Identification and estimation of causal effects with outcomes truncated by death

By LINBO WANG

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

linbowang@g.harvard.edu

XIAO-HUA ZHOU

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

azhou@u.washington.edu

AND THOMAS S. RICHARDSON

Department of Statistics, University of Washington, Seattle, Washington 98195, U.S.A.

thomasr@u.washington.edu

Summary

It is common in medical studies that the outcome of interest is truncated by death, meaning that a subject has died before the outcome could be measured. In this case, restricted analysis among survivors may be subject to selection bias. Hence, it is of interest to estimate the survivor average causal effect, defined as the average causal effect among the subgroup consisting of subjects who would survive under either exposure. In this paper, we consider the identification and estimation problems of the survivor average causal effect. We propose to use a substitution variable in place of the latent membership in the always-survivor group. The identification conditions required for a substitution variable are conceptually similar to conditions for a conditional instrumental variable, and may apply to both randomized and observational studies. We show that the survivor average causal effect is identifiable with use of such a substitution variable, and propose novel model parameterizations for estimation of the survivor average causal effect under our identification assumptions. Our approaches are illustrated via simulation studies and a data analysis.

Some key words: Causal inference; Instrumental variable; Model parameterization; Principal stratification; Survivor average causal effect.

1. Introduction

In medical studies, researchers are often interested in evaluating risk factors for a non-mortality outcome. However, this outcome may be truncated by death and hence undefined if some subjects die before the follow-up assessment. For example, suppose we are interested in estimating the effect of smoking on memory decline in an aged population. If a subject dies before the follow-up memory test is administered, then his/her memory score at the follow-up visit is undefined. Direct comparisons between smokers and non-smokers among observed survivors are subject to selection bias, as non-smokers are more likely to survive to the follow-up assessment, and survival is associated with memory decline (Rosenbaum, 1984; Robins & Greenland, 1992). More

fundamentally, direct comparisons between observed survivors are not causally interpretable, as they compare outcomes from different subpopulations at baseline (Rubin, 2006). Robins (1986, § 12.2) proposed estimating the average causal effect in the always-survivor group, the group of subjects who would survive if they chose to receive either exposure at baseline. The always-survivor group was later termed a principal stratum by Rubin (1998) and Frangakis & Rubin (1999, 2002), and the contrast within this group is the survivor average causal effect. This contrast is causally interpretable because membership of the always-survivor group is defined at baseline, and since these subjects always survive, their subsequent outcomes under both treatments are well-defined. Alternative estimands are discussed by Kurland et al. (2009) and Weuve et al. (2012).

The survivor average causal effect is not identifiable without further assumptions (Zhang & Rubin, 2003). Large-sample bounds for the survivor average causal effect have been derived under minimal assumptions (Zhang & Rubin, 2003; Imai, 2008; Long & Hudgens, 2013). In order to identify this effect, it is common to perform a sensitivity analysis by assuming a class of identification conditions indexed by a sensitivity parameter (Gilbert et al., 2003; Hayden et al., 2005; Shepherd et al., 2006; Egleston et al., 2007; Jemiai et al., 2007; Chiba & VanderWeele, 2011).

Alternatively, identification of the survivor average causal effect can be based on covariate information. For example, Tchetgen Tchetgen (2014) introduced a variant of the survivor average causal effect that is identified when risk factors of survival are available in post-exposure follow-ups. The resulting causal contrast is for a principal stratum defined in terms of particular post-exposure risk factors. Hence, in principle, the causal estimand can differ depending on the set of post-exposure risk factors used in the analysis. Ding et al. (2011) took a different approach to identifying the survivor average causal effect in a randomized study setting. They proposed a semiparametric identification method based on a baseline variable whose distribution is informative of the membership of the always-survivor group. Using this baseline variable, they showed that the survivor average causal effect was identifiable under their assumptions. However, as pointed out by Tchetgen Tchetgen (2014), their assumptions essentially require that there be no common causes of the survival and outcome processes, which is very unlikely even in randomized studies.

In this article, we relax the identification assumptions of Ding et al. (2011) by employing more detailed covariate information. In contrast to the estimand of Tchetgen Tchetgen (2014), our causal parameter is defined independently of the covariates incorporated in the analysis. Furthermore, the proposed approach is applicable to both randomized trials and observational studies, and allows for measured common causes of the survival and outcome processes. We also discuss possible violations of, and alternative approaches to, our identification model. To estimate the survivor average causal effect in practice, we impose additional distributional assumptions. This is challenging for two reasons. First, unlike the standard observational study setting, in our identification framework the baseline covariates may be common causes between any pair of treatment, survival and outcome. As we explain in § 4, this unique role for baseline covariates makes the standard propensity score methods inappropriate in our setting. Second, novel model parameterizations are needed to make our distributional assumptions compatible with our identification assumptions.

2. DATA STRUCTURE, NOTATION AND CAUSAL ESTIMAND

Consider a medical study with a single follow-up visit. Let Z be the exposure indicator and let W denote covariates observed at baseline. We assume that each subject has two potential

Table 1. *Patient survival types*

$S(1)$	$S(0)$	Survival type	G	Description
1	1	Always-survivor	LL	The subject always survives, regardless of exposure status
1	0	Protected	LD	The subject survives if exposed, but dies if not exposed
0	1	Harmed	DL	The subject dies if exposed, but survives if not exposed
0	0	Doomed	DD	The subject always dies, regardless of exposure status

survival outcomes, $S(1)$ and $S(0)$, defined as the survival statuses at the follow-up visit that would have been observed had the subject been exposed and unexposed, respectively. Similarly, we let $Y(1)$ and $Y(0)$ denote the potential non-mortality outcomes under exposure and nonexposure, respectively. We assume that for $z = 0, 1$, $Y(z)$ takes real values only if $S(z) = 1$. We extend the definition of $Y(z)$ so that it takes the constant value $*$ if $S(z) = 0$.

We use G to denote the survival type as defined in Table 1, in which L represents live and D represents die. One can see from Table 1 that there exists a one-to-one mapping between the survival type and the bivariate potential survival $\{S(1), S(0)\}$, so G can be interpreted as an abbreviation for $\{S(1), S(0)\}$.

We adopt the axiom of consistency such that the observed outcome Y satisfies $Y = ZY(1) + (1 - Z)Y(0)$ and the observed survival S satisfies $S = ZS(1) + (1 - Z)S(0)$. The observed samples $O_i = (Z_i, W_i, S_i, Y_i)$ $(i = 1, \ldots, N)$ are independently drawn from an infinite superpopulation.

Throughout this article, we assume that there is no interference between study subjects with respect to both the survival S and the non-mortality outcome Y, and that there is only one version of exposure (Rubin, 1980).

