Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death

MICHELLE SHARDELL*

*Department of Epidemiology and Public Health, University of Maryland 660 West Redwood Street,
Baltimore, MD 21201, USA*

mshardel@epi.umaryland.edu

GREGORY E. HICKS

Department of Physical Therapy, University of Delaware 303 McKinly Lab, Newark, DE 19716, USA

LUIGI FERRUCCI

National Institute on Aging, 3001 S Hanover Street, Baltimore, MD 21225, USA

SUMMARY

Motivated by aging research, we propose an estimator of the effect of a time-varying exposure on an outcome in longitudinal studies with dropout and truncation by death. We use an inverse-probability weighted (IPW) estimator to derive a doubly robust augmented inverse-probability weighted (AIPW) estimator. IPW estimation involves weights for the exposure mechanism, dropout, and mortality; AIPW estimation additionally involves estimating data-generating models via regression. We demonstrate that the estimators identify a causal contrast that is a function of principal strata effects under a set of assumptions. Simulations show that AIPW estimation is unbiased when weights or outcome regressions are correct, and that AIPW estimation is more efficient than IPW estimation when all models are correct. We apply the method to a study of vitamin D and gait speed among older adults.

Keywords: Causal inference; Longitudinal data analysis, Missing data; Observational studies.

1. INTRODUCTION

Longitudinal studies in aging often involve time-varying confounding, dropout, and truncation by death. Modern methods (e.g. [Robins, 1998, 1999](#); [Robins and others, 2000](#)) handle the first two issues, but truncation by death is more vexing. Dropout and death both lead to attrition, but dropout renders outcomes unobserved whereas death renders them undefined.

We are motivated to estimate the effect of vitamin D on physical function in the Invecchiare in Chianti study (“Aging in Chianti”; shortened, “InChianti”), a longitudinal study of decline in old age comprising older Italians in which many participants dropped out or died before the study ended ([Ferrucci and others, 2000](#)). Vitamin D is a timely topic due to the 2011 Institute of Medicine (IoM) report on vitamin D reference

*To whom correspondence should be addressed.

intakes (Institute of Medicine, 2011); the Endocrine Society’s guidelines (Holick and others, 2011); and ensuing debate (Holick and others, 2012; Rosen and others, 2012). Both groups agree that more evidence is needed linking vitamin D to physical function, a key outcome for older adults, but they disagree on the target level for serum 25-hydroxyvitamin D [25(OH)D] (the gold standard vitamin D measure) and rationale thereof. The IoM recommends 25(OH)D ≥ 20 ng/mL for bone health and cites insufficient evidence for improved extra-skeletal health at higher levels, but the Endocrine Society recommends ≥ 30 ng/mL for bone health and cites evidence for safety at this level. Given the need for evidence on extra-skeletal health (e.g. physical function) and lack of a formal causal inference of longitudinal studies on this topic, we aim to estimate the effect of serum 25(OH)D on physical function using guidelines from the IoM (≥ 20 ng/mL) and Endocrine Society (≥ 30 ng/mL). Since 25(OH)D in InChianti is time-varying, a method is needed to address this issue along with dropout and death.

Multiple methods have been proposed to handle truncation by death in analysis. Conditioning on being alive (Kurland and Heagerty, 2005; Shardell and Miller, 2008; Shardell and others, 2010) and unconditional analysis (Rajan and Leurgans, 2010) estimate non-causal associations, namely, mean differences in physical function by 25(OH)D among survivors, and the full cohort, respectively. In contrast, principal stratification (PS) estimates the “survivor average causal effect” (SACE), a causal contrast in a latent subgroup who would survive regardless of exposure (Eggleston and others, 2007; Frangakis and Rubin, 2002; Hayden and others, 2005; Lee and others, 2010; Rubin, 2006). PS is well motivated by mechanistic studies (Jemai and others, 2007; Shepherd and others, 2006), but some question if SACE is of genuine scientific interest or a stop-gap to avoid ill-defined counterfactuals of vital status (Pearl, 2011).

Ill-defined counterfactuals are why few options exist for causal inference with truncation by death, which forms a barrier to applying formal causal inference to studies with high mortality. However, some have used inverse-probability weighting (IPW) of survivors (e.g. Glymour and others, 2010; HIV-CAUSAL Collaboration, 2012). A concern is that IPW estimation requires counterfactuals of vital status. Tchetgen Tchetgen and others (2012a), hereafter TGSW, showed that in point exposure studies, IPW can estimate a causal contrast without counterfactuals of vital status under a set of assumptions. In this paper, we extend TGSW’s exposition to longitudinal studies.

Unbiased IPW estimation requires correctly modeling dropout, mortality, and exposure weights. Thus, we propose doubly robust augmented IPW (AIPW) estimation using outcome regressions. AIPW estimation is consistent when weights or outcome regressions are correct (Bang and Robins, 2005; Robins, 2000), and AIPW is more efficient than IPW when all models are correct.

