



Published in final edited form as:  
*Biometrika*. 2008 ; 95(1): 35–47.

## Population intervention models in causal inference

ALAN E. HUBBARD and MARK J. VAN DER LAAN

Division of Biostatistics, University of California, Berkeley, California 94720, U.S.A.  
hubbard@stat.berkeley.edu laan@stat.berkeley.edu

### Summary

We propose a new causal parameter, which is a natural extension of existing approaches to causal inference such as marginal structural models. Modelling approaches are proposed for the difference between a treatment-specific counterfactual population distribution and the actual population distribution of an outcome in 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 population intervention models. We focus on intervention models estimating the effect of an intervention in terms of a difference and ratio of means, called risk difference and relative risk if the outcome is binary. We provide a class of inverse-probability-of-treatment-weighted and doubly-robust estimators of the causal parameters in these models. The finite-sample performance of these new estimators is explored in a simulation study.

### Keywords

Attributable risk; Causal inference; Confounding; Counterfactual; Doubly-robust estimation;  $G$ -computation estimation; Inverse-probability-of-treatment-weighted estimation

## 1. Introduction

Robins (2000) proposed estimators for a new class of models that relate the distribution of counterfactuals in a population to the universal application of, or exposure to, a treatment or a risk factor. For example, Robins proposed estimators 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,  $V$ , such as gender. Several estimators have been proposed for this parameter, including the likelihood-based  $G$ -computation estimator, and those based on an estimating equation approach, namely the inverse-probability-of-treatment-weighted estimator and the doubly-robust extension (van der Laan & 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 respectively an estimator of the conditional distribution of the outcome given the treatment and confounders, an estimator of the treatment assignment distribution, or both. 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.

The estimating equation approach provides a convenient mathematical method for deriving a class of estimators for a wide variety of causal inference problems. However, before applying the machinery (van der Laan & 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. For instance, marginal structural models compare the prevalence of disease in a population in which everyone smokes three packs of cigarettes against the prevalence in the same population in which everyone smokes two packs and so on. 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, 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, such as  $E(Y_a) - E(Y)$  or  $E(Y_a)/E(Y)$ , where  $a$  is a target level of the risk factor. Epidemiologists 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.

## 2. Data, model and parameters of interest

To define our model, we use the causal inference framework originally proposed by Rubin (Rubin, 1978) treating causal inference as a missing-data problem. 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  $O_1, \dots, O_n$  represents  $n$  independent and identically distributed observations of  $O$ . We are concerned with estimation and testing of the effect of an intervention,  $A = a$ , 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  $(Y_a : a \in A)$ , so that  $X = \{(Y_a : a \in A), W\}$  represents a full data random variable with a population distribution  $F_{X0}$ . Here  $A$  denotes the set of possible treatments. Secondly, one assumes that the observed data,  $O = (W, A, Y = Y_A)$ , correspond to observing the  $A$ -specific component of the collection of treatment-specific data structures  $X = \{X_a \equiv (Y_a, W) : a \in A\}$  corresponding to the treatment  $A$  the subject actually received; that is, treatment assignment serves as a censoring variable. Here  $A$  is short-hand for a vector of censoring indicators, where the 0's and a single 1 correspond to unobserved counterfactuals,  $X_a, a \neq A$ , and the observed counterfactual,  $X_A$ , respectively. This assumption was originally introduced by Rubin (1986) as the stable-unit-treatment-value assumption and has also been referred to as the consistency assumption. Since  $O$  is 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  $X$ ; that is  $O \sim P_{F_{X0}, g_0}$ , where  $g_0(a | X) \equiv \text{pr}(A = a | X)$  is referred to as the treatment mechanism. In order to have an identifiable causal parameter we will also rely on the so-called randomization assumption on  $g_0$ :

$$g_0(a|X) = g_0(a|W),$$

or, equivalently,  $A$  is conditionally independent of  $X$ , given  $W$ . Within Rubin's framework, this assumption is known as strong ignorability; it is also often referred to as the assumption of no unmeasured confounding. Under these assumptions, the density of the observed data structure with respect to an appropriate dominating measure factorizes as follows:

$$p_0(O) = f_w(W) f_{Y|A,W}(Y|A,W) g_0(A|W).$$

If  $g_0(a | W) > 0, F_W$  almost everywhere, then it also follows that the first two factors of the observed data density identify the counterfactual distribution of  $(Y_a, W)$ :

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

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

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

$$\psi_0(a, V) = \Psi(F_{x_0, g_0})(a, V) \equiv \Phi(F_{y_a|V}, F_{y|V}),$$

for some known functional  $\Phi$ . In words,  $\psi_0$  measures the effect of setting  $A = a$  for everybody in our population on the distribution of  $Y$ , within stratum  $V = v$ . A  $\Phi$ -specific intervention model is now defined as a model on this parameter  $\psi_0$ :