Rubin (2000) noted that the survivors observed in the exposed group are from a mixture of the always-survivor and protected strata, while the survivors observed in the nonexposed group are from a mixture of always-survivor and harmed strata. As a result, direct comparisons between different exposure groups among observed survivors are not causally meaningful, as these people are from different subpopulations at baseline. To address this issue, since the always-survivor stratum is the only group for which $Y(1)$ and $Y(0)$ both take real values, we define our causal estimand to be the average causal effect in this stratum:

$$\Delta_{\mathrm{LL}} = E\{Y(1) - Y(0) \mid G = \mathrm{LL}\}.$$

This estimand is also known as the survivor average causal effect.

3. Identification of the survivor average causal effect

3·1. *The identification problem*

Although causally interpretable, the survivor average causal effect is in general not identifiable as it depends on the potential outcomes $Y(1), Y(0), S(1)$ and $S(0)$. Furthermore, it is not identifiable even under assumptions that are sufficient for identification in other causal contexts:

Assumption 1 (Monotonicity). $S(1) \geqslant S(0)$ almost surely.

The monotonicity assumption may be plausible in some observational studies. For example, in studies evaluating the effect of smoking on memory decline, it is widely believed that smoking is bad for overall health and hence overall survival. This assumption tends to be questionable in

randomized clinical trials with acute diseases, because typically a clinical trial would be unethical if the researchers believe that one treatment benefits survival a priori. To address this issue, we relax this assumption later in § 3·4.

Assumption 2 (*S*-ignorability). The treatment assignment is independent of the potential survival statuses given the observed covariates W, so that $Z \perp\!\!\!\perp S(z) \mid W$ for $z = 0, 1$.

Assumption 3 (*Y*-ignorability). The treatment assignment in the always-survivor stratum is independent of the potential outcomes given observed covariates W, so that $Z \perp\!\!\!\perp Y(z) \mid W, G = $ LL for $z = 0, 1$.

Assumptions 2 and 3 are similar to the weakly ignorable treatment assignment assumption (see, e.g., Imbens, 2000). Under Assumption 3, we have

$$E\{Y(z) \mid G = \mathrm{LL}\} = \frac{E_W(\mu_{z,\mathrm{LL},W} \pi_{\mathrm{LL}|W})}{E_W(\pi_{\mathrm{LL}|W})}, \tag{1}$$

where $\mu_{z,g,W} = E(Y \mid Z = z, G = g, W)$ and $\pi_{g|W} = \mathrm{pr}(G = g \mid W)$. Under Assumptions 1 and 2, $Z \perp\!\!\!\perp \{S(1), S(0)\} \mid W$, so the W-specific survival-type probabilities are identified by

$$\pi_{\mathrm{LL}|W} = \mathrm{pr}(S = 1 \mid Z = 0, W), \quad \pi_{\mathrm{LD}|W} = \mathrm{pr}(S = 1 \mid Z = 1, W) - \mathrm{pr}(S = 1 \mid Z = 0, W); \tag{2}$$

$\mu_{0,\mathrm{LL},W}$ can be identified by

$$\mu_{0,\mathrm{LL},W} = E(Y \mid Z = 0, S = 1, W). \tag{3}$$

However, $\mu_{1,\mathrm{LL},W}$ is not identifiable from the observed data. In fact, as noted by Zhang & Rubin (2003), the observed data in the group $\{i : Z_i = 1, S_i = 1, W_i = w_i\}$ can be written as a mixture of two distributions from the LL and LD strata:

$$F(y \mid Z = 1, S = 1, W) = \sum_{g=\mathrm{LL},\mathrm{LD}} p_{g|1,W,s=1} F(y \mid Z = 1, G = g, W),$$

where F denotes the cumulative distribution function, and for $g = $ LL or LD, $p_{g|z,W,s=1} \equiv \mathrm{pr}(G = g \mid W)/\mathrm{pr}(S = 1 \mid Z = z, W)$ is identifiable from data. Unless the mixing probability $p_{\mathrm{LL}|1,w,s=1}$ is 1 or 0, $\mu_{1,\mathrm{LL},W}$ is only partially identifiable in that there is a range of values for $\mu_{1,\mathrm{LL},W}$ that are compatible with the observed data distribution.

3·2. *Identifying the survivor average causal effect using a substitution variable*

To identify the survivor average causal effect, without loss of generality we assume that the baseline covariates W can be written as (X, A). We propose an identification framework in which the role of X is similar to a confounder while that of A is similar to an instrument. Specifically, we make the following assumptions on A and X:

Assumption 4 (Exclusion restriction). $A \perp\!\!\!\perp Y(1) \mid Z = 1, G, X$;

Assumption 5 (Substitution relevance). $A \not\!\perp\!\!\!\perp G \mid Z = 1, S = 1, X$.

The survival type G is a latent baseline variable that satisfies Assumptions 4 and 5. For this reason, any variable A satisfying Assumptions 4 and 5 is called a substitution variable for G. The
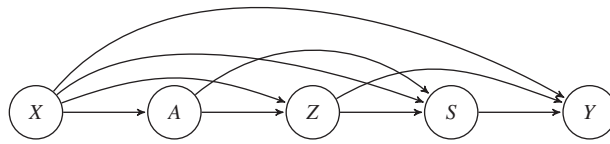
Fig. 1. The simplest causal diagram associated with the structural equations (4).

conditions for a substitution variable are similar to those for an instrumental variable (Angrist et al., 1996). Specifically, Assumption 4 is similar to the exclusion restriction assumption in an instrumental variable analysis in that they both capture the notion of A having no direct effect on Y, and Assumption 5 is similar to the instrumental relevance assumption; in particular, they both require that A be a relevant variable. We clarify that, similar to the instrumental relevance assumption, Assumption 5 needs to hold for all possible values of X. As we illustrate in § 3·3, even in a randomized study setting, the inclusion of covariate information X makes Assumptions 4 and 5 more plausible.

Theorem 1 below states that the survivor average causal effect is identifiable with a substitution variable for the survival type. Proofs of the theorems can be found in the Supplementary Material.

THEOREM 1. *Under Assumptions 1–5,* Δ_{LL} *is identifiable.*

When A takes more than two values, Δ_{LL} may be over-identified. In this case, one may falsify our identification assumptions using specification tests such as the Sargan–Hansen test (Hansen, 1982).

