention is always harmful, then the switch relative risk reduces to the relative risk. For a more detailed discussion on the switch relative risk we refer to van der Laan et al. (2005).

Class of estimating functions for switch relative risk intervention models: binary outcome

For a binary Y, define

$$\begin{aligned} m^*(a,V \mid \beta,\eta) &\equiv \eta(V)\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)}(a,V) \\ &+ (1-(1-\eta(V))\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)^c}(a,V). \end{aligned}$$

This represents the marginal structural model $E(Y_a \mid V) = m^*(a,V \mid \beta)$ for given η implied by the model (13) on the switch relative risk ψ_{0SRR}.

We propose the following class of estimating functions for β_0 indexed by arbitrary functions $h_1(A,V)$ and $h_2(V)$, and depending on two nuisance parameters g_0 and $Q_0(A,W) \equiv E_0(Y \mid A,W)$:

$$\begin{aligned} D_{h_1,h_2}(O \mid g,Q,\beta) \equiv & \frac{h_1(A,V)}{g(A \mid W)}(Y - m^*(A,V \mid \beta,h_2) - \\ & \frac{h_1(A,V)}{g(A \mid W)}(Q(A,W) - m^*(A,V \mid \beta,h_2(V))) \\ & + \sum_{a\in\mathcal{A}} h_1(a,V)(Q(a,W) - m^*(a,V \mid \beta,h_2(V))) \\ & - \sum_{a\in\mathcal{A}} h_1(a,V)\left\{Y\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)}(a,V) + (1-Y)\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)^c}(a,v)\right\} \\ & + E\left(\sum_{a\in\mathcal{A}} h_1(a,V)\left\{h_2(V)\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)}(a,V) + (1-h_2(V))\theta(a,V \mid \beta)I_{\mathcal{A}(\beta)^c}(a,v)\right\}\right). \end{aligned}$$

Our proposed choice for (h_1,h_2) is given by:

$$\begin{aligned} h_1^*(A,V) &\equiv \frac{d}{d\beta} m*(A,V \mid \beta,h_2^*)g_0(A \mid V) \\ h_2^*(V) &\equiv E(Y \mid V). \end{aligned}$$

Estimating the Influence Curve for Asymptotic inference.

Our estimator β_n will be asymptotically consistent if either g_n is consistent for g_0 or Q_n is consistent for Q_0. Regarding statistical inference, we will first consider the case in which one is willing to assume that g_n consistently estimates g_0. If g_n converges to g_0 for $n \to \infty$, then it is straightforward to show that, under regularity conditions, β_n is asymptotically linear with influence curve

$$\begin{aligned} IC(O_i \mid P_0) &= -c_0^{-1} \{D_{h_1,h_2}(O \mid g_0, Q_1, \beta_0) + IC_{nuis}(O \mid P_0)\} \\ &\quad -c_0^{-1} \left\{ \sum_a h_1(a,V) h_2(V) - E(\sum_a h_1(a,V) h_2(V)) \right\}, \end{aligned}$$

where IC_{nuis} is the influence curve of $\Phi(G_n)$ as an estimator of $\Phi(G) \equiv P_{F_{X0},G} \frac{h_1(A,V)}{g_0(A|W)}(Q_1(A,W) - Y)$, and

$$c_0 \equiv \frac{d}{d\beta_0} E_0 D_{h_1,h_2}(O \mid g_0, Q_1, \beta_0).$$

In the special case that $h_2 = h_2^*$, then it can be shown that

$$IC_{nuis} = -\Pi(D_{h_1}(\cdot \mid Q_1, g_0, \eta_0, \beta_0) \mid T_2(P_0)),$$

where $D_{h_1}(O \mid Q_1, g_0, \eta_0, \beta_0)$ is the double robust estimating function for the standard MSM corresponding with the intervention model with η_0 being known (see (6) for its definition in the additive risk intervention model). $T_2(P_0) \subset \{\phi(A,W) : E(\phi(A,W) \mid W) = 0\}$ is the tangent space of the model $\mathcal{G}$ for g_0 at P_0, and $\Pi(\cdot \mid T_2(P_0))$ is the projection operator onto this subspace $T_2(P_0)$ within the Hilbert space $L_0^2(P_0)$.

In general, it seems to be reasonable approach to use $IC(O \mid P_0)$ with this particular choice of IC_{nuis} as an influence curve of our double robust estimator β_n. It can also be expected that the contribution of IC_{nuis} in this influence curve decreases the variance of the influence curve (since it subtracts a $T_2(P_0)$ component of $D_{h_1}(O \mid Q_1, g_0, \eta_0, \beta_0)$). As a consequence, a typically conservative influence curve is given by

$$\begin{aligned} IC(O_i \mid P_0) &= -c_0^{-1} \left\{ D_{h_1,h_2}(O \mid g_0, Q_1, \beta_0) + \sum_a h_1(a,V) h_2(V) \right. \\ &\quad \left. - \; E\left(\sum_a h_1(a,V) h_2(V) \right) \right\}, \end{aligned}$$

which corresponds with the influence curve one would obtain if one uses $g_n = g_0$, as if g_0 is known.