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

for some Euclidean parameterization  $\beta \rightarrow m(a, V | \beta)$ . The following sections present estimators for two classes of parameters, namely the additive risk and relative risk:

$$\begin{aligned} \psi_{0,AR}(a, V) &= E(Y_a | V) - E(Y | V) = m(a, V | \beta_0), \\ \psi_{0,RR}(a, V) &= \frac{E(Y_a | V)}{E(Y | V)} = m(a, V | \beta_0). \end{aligned} \tag{2}$$

Note that the structural nested mean models of Robins (1989, 1994) provide parameters that compare the mean of an outcome, within strata of  $A$  and  $W$ , under the ‘natural’ population distribution of treatment versus that under an intervention when  $A = 0$  for all subjects, that is,  $E(Y | A, W) - E(Y_0 | A, W)$ . Thus, the parameters discussed here capture a different effect, specifically not conditioning on  $A$ . Also, as discussed in the Appendix of van der Laan et al. (2007), 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 for the case of only one target intervention of interest,  $a^*$ . For instance, if  $A$  measures cigarette smoking, we would only be interested in the intervention  $a^* = 0$ . We describe a direct estimating equation approach for the additive risk that enjoys the very attractive property of being consistent even when the model for  $E(Y | V)$  is misspecified. Note that, if  $m(V | \beta) = E(Y_{a^*} | V) - E(Y | V)$ , then  $E(Y_{a^*} | V) = m(V | \beta) + E(Y | V)$  and thus, treating  $E(Y | V)$  as known, we can use the estimating functions already provided by Robins & Rotnitzky (1992) for models  $E(Y_{a^*} | V) = m(V | \beta) + E(Y | V)$ . For instance, the inverse-probability-of-censoring-weighted estimating function for  $E(Y_{a^*} | V)$  is

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

where  $\eta(V) \equiv E(Y | V)$  and  $h(V)$  is some arbitrary function of  $V$ . The estimator, defined as the solution of the estimating equation  $0 = \sum_{i=1}^n D_{h,IPCW}(0 | g_n, \eta_n, \beta)$ , is consistent if  $g_n(A | W)$  consistently estimates  $g_0(A | W)$  and  $\eta_n(V)$  consistently estimates  $\eta_0(V)$ , where the subscript, 0, always refers to the true parameter. The doubly-robust estimating function extension to this estimator is given by

$$\begin{aligned} D_{h,DR}(0 | g, \eta, Q, \beta) &= \frac{h(V)I(A=a^*)}{g(a^*|W)} \{Y - \eta(V) - m(V | \beta)\} \\ &\quad - \frac{I(A=a^*) - g(a^*|W)}{g(a^*|W)} h(V) \{Q(W, A) - \eta(V) - m(V | \beta)\} \end{aligned} \tag{4}$$

(van der Laan & Robins, 2002, p. 35), where  $Q(A, W) \equiv E(Y | A, W)$ . The doubly-robust estimator, defined as the solution of the estimating equation  $0 = \sum_{i=1}^n D_{h,DR}(0 | g_n, \eta_n, Q_n, \beta)$ , is consistent if  $\eta_n(V)$  consistently estimates  $\eta_0(V)$  and either  $Q_n(W, A)$  consistently estimates  $Q_0(A, W)$  or  $g_n(A | W)$  consistently estimates  $g_0(A | W)$ .

Both of these estimators thus 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  $D_{h,DR}$  estimating functions yields consistent estimators even when  $\eta(V)$  is misspecified. To be specific, we propose estimators based on the following estimating functions, now indexed by  $h$  and  $\eta_1$ :

$$D_{h,\eta_1,IPCW}(0|g,\beta) = D_{h,IPCW}(0|g,\eta_1,\beta) + C_h(O|\eta_1,\beta) \tag{5}$$

$$D_{h,\eta_1,DR}(0|g,Q,\beta) = D_{h,DR}(0|g,\eta_1,Q,\beta) + C_h(O|\eta_1,\beta), \tag{6}$$

where  $C_h(O | \eta_1, \beta) = -Yh(V) + E\{h(V)\eta_1(V)\}$  and  $\eta_1$  represents an index of the estimating function with the optimal choice being  $\eta_0 = E(Y | V)$ . In the estimating equation  $\eta_1$  can now be replaced by an estimator of  $\eta_0$ , which is not necessarily consistent, and  $C_h$  can be obtained by finding the expected values of the original estimating functions  $D_{h,\eta_1,IPCW}(0 | g, \beta)$  and  $D_{h,\eta_1,DR}(0 | g, Q, \beta)$ , assuming that all parameters other than  $\eta$  are known. It turns out that the resulting ‘leftover’ part can be easily estimated nonparametrically by  $C_h(O | \eta_m, \beta)$ , where  $\eta_m$  is an estimator converging to  $\eta_1$ .