3·3. *A nonparametric structural equation model with independent errors*

The key assumptions in our identification model are Assumptions 2–4. These are implied by a certain nonparametric structural equation model with independent errors (Pearl, 2009), given by

$$
\begin{aligned}
X &= f_X(\epsilon_X), \quad A = f_A(X, \epsilon_A), \quad Z = f_Z(X, A, \epsilon_Z), \\
S &= f_S(X, A, Z, \epsilon_S), \quad Y = f_Y(X, Z, S, \epsilon_Y),
\end{aligned}
\tag{4}
$$

where the error terms $\epsilon_X, \epsilon_A, \epsilon_Z, \epsilon_S$ and ϵ_Y are jointly independent. Here the error terms can be interpreted as the set of one-step-ahead counterfactuals; for example, ϵ_Z can be interpreted as $\{Z(a, x) : a \in \mathcal{A}, x \in \mathcal{X}\}$, where $\mathcal{A}$ and $\mathcal{X}$ are the sample spaces for A and X, respectively. Figure 1 gives the simplest causal diagram associated with (4). There is no directed edge from A to Y, encoding the notion that A has no direct effect on Y. We remark that Assumptions 2–4 also follow from (4) with dependent errors that allow for unmeasured confounding between nodes in Fig. 1. See the Supplementary Material for examples.

In contrast to the identification assumptions in Ding et al. (2011), we include baseline covariates X in our identification framework. Our approach is applicable to observational studies, since we allow for the edge $X \to Z$. Moreover, we allow for edges $X \to S$ and $X \to Y$, thereby avoiding the assumption that there are no common causes of the survival and outcome processes, which is very restrictive even for randomized studies.

Remark 1. The assumptions of a nonparametric structural equation model with independent errors cannot be tested, even in principle, via any randomized experiment on the measured variables (Robins & Richardson, 2010).

3·4. *Alternatives to the individual-level monotonicity assumption*

We now consider alternative stochastic monotonicity assumptions in place of the individual-level Assumption 1. These alternatives were used by Roy et al. (2008) in the context of noncompliance and by Lee et al. (2010) in the context of truncation by death.

We characterize the association between the potential survivals $S(1)$ and $S(0)$ by

$$\rho(W) = \frac{\mathrm{pr}(G = \mathrm{LL} \mid W) - \mathrm{pr}\{S(1) = 1 \mid W\} \, \mathrm{pr}\{S(0) = 1 \mid W\}}{\min[\mathrm{pr}\{S(1) = 1 \mid W\}, \, \mathrm{pr}\{S(0) = 1 \mid W\}] - \mathrm{pr}\{S(1) = 1 \mid W\} \, \mathrm{pr}\{S(0) = 1 \mid W\}}.$$

Note that $\rho(W)$ is no larger than 1 and differs from the correlation between $S(1)$ and $S(0)$ by a nonnegative identifiable factor. In particular, the sign of $\rho(W)$ coincides with the sign of the conditional correlation $\mathrm{cor}\{S(1), S(0) \mid W\}$. In practice, one might be willing to assume that the potential survivals under active treatment and control are not negatively correlated, such that

$$\mathrm{pr}\{S(1) = 1 \mid S(0) = 1, W\} \geqslant \mathrm{pr}\{S(1) = 1 \mid S(0) = 0, W\}$$

almost surely. This gives rise to the following stochastic monotonicity assumption.

Assumption 6 (Stochastic monotonicity). We have that $0 \leqslant \rho(W) \leqslant 1$ almost surely.

Note that $\rho(W) = 0$ if and only if conditional on observed covariates W, $S(1)$ is not correlated with $S(0)$. On the other hand, $\rho(W) = 1$ if and only if $\mathrm{pr}(G = \mathrm{LL} \mid W) = \min[\mathrm{pr}\{S(1) = 1 \mid W\}, \mathrm{pr}\{S(0) = 1 \mid W\}]$. We emphasize that Assumption 6 does not reduce to Assumption 1 when $\rho(W) = 1$; in particular, assuming a value for $\rho(W) \in [0, 1]$ does not specify which treatment is more beneficial for survival. This is particularly appealing when the monotonicity assumption is not plausible a priori due to ethical reasons.

Theorem 2 states that under Assumptions 2–6, if we further assume that Assumptions 4 and 5 hold in the control arm, then given $\rho(W)$, the survivor average causal effect is identifiable from the observed data. The additional conditional independence condition (i) in Theorem 2 also follows from the structural equation model (4) with independent errors.

THEOREM 2. *Suppose that Assumptions 2–6 hold and that:*

(i) $A \perp\!\!\!\perp Y(0) \mid Z = 0, G, X;$
(ii) *for all $x \in \mathcal{X}$, $A \not\!\perp\!\!\!\perp G \mid Z = 0, S = 1, X = x$.*

Then Δ_{LL} is identifiable given $\rho(W)$.

3·5. *Relaxing the exclusion restriction assumption*

In practice, the ignorability assumptions, Assumptions 2 and 3, may be questionable if there are unmeasured confounders between the exposure Z and either the survival S or the non-mortality outcome Y. Similarly, the exclusion restriction assumption, Assumption 4, may be questionable if the set of observed covariates X is not rich enough to contain all common causes of the non-mortality outcome and any one of the treatment, substitution variable and survival. There is extensive literature on relaxing the ignorability assumptions for causal inference in observational studies (see, e.g., Rosenbaum & Rubin, 1983; Imbens, 2003; Ding & Vanderweele, 2014), which may be applied here with slight modifications. Hence, in this article, we focus on alternative assumptions to relax the exclusion restriction assumption.

Specifically, we propose an alternative no-interaction assumption, which says that when conditioning on the baseline variables X, neither Z nor G modifies the effect of A on Y.

Assumption 7 (No interaction). For all x, a_0 and a_1,

$$E(Y \mid Z = 1, G = \mathrm{LD}, X = x, A = a_1) - E(Y \mid Z = 1, G = \mathrm{LD}, X = x, A = a_0) \tag{5}$$

$$= E(Y \mid Z = 1, G = \mathrm{LL}, X = x, A = a_1) - E(Y \mid Z = 1, G = \mathrm{LL}, X = x, A = a_0) \tag{6}$$

$$= E(Y \mid Z = 0, G = \mathrm{LL}, X = x, A = a_1) - E(Y \mid Z = 0, G = \mathrm{LL}, X = x, A = a_0). \tag{7}$$

In contrast, Assumption 4, implies that the comparisons in (5) and (6) are both zero.

This no-interaction assumption may be plausible when the substitution variable affects the outcome independently of other disease processes. For example, suppose that one is interested in studying the effects of statin usage on memory decline, taking baseline dementia status as the substitution variable. It is likely that baseline dementia status will affect the outcome directly, as memory decline is irreversible: subjects with dementia are more likely to experience a decline in memory. However, it seems reasonable to assume that neither statin usage, which is typically due to high blood cholesterol levels, nor the latent survival statuses, which largely reflect baseline cholesterol levels, would interact, on the additive scale, with the process of memory decline.