Statistical inference for β_0 can now be based on $\beta_n \sim N(\beta_0, \Sigma_n)$, where

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} \hat{IC}(O_i)\hat{IC}(O_i)^\top$$

and $\hat{IC}(O)$ is the estimate of the influence curve $IC(O \mid P_0)$ obtained by plugging in our estimates of g_0, Q_0, h_1^*, h_2^* and β_0. As mentioned in section 4, one can avoid deriving the influence curve of the estimator by using the bootstrap to estimate the distribution of $\sqrt{n}(\beta_n - \beta_0)$.

R-Code for DR-IPTW estimator for Additive Risk (7)

The R-function below is just example code implemented for a specific choice for $m(a, V|\beta) = E[Y|V] - E[Y_a|V]$, for a continuous V and W, and for a binary A and Y. The function returns the Euclidean norm of the sum of the estimating equation and can be used with the quasi-Newton multi-dimensional optimization R-function *optim* (R Development Core Team (2005)) to derive the estimate that solves this estimating equation. It can easily be generalized to for any choice of $m(a, V|\beta)$ and different data structures.

```
driptw<-function(beta,A,W,V,Y) {
# Makes design matrix dm
ta<-table(A)
a<-as.numeric(names(ta))
na<-length(a)
V2<-V^2
AV<-A*V
dm<-cbind(1,A,V,V2,AV)
# PIM model
mav<-beta[1]+beta[2]*A+beta[3]*V+beta[4]*V2+beta[5]*AV
# Estimates treatment model
glm.A<-glm(A~W*V,family=binomial)
pred.A<-predict(glm.A,type="response")
### Estimates E[Y|V]
glm.YV<-glm(Y~V,family=binomial)
pred.YV<-predict(glm.YV,type="response")
```

```
### Estimates E[Y|W,V,A]
glm.YAW<-glm(Y~W+V+V2+A+AV,family=binomial)
pred.YAW<-predict(glm.YAW,type="response")
### Weights
wt.A<-A/pred.A+(1-A)/(1-pred.A)
### Components of Estimating Equation
term1<-dm*(wt.A*(Y-pred.YV+mav))
term2<-dm*(wt.A*(pred.YAW-pred.YV+mav))
term3<-matrix(0,dim(term2)[1],dim(term2)[2])
term4<-matrix(0,dim(term2)[1],dim(term2)[2])
term5<-matrix(0,dim(term2)[1],dim(term2)[2])
for(i in 1:na) {
 aa<-rep(a[i],length(V))
 aV<-aa*V
 dm2<-cbind(1,aa,V,V2,aV)
 mav2<-beta[1]+beta[2]*aa+beta[3]*V+beta[4]*V2+beta[5]*aV
 pred.YaW<-predict(glm.YAW,type="response",newdata=
   data.frame(W=W,V=V,V2=V2,A=aa,AV=aV))
 term3<-term3+dm2*(pred.YaW-pred.YV+mav2)
 term4<-term4+dm2*Y
 term5<-term5+dm2*pred.YV
}
sumt5<-t(matrix(rep(apply(term5,2,mean),length(V)),
     dim(term2)[2],dim(term2)[1]))
total<-term1-term2+term3-term4+sumt5
norm.crt<-apply(total,2,sum)
#Returns Euclidean norm of sum of estimating equation
sum(norm.crt^2) }
```

References

Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

James M. Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials (Minneapolis, MN, 1997)*, pages 95–133. Springer, New York, 2000a.

J.M. Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach in causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley, editors, *Health Service Methodology: A Focus on AIDS*, pages 113–159. U.S. Public Health Service, National Center for Health SErvices Research, Washington D.C., 1989.

J.M. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23:2379–2412, 1994.

J.M. Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association*, 2000b.

J.M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*, Methodological issues. Bikhäuser, 1992.

J.M. Robins and A. Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *JASA*, 90:122–129, 1995.

M. J. van der Laan, A.E. Hubbard, and N.P. Jewell. Estimation of treatment effects in randomized trials with noncompliance and a dichotomous outcome. Technical Report 157, Division of Biostatistics, University of California, Berkeley, 2005.

M.J. van der Laan and J.M. Robins. *Unified methods for censored longitudinal data and causality.* Springer, New York, 2002.