## 4. Estimating intervention effects over all possible interventions

### 4.1. Background

In this section, we propose estimating functions for more general models that simultaneously capture the effect of a number of different interventions, corresponding to all the different values that the treatment variable  $A$  can assume. We focus on models for the additive risk difference and the relative risk, see (2), for which we present inverse-weighting-based estimating functions as well as their doubly-robust extensions. Since we now observe one counterfactual outcome of interest for each observation, the former will be referred to as inverse-probability-of-treatment-weighted rather than inverse-probability-of-censoring-weighted estimating functions.

The estimators based on these estimating functions are consistent if the user can supply a consistent estimator of the treatment mechanism  $g_0(A | W)$ . In order to deal with the curse of dimensionality, it will generally be necessary to assume a model  $G$  within which one can then obtain a maximum likelihood estimator,

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

or versions of this involving regularization and/or model selection. Our proposed doubly-robust estimators are consistent if  $g_0(A | W)$  or  $E(Y | A, W)$  is estimated consistently;  $E(Y | A, W)$  can be estimated using straightforward regression methods.

Note that the models discussed in this section can also be applied if one is interested in a single target intervention  $a^*$ . By borrowing information from other interventions to estimate the effect at the intervention of interest, the resulting estimators can achieve a smaller variance than those described in § 3. However, they are likely to suffer from greater bias if the model is incorrectly specified.

### 4.2. Estimation and inference

In each of the models we will apply the following general strategy for deriving a class of estimating functions, using as an illustrative example the additive risk intervention model. Our general strategy corresponds to the estimating function methodology presented in van der Laan & Robins (2002); estimating functions are implied by the so-called orthogonal complement of the nuisance tangent space, which is equivalent to the class of all influence curves of regular

asymptotically linear estimators. Such estimating functions are orthogonal in the  $L^2(P)$  Hilbert space to all scores implied by fluctuations of the nuisance parameters. Such estimating functions always have first-order robustness with respect to specifying a variation-independent nuisance parameter; in many settings, these estimating functions remain unbiased even under complete misspecification of nuisance parameters. This has important practical consequences, such as in testing a null hypothesis about the additive risk, since the validity of such a test is not affected by bias in the model for  $\eta(V)$ .

**Stage 1**—The population parameter  $\eta_0$ , which is used for comparison by the intervention model, implies a marginal structural model for the conditional distribution of  $Y_a$ , given  $V$ . For example, for a given  $\eta_0(V)$ , the additive risk intervention model implies the marginal structural model  $E(Y_a | V) = \eta_0(V) + m(a, V | \beta_0)$ .

**Stage 2**—Given  $\eta_0$ , this marginal structural model implies a class of inverse-probability-of-treatment-weighted and doubly-robust estimating functions as established in previous literature. For example, the class of doubly-robust estimating functions for  $E(Y_a | V) = \eta_0(V) + m(a, V | \beta_0)$ , with  $\eta_0$  treated as known, is given by

$$D_{h_1}(O|g, Q, \eta_0, \beta) \equiv \frac{h_1(A, V)}{g(A|W)} \{Y - \eta_0(V) - m(A, V|\beta)\} - \frac{h_1(A, V)}{g(A|W)} \{Q(A, W) - \eta_0(V) - m(A, V|\beta)\} + \sum_{a \in \mathcal{A}} h_1(a, V) \{Q(a, W) - \eta_0(V) - m(a, V|\beta)\}, \tag{7}$$

where the index  $h_1$  can be an arbitrary function of  $A$  and  $V$ . Note that this estimating function depends on the unknown nuisance parameters  $g_0(A | W)$  and  $Q_0(A, W)$ . The first term of this estimating function represents the inverse-probability-of-treatment-weighted estimating function, and the last two terms are its projection on to the nuisance tangent space  $T_{RA} = [\varphi(A, W) : E\{\varphi(A, W) | W\} = 0]$  of the treatment mechanism only assuming the randomization assumption. This estimating function is doubly-robust with respect to misspecification of  $g_0$  and  $Q_0$  in the sense that, if  $\max_a \{h_1(a, V)/g_1(a | W)\} < \infty$ ,  $F_{W0}$  almost everywhere, and either  $g_1 = g_0$  or  $Q_1 = Q_0$ , then

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

**Stage 3**—To make the estimator robust to misspecification of  $\eta$ , determine the influence curve of the estimator  $\beta_n$ , solving  $0 = \sum_i D_{h_1}(O_i | g_0, Q_0, \eta_n, \beta)$  for an appropriate estimator  $\eta_n$  of  $\eta_0$  under a nonparametric model for  $\eta_0$ . We treat this class of influence curves, ignoring standardizing constants and matrices, as a new class of estimating functions. This results in a corrected class of estimating functions given by

$$D_{h_1}(O|g, Q, \eta, \beta_0) + C_{h_1}(O|\eta, \beta_0).$$

The latter term actually corresponds to the influence curve of the estimator  $E_0 D_{h_1}(O | g_0, Q_0, \eta_n, \beta_0)$  of the parameter  $E_0 D_{h_1}(O | g_0, Q_0, \eta_0, \beta_0)$ , with the parameters other than  $\eta_0$  treated as known. The influence curve is defined with regard to a regular, asymptotically linear estimator,  $\mu_n$ , of parameter  $\mu_0$  as follows:  $\mu_n - \mu_0 = n^{-1} \sum_{i=1}^n IC_i(O_i) + o_p(n^{-1/2})$ . In the additive risk intervention model,

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

where this class of influence curves is indexed by  $h_1$ , with  $\beta$  treated as known, and  $\eta$  is the only unknown parameter.