To relax Assumption 4 in the absence of X, Ding et al. (2011) also assumed that (5) is equal to (6). Moreover, they placed linearity assumptions on $E(Y \mid Z = 1, G, A)$ and required that the substitution variable A have at least three categories or be continuous. Both assumptions may be overly restrictive in practice; for instance, they may not hold in the example of statin usage and memory decline described above, taking A to be dementia status at baseline. This is because current knowledge on dementia suggests that baseline cognitive status has a nonlinear influence on the deterioration of cognitive score (see, e.g., Sperling et al., 2014); in addition, baseline dementia status is often coded as a binary variable. To mitigate the need for basing identification on a linearity assumption and allow for binary substitution variables, we instead assume that (6) is equal to (7); that is, Z is not an effect modifier for the conditional effect of A on Y. This is similar in spirit to the no-interaction assumption in causal mediation analysis (e.g., Robins, 2003). In particular, it is guaranteed to hold under the strong null of no treatment effect of Z on Y.

The inclusion of baseline covariates in Assumption 7 has important implications for the design of studies with truncation by death: every possible cause of the outcome should be measured, if feasible, to try to ensure that no residual effect modification remains within strata defined by X.

Theorem 3 states that if we replace the exclusion restriction assumption, Assumption 4, with Assumption 7, then the survivor average causal effect is still identifiable.

THEOREM 3. *Under Assumptions* 1–3, 5 *and* 7, Δ_{LL} *is identifiable.*

4. MODEL PARAMETERIZATION

In the previous section we showed that the survivor average causal effect is identifiable under various identification assumptions. When the covariates W take only a few discrete values within each level of X, the identification formulae in the proofs of Theorems 1–3 imply plug-in estimators in the case of binary A and generalized method of moment estimators in the case where A takes more than two discrete values (Hansen, 1982; Ding et al., 2011). When the covariates W are continuous or high-dimensional, however, we need to impose additional distributional assumptions. We first note that the identification assumptions imply certain constraints on the observed data law, as described in Proposition 1.

Proposition 1.

(I) *The assumptions of Theorem* 1 *imply the following constraints on the law of* (Z, S, Y, W) :

$$\text{pr}(S = 0 \mid Z = 1, X, A) \leqslant \text{pr}(S = 1 \mid Z = 1, X, A); \tag{8}$$

for all x, $E(Y \mid Z = 1, S = 1, X = x, A = a)$ *is bounded as a function of* a; $\quad$ (9)

for all x, $\dfrac{\text{pr}(S = 1 \mid Z = 0, X = x, A = a)}{\text{pr}(S = 1 \mid Z = 1, X = x, A = a)}$ *is not constant as a function of* a. $\quad$ (10)

(II) *The assumptions of Theorem* 2 *imply* (9) *and that*

for all x, $E(Y \mid Z = 0, S = 1, X = x, A = a)$ *is bounded as a function of* a. $\quad$ (11)

(III) *The assumptions of Theorem* 3 *imply* (8) *and that*

for all x, $E(Y \mid Z = 1, S = 1, X = x, A = a) - E(Y \mid Z = 0, S = 1, X = x, A = a)$

is bounded as a function of a. $\quad$ (12)

To ensure that the modelling assumptions are compatible with the constraints in Proposition 1, we avoid imposing distributional assumptions directly on the observed data. Instead, we make the following distributional assumptions on the law of $\{Z, W, S(1), S(0), Y(1), Y(0)\}$, which are compatible with the identification assumptions. For brevity, here we only give models under the assumptions of Theorem 1. Parameterizations for estimating Δ_{LL} under alternative identification assumptions which are similar in spirit are described in the Appendix.

Model 1. The expectation $E\{Y(0) \mid Z = 0, G = \text{LL}, X, A\}$ is known up to a finite-dimensional parameter α_1; that is, $E\{Y(0) \mid Z = 0, G = \text{LL}, X, A\} = m_1(X, A; \alpha_1)$, where $m_1(\cdot, \cdot; \alpha_1)$ is a known function and α_1 is an unknown parameter. Specifically, for the simulations and data example we consider

$$m_1(X, A; \alpha_1) = \alpha_{10} + X^{\text{T}} \alpha_{11} + A \alpha_{12}. \tag{13}$$

Model 2. The expectation $E\{Y(1) \mid Z = 1, X, G, A\}$ is known up to a finite-dimensional parameter α_2; that is, $E\{Y(1) \mid Z = 1, X, G, A\} = m_2(X, G; \alpha_2)$, where G takes values in $\{\text{LL}, \text{LD}\}$, $m_2(\cdot, \cdot; \alpha_2)$ is a known function and α_2 is an unknown parameter. Note that due to Assumption 4, $E\{Y(1) \mid Z = 1, X, G, A\}$ does not depend on the value of A. Specifically, for the simulations and data example we code LL to be 1 and LD to be 0, and consider

$$m_2(X, G; \alpha_2) = \alpha_{20} + X^{\text{T}} \alpha_{21} + G \alpha_{23}. \tag{14}$$

Model 3. The probability $\text{pr}\{S(1) = 1 \mid X, A\}$ is known up to a finite-dimensional parameter β_1; that is, $\text{pr}\{S(1) = 1 \mid X, A\} = \theta_1(X, A; \beta_1)$, where $\theta_1(\cdot, \cdot; \beta_1)$ is a known function and β_1 is an unknown parameter. Specifically, for the simulations and data example we consider

$$\theta_1(X, A; \beta) = \text{expit}(\beta_{10} + X^{\text{T}} \beta_{11} + A \beta_{12}). \tag{15}$$

Model 4. The probability $\text{pr}\{S(0) = 1 \mid X, A\}/\text{pr}\{S(1) = 1 \mid X, A\}$ is known up to a finite-dimensional parameter γ; that is, $\text{pr}\{S(0) = 1 \mid X, A\}/\text{pr}\{S(1) = 1 \mid X, A\} = \theta_{0/1}(X, A; \gamma)$,

where $\theta_{0/1}(\cdot, \cdot; \gamma)$ is a known function and γ is an unknown parameter. Specifically, for the simulations and data example we consider

$$\theta_{0/1}(X, A; \gamma) = \text{expit}(\gamma_0 + X^{\mathrm{T}}\gamma_1 + A\gamma_2). \tag{16}$$

Models 3 and 4 may be replaced by distributional assumptions on $\text{pr}\{S(z) \mid X, A\}$ ($z = 0, 1$) (e.g., Lee et al., 2010). However, constraint (8) implies that $\text{pr}\{S(0) = 1 \mid X, A\}$ lies in the range $[0, \text{pr}\{S(1) = 1 \mid X, A\}]$. Hence the model parameters for $\text{pr}\{S(1) = 1 \mid X, A\}$ live in a constrained space, making estimation and asymptotic analysis difficult. To avoid such constraints, we reparameterize our models as in Models 3 and 4.