**Stage 4**—Finally, in each of our three intervention models, the corrected estimating functions remain unbiased if  $\eta$  is misspecified. This allows us to treat  $\eta$  as another index in the class of estimating functions, which results in the final class of proposed estimating functions indexed by functions  $h_1, h_2$ :

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

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

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

For each of our models we propose a particular choice  $(h_1, h_2)$  depending on the true data-generating distribution, which can easily be estimated. Misspecification of this choice does not affect the unbiasedness of the estimating function, but only results in using a different, suboptimal, unbiased estimating function. As a consequence, a different choice from the optimal  $h_1$  and  $h_2$  does not affect the consistency and asymptotic linearity (van der Laan & Robins, 2002) of the resulting estimator of  $\beta_0$ .

### 4.3. Class of estimating functions for additive risk intervention models

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

$$\begin{aligned} D_{h_1, h_2}(O|g, Q, \beta) &\equiv \frac{h_1(A, V)}{g(A|W)} \{Y - h_2(V) - m(A, V|\beta)\} \\ &\quad - \frac{h_1(A, V)}{g(A|W)} \{Q(A, W) - h_2(V) - m(A, V|\beta)\} \\ &\quad + \sum_{a \in \mathcal{A}} h_1(a, V) \{Q(a, W) - h_2(V) - m(a, V|\beta)\} \\ &\quad - \sum_{a \in \mathcal{A}} h_1(a, V) Y + E \left\{ \sum_{a \in \mathcal{A}} h_1(a, V) h_2(V) \right\}. \end{aligned} \tag{8}$$

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

$$\begin{aligned} h_1(A, V) &\equiv \frac{\partial}{\partial \beta} m(A, V|\beta) g_0(A|W), \\ h_2(V) &\equiv E(Y|V). \end{aligned} \tag{9}$$

The optimal choice for  $(h_1, h_2)$  remains to be investigated. Our goal was to propose choices that are sensible and relatively simple. The natural choice for  $h_2$  is  $E(Y | V)$ . The choice of  $h_1$  is motivated by the discussion in Chapter 6 of van der Laan & Robins (2002), with  $g_0(A | V)$  added as a stabilizing factor that can ameliorate the impact of very small estimated treatment probabilities  $g_n(A | W)$ , as also suggested in Robins et al. (2000). A more general discussion of the choice of  $h$  in the context of nonparametric models of causal effects can be found in Neugebauer & van der Laan (2006). Finally, to simplify the notation, the dependence of the estimating function on the unknown mean, in the last term, is not reflected in our notation.

### 4.4. Class of estimating functions for relative risk intervention models

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

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

Our proposed choices for  $(h_1, h_2)$ , similar to those discussed for the additive risk model, are given by

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

#### 4.5 General mean intervention models

We now consider a general mean intervention model corresponding to a general choice of  $\Phi \{E(Y_a | V), E(Y | V)\} = m(a, V | \beta)$ . Given  $\eta_0(V)$ , this model implies a marginal structural model  $E(Y_a | V) = f\{\eta_0(V), m(a, V | \beta)\}$  for some bivariate real-valued function  $f$ . Our general strategy now results in the following class of estimating functions for  $\beta_0$ :

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

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

It follows that  $E_0 D_{h_1}(O | g_0, Q_0, \eta_1, \beta_0)$  only equals zero at a misspecified  $\eta_1$  if  $x \rightarrow f(x, y)$  is linear, which holds precisely for the additive and relative risk models, but fails to hold for most other functions of interest. 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  $\eta_0(V)$ , beyond consistent estimation of either  $g_0(A | W)$  or  $Q_0(A, W)$ .

#### 4.6. Estimation and asymptotic inference

Given estimators  $h_{1n}, h_{2n}, g_n$  and  $Q_n$  of  $h_1, h_2, g_0$  and  $Q_0$ , the estimator  $\beta_n$  for the additive and relative risk models is defined as the solution of the estimating equation

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

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

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

where

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

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

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

and  $\text{EIC}(O)$  is an estimator of the influence curve  $\text{IC}(O | P_0)$  of our estimator. The Appendix contains a derivation of the influence curve when  $g_n$  consistently estimates  $g_0$ . In general, we recommend using the bootstrap to estimate the distribution of  $n^{-1/2}(\beta_n - \beta_0)$ .