To derive the maximum likelihood estimator, we note that Models 1–4 correspond to the following modelling constraints on the observed data distribution:

$$\text{pr}(S = 1 \mid Z = 1, X, A) = \theta_1(X, A; \beta_1),$$
$$\text{pr}(S = 1 \mid Z = 0, X, A) = \theta_1(X, A; \beta_1)\theta_{0/1}(X, A; \gamma),$$
$$E(Y \mid Z = 1, S = 1, X, A) = \theta_{0/1}(X, A; \gamma)m_2(X, 1; \alpha_2) + \{1 - \theta_{0/1}(X, A; \gamma)\}m_2(X, 0; \alpha_2),$$
$$E(Y \mid Z = 0, S = 1, X, A) = m_1(X, A; \alpha_1).$$

These models are compatible with the testable implications described in Proposition 1. Ordinary least squares and maximum likelihood estimation may be used for parameter estimation. The survivor average causal effect Δ_{LL} can then be estimated using (1), in which $\mu_{z,\mathrm{LL},W}$ can be estimated using (13) and (14), $\pi_{\mathrm{LL}|W}$ can be estimated using the product of (15) and (16), and the integration is taken with respect to the empirical distribution of W. Using M-estimation theory one can show that the resulting estimate of Δ_{LL} is consistent and asymptotically normally distributed.

Remark 2. An alternative popular approach to estimating causal effects is based on the propensity score $e(X)$, defined as the probability of assignment to exposure conditioning on baseline covariates. However, although the propensity score is sufficient for summarizing the effect of X on Z in the sense that $Z \perp\!\!\!\perp X \mid e(X)$, it is not sufficient for summarizing the effect of X on A. Thus, propensity score methods are not directly applicable to estimating the survivor average causal effect under our identification assumptions.

5. SIMULATION STUDIES

5·1. *Settings following Assumptions* 1–5

We first consider simulation settings in which data were generated according to Fig. 1. The baseline covariates $X = (X_1, X_2, X_3)$ are a combination of discrete and continuous variables: X_1 is a discrete variable taking values 1 or -1 with probability $1/2$ each, and (X_2, X_3) follows a multivariate normal distribution $N(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0\cdot5 \\ 0\cdot5 & 1 \end{pmatrix}.$$

Conditional on X, the substitution variable A was generated from a Bernoulli distribution such that $\text{pr}(A = 1 \mid X) = \text{expit}(X^{\mathrm{T}}u)$, where $u = (1, 1, 1)^{\mathrm{T}}/2$. The exposure variable Z was generated

Table 2. *Bias* ($\times 100$) *and standard error* ($\times 100$, *in parentheses*) *for various methods of estimating* Δ_{LL} *under settings following Assumptions* 1–5, *in which* Prop-ER *and* Prop-NI *are expected to be consistent. Here* $\delta_1 = 1$ *corresponds to the presence of confounding between* Z *and* Y, *and* $\delta_2 = 1$ *corresponds to the presence of confounding between* S *and* Y; *results are based on* 500 *simulated datasets*

			Exclusion restriction: true			
Sample size	δ_1	δ_2		Estimation method		
			Naive	DGYZ	Prop-ER	Prop-NI
200	0	0	73 (0·61)	3800 (3300)	7·2 (7·7)	32 (1·9)
		1	46 (0·67)	−1100 (1000)	10 (1·2)	2·9 (0·91)
	1	0	80 (0·85)	160 (4·7)	−7·6 (27)	45 (2·4)
		1	40 (0·83)	35 (8·4)	35 (1·5)	28 (1·2)
1000	0	0	73 (0·27)	410 (7·0)	−4·3 (1·5)	8·3 (1·1)
		1	48 (0·29)	−320 (860)	3·3 (0·56)	1·6 (0·41)
	1	0	80 (0·36)	180 (1·6)	−8·3 (3·5)	9·7 (1·3)
		1	41 (0·36)	79 (2·5)	11 (0·97)	9·3 (0·58)
5000	0	0	73 (0·11)	380 (2·1)	−0·16 (0·58)	1·7 (0·48)
		1	48 (0·13)	−1800 (810)	0·72 (0·27)	0·37 (0·21)
	1	0	81 (0·17)	180 (0·65)	−1·1 (0·62)	1·9 (0·61)
		1	41 (0·16)	83 (1·0)	2·6 (0·50)	2·0 (0·24)

Naive, linear regression among observed survivors; DGYZ, the method of Ding et al. (2011); Prop-ER, the proposed method under the exclusion restriction assumption; Prop-NI, the proposed method under the no-interaction assumption.

following a logistic model $\mathrm{pr}(Z = 1 \mid X, A) = \mathrm{expit}(\delta_1 X u + \delta_1 A)$, where δ_1 is a parameter taking values 0 or 1. The survival type G was generated from a multinomial distribution such that $\mathrm{pr}(G = \mathrm{LL} \mid X, A) = \mathrm{expit}(\gamma_0 + X^{\mathrm{T}} \gamma_1 + A \gamma_2) \, \mathrm{expit}(\beta_{10} + X^{\mathrm{T}} \beta_{11} + A \beta_{12})$, $\mathrm{pr}(G = \mathrm{LD} \mid X, A) = \{1 - \mathrm{expit}(\gamma_0 + X^{\mathrm{T}} \gamma_1 + A \gamma_2)\} \, \mathrm{expit}(\beta_{10} + X^{\mathrm{T}} \beta_{11} + A \beta_{12})$ and $\mathrm{pr}(G = \mathrm{DD} \mid X, A) = 1 - \mathrm{expit}(\beta_{10} + X^{\mathrm{T}} \beta_{11} + A \beta_{12})$, where $(\beta_{10}, \beta_{11}, \beta_{12}) = (4, \delta_2, \delta_2, \delta_2, 2)/2$, $(\gamma_0, \gamma_1, \gamma_2) = (0, -3\delta_2, \delta_2, \delta_2, 2)/2$ and δ_2 is a parameter taking values 0 or 1. The potential outcomes $Y(z)$ were generated from the following normal distributions: $Y(1) \mid G = \mathrm{LL}, X, A \sim N(5 + X^{\mathrm{T}} u, 0.5^2)$, $Y(1) \mid G = \mathrm{LD}, X, A \sim N(7 + X^{\mathrm{T}} u, 0.5^2)$ and $Y(0) \mid G = \mathrm{LL}, X, A \sim N(4 + X^{\mathrm{T}} u, 0.5^2)$. The observed survival S and observed outcome Y follow from these by the consistency assumption. Under our settings, the treatment assignment and outcome are confounded when $\delta_1 = 1$, whereas the survival and outcome processes are confounded when $\delta_2 = 1$. The true value for Δ_{LL} is 1.

We compared four methods for estimating Δ_{LL}: linear regression of Y on Z, X and A among observed survivors; the estimation method of Ding et al. (2011); the proposed method under the exclusion restriction assumption (Assumption 4), i.e., the estimation method using models 1–4; the proposed method under the no-interaction assumption (Assumption 7), i.e., the estimation method using models 3, 4 and 7. We differ from Ding et al. (2011) in both the identification conditions and the estimation method. Table 2 summarizes the results. The naive regression method is biased in all settings because of selection bias. In the presence of confounding between Z and Y or between S and Y, i.e., when $\delta_1 \neq 0$ or $\delta_2 \neq 0$, the estimator of Ding et al. (2011) is inconsistent. However, even in the absence of confounding between Z and Y or between S and Y, i.e., when $\delta_1 = \delta_2 = 0$, the estimator of Ding et al. (2011) can be unstable, especially if the sample size is small. Although in this case the substitution relevance assumption of Ding et al. (2011) holds since $\mathrm{pr}(G = \mathrm{LL} \mid Z = 1, S = 1, A = 1) \neq \mathrm{pr}(G = \mathrm{LL} \mid Z = 1, S = 1, A = 0)$, for some simulated samples the estimates of $\mathrm{pr}(G = \mathrm{LL} \mid Z = 1, S = 1, A = 1)$ can be very close to the estimates of $\mathrm{pr}(G = \mathrm{LL} \mid Z = 1, S = 1, A = 0)$, leading to instability in the

Table 3. *Bias* (×100) *and standard error* (×100, *in parentheses*) *for various methods of esti-mating* Δ_{LL} *under settings in which the exclusion restriction assumption is violated. In these settings* Prop-NI *is expected to be consistent but* Prop-ER *is not. Here* $\delta_1 = 1$ *corresponds to the presence of confounding between* Z *and* Y, *and* $\delta_2 = 1$ *corresponds to the presence of confounding between* S *and* Y; *results are based on* 500 *simulated datasets*

			Exclusion restriction: false			
Sample size	δ_1	δ_2		Estimation method		
			Naive	DGYZ	Prop-ER	Prop-NI
200	0	0	73 (0·61)	5500 (4700)	160 (11)	32 (1·9)
		1	46 (0·67)	−1700 (1500)	16 (1·7)	2·9 (0·91)
	1	0	80 (0·85)	66 (7·6)	140 (13)	45 (2·4)
		1	40 (0·83)	−140 (13)	54 (2·0)	28 (1·2)
1000	0	0	73 (0·27)	580 (10)	160 (1·9)	8·3 (1·1)
		1	48 (0·29)	−540 (1300)	8·1 (0·62)	1·6 (0·41)
	1	0	80 (0·36)	94 (2·4)	170 (5·3)	9·7 (1·3)
		1	41 (0·36)	−62 (3·8)	31 (1·6)	9·3 (0·58)
5000	0	0	73 (0·11)	540 (3·1)	160 (0·69)	1·7 (0·48)
		1	48 (0·13)	−2800 (1200)	4·7 (0·27)	0·37 (0·21)
	1	0	81 (0·17)	95 (0·94)	160 (0·77)	1·9 (0·61)
		1	41 (0·16)	−56 (1·6)	20 (0·47)	2·0 (0·24)

causal effect estimates. In contrast, the proposed estimators are more stable, and have biases that decrease with sample size in all the settings considered here.

5·2. *Settings with violation of exclusion restriction*

We now evaluate the sensitivity of the proposed method to departures from the exclusion restriction assumption, Assumption 4. We generated data in the same way as in § 5·1, except that the potential outcome $Y(z)$ was generated from the following normal distributions: $Y(1) \mid G = \mathrm{LL}, X, A \sim N(5 + X^{\mathrm{T}}u + A, 0\cdot5^2)$, $Y(1) \mid G = \mathrm{LD}, X, A \sim N(7 + X^{\mathrm{T}}u + A, 0\cdot5^2)$ and $Y(0) \mid G = \mathrm{LL}, X, A \sim N(4 + X^{\mathrm{T}}u + A, 0\cdot5^2)$. The method of Ding et al. (2011) for dealing with exclusion restriction violations is not applicable here, as it requires the substitution variable A to either be continuous or have at least three categories.

The simulation results are presented in Table 3. As expected, the proposed estimator assuming no interaction is consistent for all simulation settings considered here, whereas both the estimator of Ding et al. (2011) and the proposed estimator assuming the exclusion restriction assumption are biased for estimating Δ_{LL}.

6. APPLICATION TO A SOUTHWEST ONCOLOGY GROUP TRIAL

We illustrate the application of the proposed methods using data from a randomized phase III trial to compare docetaxel plus estramustine with mitoxantrone plus prednisone in men with metastatic, hormone-independent prostate cancer (Petrylak et al., 2004). The dataset we use, which was created by Ding et al. (2011), contains observations on 487 men aged 47 to 88. Of these subjects, 258 were randomly assigned to receive docetaxel plus estramustine and 229 were randomly assigned to receive mitoxantrone plus prednisone. In our analysis, we are interested in comparing these two treatments in terms of health-related quality of life one year after receiving treatment.

Table 4. *Survivor average causal effect of docetaxel plus estramustine treatment on health-related quality of life*

Estimation method	Point estimate	Bootstrapped SE	2·5%	50%	97·5%
DGYZ	7·01	3·09	1·81	6·56	13·64
Prop-ER	3·06	11·79	−15·15	3·29	22·60
Prop-NI	2·73	3·82	−3·97	2·95	10·83

DGYZ, results adapted from Ding et al. (2011); Prop-ER, results estimated using the proposed method under the exclusion restriction assumption; Prop-NI, results estimated using the proposed method under the no-interaction assumption.

A naive analysis shows that among patients who survived to one year after receiving the assigned treatment, the quality of life for those assigned to the docetaxel plus estramustine group is higher by 2·46 units than those assigned to the mitoxantrone plus prednisone group; the 95% confidence interval is $[−3·31, 8·24]$. However, this estimate is not causally interpretable, as subjects who would survive if assigned to docetaxel plus estramustine are potentially different from subjects who would survive if assigned to mitoxantrone plus prednisone. Moreover, as reported by Petrylak et al. (2004), docetaxel plus estramustine is beneficial for overall survival compared to mitoxantrone plus prednisone. The direct comparison among observed survivors is therefore also subject to selection bias. Instead, we apply the method of Ding et al. (2011) and our proposed methods to deal with truncation due to death. To account for possible common causes of the survival and outcome processes, we adjust for the following variables in the proposed methods: age, race, type of prognosis, bone pain and performance status. Following Ding et al. (2011), we use the baseline quality of life as the substitution variable, and the change in quality of life over the one-year period as the outcome.