### 5. Simulation study

To examine the finite-sample properties of our estimators, simulation studies were conducted to evaluate the performance of estimators of models of the additive risk. In the notation introduced in § 2, the data consist of independent and identically distributed realizations of  $O = (W = (V, Z), A, Y)$ . Interest lies in estimating the projection of the risk difference  $E(Y | V) - E(Y_a | V)$  on to the working model  $m(a, V | \beta) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 (V - 3)_+ + \beta_4 a V + \beta_5 a (V - 3)_+$ , where  $(V - 3)_+$  is a linear spline term. This model was chosen because we wanted to examine the performance under situations in which the model for the additive risk does and does not closely approximate the true additive risk as a function of  $V$ . Our model above accomplishes this if, when  $a = 0$ , the model closely approximates the true function, whereas, for  $a = 1$ , the model is clearly misspecified; see Figs 1 and 2. A linear model was used to estimate the additive risk, which does not lead to difficulties in estimation given that our estimating approach does not require this model to constrain the risk difference to the interval  $(-1, 1)$  and, in this case, the true additive risks are well away from these bounds. If one expects more significant additive risks and/or has  $V$  that are unbounded, then other forms of  $m$  could be chosen to avoid estimating additive risks outside  $(-1, 1)$ .

The data were generated as follows. Both  $V$  and  $Z$  were distributed uniformly over the integers 1–7. Both  $Y$  and  $A$  are binary, following the conditional distributions

$$\begin{aligned} \text{logit}\{\text{pr}(A=1|W)\} &= Z - V + 0.5ZV \\ \text{logit}\{\text{pr}(Y=1|W,A)\} &= -2 + Z + 2V^2 + 0.5A + 0.25AV^2. \end{aligned}$$

The performance of three estimators is examined: an inverse-probability-of-treatment-weighted estimator, referred to as IPTW, without the correction factor,  $C_{h1}(O | \eta_0, \beta_0)$ , which is derived from the estimating function given by the first term of (7); a doubly-robust estimator, referred to as DR1, that is defined by the estimating function (4) and also does not use the correction factor; and another doubly-robust estimator, referred to as DR2, that is defined by the estimating function (8), which has the correction factor that should make the estimator robust to misspecification of  $\eta$ . For all these estimators, we use  $h_1(A, V) = \partial m(A, V | \beta) / \partial \beta = (1, A, V, (V - 3)_+, AV, A(V - 3)_+)$ . The functions  $g(A | W)$  and  $\eta(V)$  are estimated through logistic regression, with the model for  $\alpha(A | W)$  correctly specified as  $\text{logit}\{g(1 | W, \alpha)\} = \alpha_0 + \alpha_1 Z + \alpha_2 V + \alpha_3 ZV$ , but the model for  $\eta(V)$  misspecified as  $\text{logit}\{\eta(V | \gamma)\} = \gamma_0 + \gamma_1 V$ ; the true model is a more complicated, approximately quadratic form. The estimator  $h_{2,n}(V) = \eta_n(V)$  obtained from this model is then used for  $h_2(V)$ . One would thus expect the estimators IPTW and DR1 of  $\beta$  to be biased, because of an inconsistent estimator of  $\eta(V)$ , whereas the estimator DR2 should be less biased. The nuisance parameter  $Q(A, W)$ , required by the estimators DR1 and DR2, is estimated through the correctly specified logistic regression model  $\text{logit}\{Q(A, W | \theta)\} = \theta_0 + \theta_1 Z + \theta_2 V^2 + \theta_3 A + \theta_4 AV^2$ . The term  $E\{\sum_{a \in \mathcal{A}} h_1(a, V) h_2(V)\}$  required by the estimator DR2 is estimated through the five-dimensional, to fit the dimension of  $\beta$ , vector of averages,

$$\widehat{E} \left\{ \sum_{a \in \mathcal{A}} h_1(a, V) h_{2,n}(V) \right\} = \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{a \in \mathcal{A}} h_1(a, V_i) h_{2,n}(V_i) \right\}.$$



Finally, we use a numerical optimization routine to find the minimum of the norm of the estimating equations with respect to  $\beta$ :

$$\beta_n = \arg \min_{\beta} \left\| \sum_{i=1}^n D_{h_1, h_2, n} (O_i | g_n, Q_n, \beta) \right\|.$$

We compare the three estimators to the true values of  $E(Y | V) - E(Y_a | V)$  for  $a = 0, 1$  and  $V = 1, \dots, 7$ , using sample sizes of  $n = 1000$ . Figure 1 compares the mean estimates, over the 1000 simulations, for the three estimators to the true values at the two values of the intervention,  $a = 0$  and  $a = 1$ , separately in Figs 1(a) and (b), respectively. In addition, the overall relative performance, both variance and bias, is displayed as the relative mean-squared error, relative to the estimator IPTW, as RMSE1 and RMSE2 for the estimators DR1 and DR2, respectively, see Figs 1(c) and (d). The same simulation is repeated for a sample size of 1000; see Fig. 2.