We first analyse this dataset under the monotonicity assumption. The point estimates and bootstrap standard errors are displayed in Table 4. The results in Ding et al. (2011) suggest that docetaxel plus estramustine had a significant causal effect on the quality of life among those patients who would survive one year after receiving treatment regardless of which treatment group they were assigned to. In contrast, after accounting for the baseline covariate information that might simultaneously impact the potential survival and the potential quality of life, we were not able to reach such a conclusion. Both of the proposed methods yield a point estimate that is much closer to zero, and their 95% confidence intervals cover zero. These results show that, even in the setting of a randomized trial, adjusting for baseline covariates can lead to different estimates for the survivor average causal effect; this should not be surprising since baseline variables may confound S and Y.

The proposed estimator Prop-ER, which assumes exclusion restriction and the monotonicity condition, is very unstable. In particular, the point estimate for the model parameter γ in Model 4 is large in magnitude, so that some fitted values of $\theta_{0/1}$ are close to the boundary. This indicates a monotonicity violation. In fact, previous analyses also suggest that the monotonicity assumption could be problematic for this dataset (Ding et al., 2011), and as we discussed earlier, for randomized trials the monotonicity assumption is not plausible a priori. For these reasons, although the overall one-year survival rate in the docetaxel plus estramustine group, 49·6% is higher than that in the mitoxantrone plus prednisone group, 38·9%, a sensitivity analysis of the monotonicity assumption is warranted. In our analysis the sensitivity parameter is π_{DL}, which has a one-to-one correspondence with the sensitivity parameter ρ but has the interpretation as the overall fraction of patients who would die one year after receiving docetaxel plus estramustine but would survive one year after receiving mitoxantrone plus prednisone. Under our Models 3
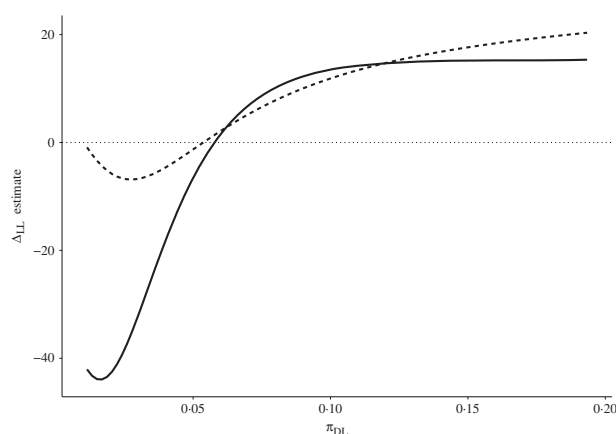
Fig. 2. Sensitivity analysis for estimating the survivor average causal effect in the Southwest Oncology Group dataset; the solid line represents the proposed method assuming exclusion restriction, and the dashed line the proposed method assuming no interaction.

and 6, the range for π_{DL} is estimated to be $[0\cdot01, 0\cdot19]$, with the lower and upper limits corresponding to $\rho = 1$ and 0, respectively. As the lower limit is greater than 0, the monotonicity assumption, Assumption 1, is not compatible with these modelling assumptions. Figure 2 summarizes the results from our sensitivity analysis. As both of the sensitivity lines cross zero, without prior knowledge of the possible values for π_{DL}, we cannot reach any definitive answer about the causal effect of docetaxel plus estramustine versus mitoxantrone plus prednisone on the quality of life one year after receiving treatment. Furthermore, the two curves from the proposed methods agree approximately unless π_{DL} is smaller than 0·05. We also note that under a different set of assumptions, Ding et al. (2011) obtained a value of π_{DL} that is much greater than our upper bound. Under our models 3 and 6, such a value is incompatible with the stochastic monotonicity assumption (Assumption 6) and would imply a negative correlation between the potential survival under docetaxel plus estramustine and the potential survival under mitoxantrone plus prednisone.

7. DISCUSSION

In this article, we have considered the identification and estimation of the survivor average causal effect. Compared with previous works that attempted to identify the survivor average causal effect using an exogenous variable, our method adjusts for baseline covariates X, which may be common causes between any pair of treatment, survival and outcome. This role for baseline covariates is unique to our identification framework for the survivor average causal effect. In contrast to the standard estimation problem of the average causal effect, even in randomized studies, the inclusion of baseline covariates is crucial for obtaining unbiased estimates of the survivor average causal effect, since there is likely to be confounding between survival and the post-survival outcome.

In our estimation approach, instead of imposing modelling assumptions on the observed data, we use parametric models for the potential outcomes. This not only simplifies the estimation procedure but also makes it easier to ensure compatibility between our identification and modelling assumptions. We also propose alternative methods to relax the monotonicity assumption and the exclusion restriction assumption. Formal inference under our alternative assumption,

Assumption 6, requires a sensitivity analysis framework that accounts for uncertainty due to both nonidentifiability and sampling variability, such as the ones based on ignorance and uncertainty regions described by Vansteelandt et al. (2006) and others. To focus on the main ideas here, we present only point estimates under Assumption 6 for fixed values of ρ.

In the current work we have considered only the binary exposure case. In practical medical studies, the exposure variable may have multiple levels. For example, the smoking variable may be coded as never smokers, past smokers and current smokers. It would therefore be interesting to generalize the proposed methods to deal with ordinal exposure status.

Acknowledgement

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of the theorems and propositions, as well as alternative causal diagrams that are compatible with the structural equations (4) but which allow for certain dependencies between error terms.

Appendix

Parameterizations under alternative identification assumptions

To estimate Δ_{LL} under the assumptions of Theorem 2, we assume Models 2 and 3 and the following:

Model 5. $E\{Y(0) \mid Z = 0, X, G, A\}$ is known up to a finite-dimensional parameter α_3; that is, $E\{Y(0) \mid Z = 0, X, G, A\} = m_3(X, G; \alpha_3)$, where G takes values in $\{\mathrm{LL}, \mathrm{DL}\}$, $m_3(\cdot, \cdot; \alpha_3)$ is a known function and α_3 is an unknown parameter. Note that because of condition (i) of Theorem 2, $m_3(\cdot, \cdot; \alpha_3)$ does not depend on the value of A. Specifically, for the simulations and data example we code LL to be 1 and DL to be 0, and consider

$$m_3(X, G; \alpha_3) = \alpha_{30} + X^{\mathrm{T}}\alpha_{31} + G\alpha_{33};$$

Model 6. $\mathrm{pr}\{S(0) = 1 \mid X, A\}$ is known up to a finite-dimensional parameter β_0; that is, $\mathrm{pr}\{S(0) = 1 \mid X, A\} = \theta_0(X, A; \beta_0)$, where $\theta_0(\cdot, \cdot; \beta_0)$ is a known function and β_0 is an unknown parameter. Specifically, for the data example we consider

$$\theta_0(X, A; \beta_0) = \mathrm{expit}(\beta_{00} + X^{\mathrm{T}}\beta_{01} + A\beta_{02}).$$