The results suggest that, although the model for the additive risk does not contain the truth, it does a reasonable job of approximating the trend of  $E(Y | V) - E(Y_a | V)$  as a function of  $V$  for  $a = 0$ ; see Fig. 1(a). However, for  $a = 1$  the model is clearly biased; see Fig. 1(b). The gain in relative efficiency is apparent for  $a = 0$  for both of the doubly-robust estimators relative to the estimator IPTW and this gain all comes from reduction in variance. In this particular case, the bias reduction resulting from the correction factor, for misspecification of  $\eta(V)$ , is small relative to the bias of the underlying additive risk model, Fig. 1(c). The biases for all three estimators are relatively large if  $a = 1$  and there is no gain over the simpler estimator IPTW; see Fig. 1(d).

The story is somewhat different for larger sample sizes. The variance now is very small relative to the bias so that all differences in performance are based on differences in bias. Thus, even those estimators with, DR2, and without the correction factor for misspecification of  $\eta(V)$ , IPTW and DR1, have similar performance at  $a = 0$ ; see Figs 2(a) and (c). For  $a = 1$ , now the two approaches without the correction factor again have identical performance, but the DR2 model now bends, see Fig. 2(b), to reduce the bias overall. However, given the limits of the estimating model for the parameter of interest, i.e. additive risk, reducing bias for some  $V$  increases it at others; the relative mean-squared error dips below 1 at  $V = 4, 5$  and  $7$  as shown in Fig. 2(d). Finally, if one repeats these simulations with  $\eta(V)$  correctly specified, DR1 and DR2 behave nearly identically, implying that there is no further bias-reduction advantage in the estimator DR2 when the estimator of  $\eta(V)$  is consistent and both estimators have similar variances.

The results suggest that the safest estimator is DR2, as it had both the benefits of lower bias and lower variance, where the relative contributions of these two sources of error change as a function of sample size.

## 6. Discussion

For some risk factors, one is interested in a single intervention corresponding to a particular level of the risk factor. In the case of smoking, for instance, we would only be interested in an intervention that eliminates the exposure completely; it would make little sense to estimate the disease distribution in a counterfactual population in which every subject smokes three cigarettes a day as that would require one to compel many people to start smoking. In other applications, such as air pollution studies, however, it does make sense to examine the intervention effect for interventions corresponding to different target levels. For such cases, we recommend the use of models of the form  $m(a, V | \beta) = E(Y_a - Y | V)$ .

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  $j = 1, \dots, p$ , and a final outcome  $Y$ . In this case, the counterfactuals of interest,  $Y_{\bar{a}}$ , are indexed by an entire treatment history,  $\bar{a} = (a(0), a(1), \dots, a(j))$ . As outlined above, if  $\eta_0(V)$  is known or can be estimated nonparametrically, we can use the existing estimating equations developed for estimating the mean of  $E(Y_{\bar{a}} | V)$ . Using the doubly-robust estimating functions presented by van der Laan & Robins (2002) for marginal structural models for time-dependent treatments, we can simply plug in  $\eta_0(V) + m(\bar{a}, V | \beta)$  for  $E(Y_{\bar{a}} | V)$ , where  $m(\bar{a}, V | \beta) = E(Y_{\bar{a}} | V)$ . In a manner analogous to the approach outlined in this article, these estimating equations can also be altered for robustness against misspecification of  $\eta(V)$ .

## Appendix

### Estimating the influence curve for asymptotic inference

Our estimator  $\beta_n$  for the general mean intervention models will be consistent if either  $g_n$  is consistent for  $g_0$  or  $Q_n$  is consistent for  $Q_0$ . Regarding statistical inference, we assume that  $g_n$  consistently estimates  $g_0$ . It is then straightforward to show that, under regularity conditions,  $\beta_n$  is asymptotically linear with influence curve

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

where  $\text{ic}_{\text{nuis}}$  is the influence curve of  $\Gamma(G_n)$  as an estimator of

$$\begin{aligned} \Gamma(G) & \equiv P_{F_{X_0, G}} \frac{h_1(A, V)}{g_0(A|W)} \{Q_1(A, W) - Y\}, \\ c_0 & \equiv \frac{d}{d\beta_0} E_0 D_{h_1, h_2}(O | g_0, Q_1, \beta_0). \end{aligned}$$

In the special case in which  $h_2 = h^*_2$ , it can be shown that

$$\text{IC}_{\text{nuis}} = -\Pi(D_{h_1}(\cdot | Q_1, g_0, \eta_0, \beta_0) | T_2(P_0)),$$

where  $D_{h_1}(O | Q_1, g_0, \eta_0, \beta_0)$  is the doubly-robust estimating function for the standard marginal structural model corresponding to the intervention model with  $\eta_0$  being known; see (7) for its definition in the additive risk intervention model. Also,  $T_2(P_0) \subset \{\varphi(A, W) : E(\varphi(A, W) | W) = 0\}$  is the tangent space of the model  $\mathcal{G}$  for  $g_0$  at  $p_0$ , and  $\Pi\{\cdot | T_2(P_0)\}$  is the projection operator on to this subspace  $T_2(P_0)$  within the Hilbert space  $L^2_0(P_0)$ .

In general, it seems to be a reasonable approach to use  $\text{ic}(O | P_0)$  with this particular choice of  $\text{ic}_{\text{nuis}}$  in this influence curve decreases the variance of the influence curve, since it subtracts a  $T_2(P_0)$  component of  $D_{h_1}(O | Q_1, g_0, \eta_0, \beta_0)$ . As a consequence, a typically conservative influence curve is given by

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

which corresponds to the influence curve one would obtain if one uses  $g_n = g_0$  as if  $g_0$  were known.

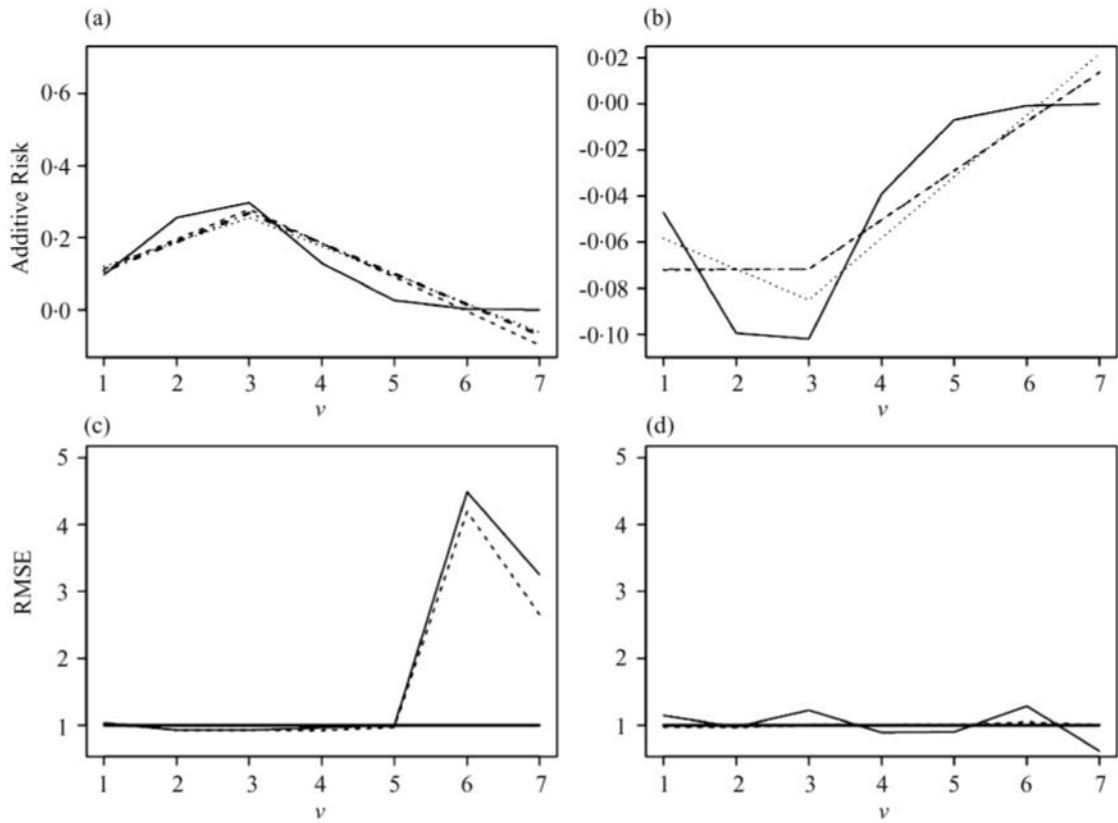
Statistical inference for  $\beta_0$  can now be based on  $\beta_n \sim N(\beta_0, \Sigma_n)$ , where

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n \text{EIC}(O_i) \text{EIC}(O_i)^T$$