Similarly, to estimate Δ_{LL} under the assumptions of Theorem 3, we assume Models 3 and 4 and the following model:

Model 7. $E(Y \mid X, A, G, Z)$ is known up to a finite-dimensional parameter α_4; that is, $E[Y \mid X, A, G, Z] = m_4(X, A, G, Z; \alpha_4)$, where G takes values in $\{LL, LD\}$, $m_4(\cdot, \cdot, \cdot, \cdot; \alpha_4)$ is a known function and α_4 is an unknown parameter. Note that by Assumption 7, $m_4(X, A, G, Z; \alpha_4)$ should not contain interaction terms within the pairs (A, Z) or (A, G). Specifically, for the simulations and data example we code LL to be 1 and LD to be 0, and consider

$$m_4(X, A, G, Z; \alpha_4) = \alpha_{40} + X^{\mathrm{T}}\alpha_{41} + A\alpha_{42} + G\alpha_{43} + Z\alpha_{44}.$$

Finally, if one wishes to relax both the monotonicity assumption, Assumption 1, and the exclusion restriction, Assumption 4, one may assume Models 3 and 6 and the following model:

Model 8. $E(Y \mid X, A, G, Z)$ is known up to a finite-dimensional parameter α_5; that is, $E(Y \mid X, A, G, Z) = m_5(X, A, G, Z; \alpha_5)$, where G takes values in $\{LL, LD, DL\}$, $m_5(\cdot, \cdot, \cdot, \cdot; \alpha_5)$ is a known function and α_5 is an unknown parameter. Note that by Assumption 7, $m_5(X, A, G, Z; \alpha_5)$ should not contain interaction terms within the pairs (A, Z) or (A, G). Specifically, for the data example we code LL to be 1 and DL and LD to be 0, and consider

$$E(Y \mid X, A, G, Z) = \alpha_{50} + X^{\mathrm{T}}\alpha_{51} + A\alpha_{52} + ZG\alpha_{53} + Z\alpha_{54} + (1 - Z)G\alpha_{55}.$$

REFERENCES

ANGRIST, J. D., IMBENS, G. W. & RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Am. Statist. Assoc.* **91**, 444–55.

CHIBA, Y. & VANDERWEELE, T. J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *Am. J. Epidemiol.* **173**, 745–51.

DING, P., GENG, Z., YAN, W. & ZHOU, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *J. Am. Statist. Assoc.* **106**, 1578–91.

DING, P. & VANDERWEELE, T. J. (2014). Generalized Cornfield conditions for the risk difference. *Biometrika* **101**, 971–7.

EGLESTON, B. L., SCHARFSTEIN, D. O., FREEMAN, E. E. & WEST, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8**, 526–45.

FRANGAKIS, C. E. & RUBIN, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86**, 365–79.

FRANGAKIS, C. E. & RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–9.

GILBERT, P. B., BOSCH, R. J. & HUDGENS, M. G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59**, 531–41.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–54.

HAYDEN, D., PAULER, D. K. & SCHOENFELD, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61**, 305–10.

IMAI, K. (2008). Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statist. Prob. Lett.* **78**, 144–9.

IMBENS, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706–10.

IMBENS, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *Am. Econ. Rev.* **93**, 126–32.

JEMIAI, Y., ROTNITZKY, A., SHEPHERD, B. E. & GILBERT, P. B. (2007). Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs. *J. R. Statist. Soc. B* **69**, 879–901.

KURLAND, B. F., JOHNSON, L. L., EGLESTON, B. L. & DIEHR, P. H. (2009). Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Statist. Sci.* **24**, 211–22.

LEE, K., DANIELS, M. J. & SARGENT, D. J. (2010). Causal effects of treatments for informative missing data due to progression/death. *J. Am. Statist. Assoc.* **105**, 912–29.

LONG, D. M. & HUDGENS, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics* **69**, 812–9.

PEARL, J. (2009). *Causality*. Cambridge: Cambridge University Press.

PETRYLAK, D. P., TANGEN, C. M., HUSSAIN, M. H. A., LARA JR, P. N., JONES, J. A., TAPLIN, M. E., BURCH, P. A., BERRY, D., MOINPOUR, C., KOHLI, M. et al. (2004). Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *New Engl. J. Med.* **351**, 1513–20.

ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Mod.* **7**, 1393–512.

ROBINS, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In *Highly Structured Stochastic Systems*, P. J. Green, N. L. Hjort & S. Richardson, eds. New York: Oxford University Press, pp. 70–81.

ROBINS, J. M. & GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–55.

ROBINS, J. M. & RICHARDSON, T. S. (2010). Alternative graphical causal models and the identification of direct effects. In *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, P. Shrout, K. M. Keyes & K. Ornstein, eds. New York: Oxford University Press, pp. 103–58.

ROSENBAUM, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J. R. Statist. Soc.* A **147**, 656–66.

ROSENBAUM, P. R. & RUBIN, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. R. Statist. Soc.* B **45**, 212–8.

ROY, J., HOGAN, J. W. & MARCUS, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9**, 277–89.

RUBIN, D. B. (1980). Comment. *J. Am. Statist. Assoc.* **75**, 591–3.

RUBIN, D. B. (1998). More powerful randomization-based p-values in double-blind trials with non-compliance. *Statist. Med.* **17**, 371–85.

RUBIN, D. B. (2000). Comment on "Causal inference without counterfactuals". *J. Am. Statist. Assoc.* **95**, 435–8.

RUBIN, D. B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Statist. Sci.* **21**, 299–309.

SHEPHERD, B. E., GILBERT, P. B., JEMIAI, Y. & ROTNITZKY, A. (2006). Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics* **62**, 332–42.

SPERLING, R., MORMINO, E. & JOHNSON, K. (2014). The evolution of preclinical Alzheimer's disease: Implications for prevention trials. *Neuron* **84**, 608–22.

TCHETGEN TCHETGEN, E. J. (2014). Identification and estimation of survivor average causal effects. *Statist. Med.* **33**, 3601–28.

VANSTEELANDT, S., GOETGHEBEUR, E., KENWARD, M. G. & MOLENBERGHS, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sinica* **16**, 953–79.

WEUVE, J., TCHETGEN TCHETGEN, E. J., GLYMOUR, M. M., BECK, T. L., AGGARWAL, N. T., WILSON, R. S., EVANS, D. A. & DE LEON, C. F. M. (2012). Accounting for bias due to selective attrition: The example of smoking and cognitive decline. *Epidemiology* **23**, 119–28.

ZHANG, J. L. & RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J. Educ. Behav. Statist.* **28**, 353–68.