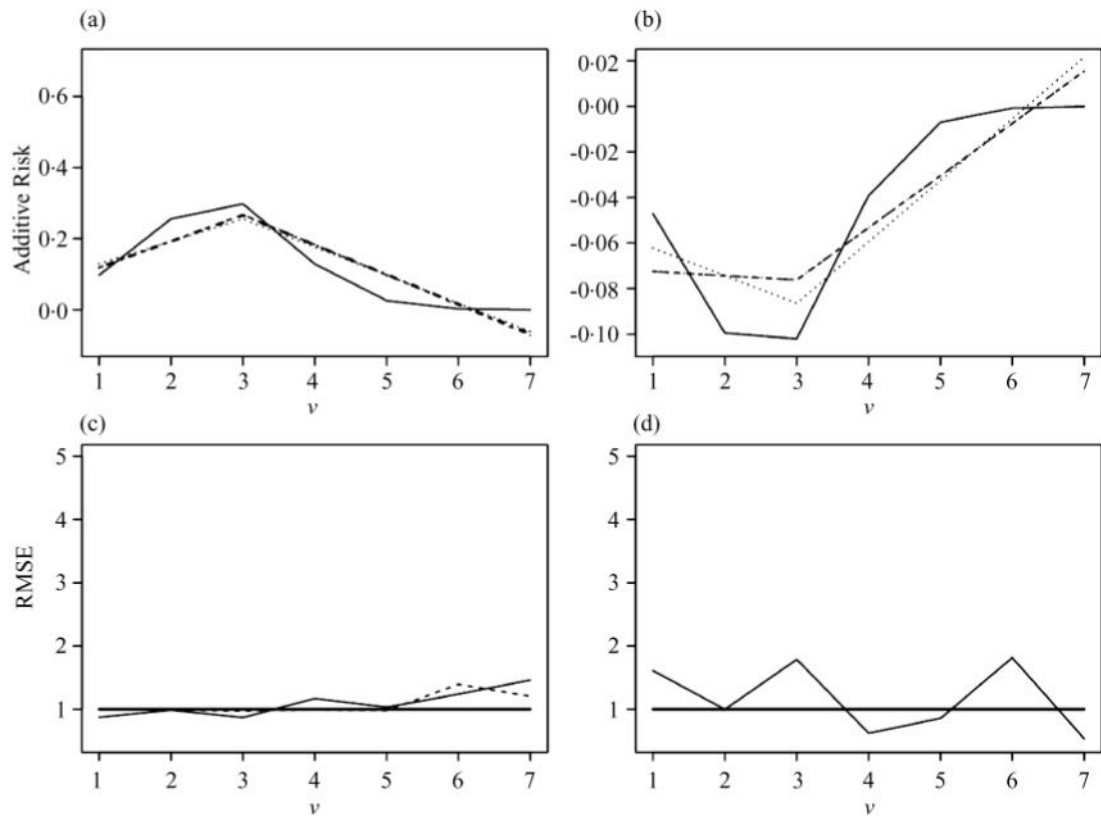
and  $\text{eic}(O)$  is the estimate of the influence curve  $\text{ic}(O | P_0)$  obtained by plugging in our estimates of  $g_0$ ,  $Q_0$ ,  $h^*_1$ ,  $h^*_2$  and  $\beta_0$ . As mentioned in § 4, one can avoid deriving the influence curve of the estimator by using the bootstrap to estimate the distribution of  $n^{1/2}(\beta_n - \beta_0)$ .

## References

- Neugebauer R, van der Laan MJ. Nonparametric causal effects based on marginal structural models. *J. Statist. Plan. Infer* 2006;137:419–34.
- Robins, JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach in causal inference in longitudinal studies.. In: Sechrest, L.; Freeman, H.; Mulley, A., editors. *Health Service Methodology: A Focus on AIDS*. U.S. Public Health Service, National Center for Health Services research; Washington, DC: 1989. p. 113-59.
- Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun. Statist* 1994;A 23:2379–412.
- Robins, JM. Marginal structural models versus structural nested models as tools for causal inference.. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer; New York: 2000. p. 95-113.
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–60. [PubMed: 10955408]
- Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers.. In: Jewell, NP.; Dietz, K.; Farewell, VT., editors. *AIDS Epidemiology: Methodological Issues*. Birkhäuser; Boston, MA: 1992. p. 24-33.
- Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann. Statist* 1978;6:34–58.
- Rubin DB. Comment on a paper by P.W. Holland. *J. Am. Statist. Assoc* 1986;81:961–2.
- van der Laan, M.; Robins, J. *Unified Methods for Censored Longitudinal Data and Causality*. Springer; New York: 2002.
- van der Laan MJ, Hubbard A, Jewell NP. Estimation of treatment effects in randomized trials with noncompliance and a dichotomous outcome. *J. R. Statist. Soc* 2007;B 69:463–82.



**Fig. 1.** Simulation study. Performance of additive risk estimators with  $n = 100$ . The estimators and true values for each  $V = v$  for (a)  $a = 0$ , (b)  $a = 1$ , for the truth, solid line, iptw, dashed line, dr1, dashed-dotted line, and dr2, dotted line. The relative mean-squared error, rmse, relative to the iptw for (c)  $a = 0$  and (d)  $a = 1$  for the DR1, dashed line, and DR2, solid line, with the solid thick black line at rmse for reference.



**Fig. 2.** Simulation study. Performance of additive risk estimators with  $n = 1000$ . The estimators and true values for each  $V = v$  for (a)  $a = 0$ , (b)  $a = 1$ , for the truth, solid line, iptw, dashed line, DR1, dashed-dotted line, and DR2, dotted line. The relative mean-squared error, rmse, relative to the iptw for (c)  $a = 0$  and (d)  $a = 1$  for the DR1, dashed line, and DR2, solid line, with the solid thick black line at  $RMSE = 1$  for reference.