2. METHODS

Consider a study of N participants with a baseline visit and $J + 1$ follow-up visits. Let $A_i(j)$ be participant i ’s exposure at visit j , $j = 0, \dots, J$; and $Y_i(j)$ be the outcome at visit j , $j = 0, \dots, J + 1$. X_i are baseline covariates, and $L_i(j)$ are time-varying covariates at visit j . Let $L(0)$ subsume $\{X, Y(0)\}$. Also, $Z_i(j)$ is vital status at visit j , $j = 0, \dots, J + 1$, where $Z_i(j) = 1$ if alive, $Z_i(j) = 0$ if dead, and $Z_i(0) = 1$ for all i . $Z_i(j) = 0$ implies $Z_i(k) = 0$ for $k \geq j$. If $Z_i(j) = 0$, then $L_i(k)$, $A_i(k)$, and $Y_i(k)$ are undefined for $k \geq j$. Finally, $R_i(j)$ is a response indicator at visit j , where $R_i(j) = 1$ if data are observed, $R_i(j) = 0$ if data are missing, and $R_i(0) = 1$ for all i . We address dropout, where $R_i(j) = 0$ implies $R_i(k) = 0$ for $k \geq j$. Suppressing subscript i , we assume data are generated in the order $L(0), A(0), R(0), Z(1), Y(1), L(1), A(1), R(1), \dots, L(J), A(J), R(J), Z(J + 1), Y(J + 1), R(J + 1)$. Let overbars denote variable history; for example, $\bar{A}(j) = \{A(0), A(1), \dots, A(j)\}$ is exposure history through visit j .

Fig. 1. Directed acyclic graph of the hypothesized pathway linking $A(0)$ to $Y(1)$. $L(0)$ comprises $A(0) - Y(1)$ confounders, $Z(1)$ is vital status, and $R(1)$ is observation status, where the absence of an arrow from $Y(1)$ to $R(1)$ encodes missing at random. The bold line is the causal pathway through which $A(0)$ can affect $Y(1)$. Dashed lines represent relations between $Z(1)$ and subsequent variables that are not of scientific interest. Under this DAG, a completers-only analysis would need to condition on $L(0)$ to remove selection bias.

2.1 Target of estimation and assumptions

We focus our exposition on the first two follow-up visits of a study to reflect the design of InChianti. We aim to estimate the causal effects of $A(0)$ on $Y(1)$ and of $\bar{A}(1)$ on $Y(2)$. We proceed by making two core assumptions: (i) no unmeasured $A - Y$ confounders and no unmeasured common causes of Z and Y and (ii) survivors' data are missing at random (MAR) (Rubin, 1976). Directed acyclic graphs (DAGs) help to exemplify these assumptions; the DAG in Figure 1 encodes a pathway linking $A(0)$ to $Y(1)$ that satisfies MAR among survivors,

$$\Pr\{R(1) = 1 \mid L(0), A(0), Y(1), Z(1) = 1\} = \Pr\{R(1) = 1 \mid L(0), A(0), Z(1) = 1\}.$$

A conventional completers-only analysis is biased by conditioning on a collider $\{Z(1)R(1) = 1\}$, an effect of multiple causes or its descendant (Hernán and others, 2004). Collider stratification bias is avoidable in Figure 1 by conditioning on $L(0)$. TGSW and Supplementary material available at *Biostatistics* online show a case where no conditioning set can remove bias.

Turning to longitudinal studies, Figure 2 shows a DAG with two follow-up visits that satisfies the MAR assumption among uncensored survivors:

$$\Pr\{R(2) = 1 \mid \bar{L}(1), \bar{A}(1), \bar{Y}(2), Z(2)R(1) = 1\} = \Pr\{R(2) = 1 \mid \bar{L}(1), \bar{A}(1), \bar{Y}(1), Z(2)R(1) = 1\}.$$

No conditioning set produces an unbiased completers-only estimate of the effect of $\bar{A}(1)$ on $Y(2)$ because $\{Y(1), L(1)\}$ is an effect of $A(0)$ and is a common cause of $Y(2)$ and $Z(2)$; that is, $\{Y(1), L(1)\}$ mediates the effect of $A(0)$ on both $Y(2)$ and $Z(2)$. An alternative method to overcome collider-stratification bias involves specifying the g-formula of Robins (1986). Estimators of the g-formula, such as IPW, have well known theoretical properties (Robins, 1986, 1998) and include computationally accessible options. Thus, in this section we specify a g-formula that addresses truncation by death. In Section 2.2, we propose a causal model and derive the interpretation of resulting causal contrasts identified by the g-formula.

Let $\bar{V}(1) = \{\bar{Y}(1), \bar{L}(1)\}$ and $\bar{v}(1) = \{\bar{y}(1), \bar{l}(1)\}$. We specify g-formulas for the first two follow-up visits:

$$g\{a(0)\} = \int E\{Y(1) \mid A(0) = a(0), Z(1) = 1, \bar{L}(0) = \bar{l}(0)\} \times dF_{L(0)}\{\bar{l}(0)\}, \quad (2.1)$$

$$g\{\bar{a}(1)\} = \int E\{Y(2) \mid \bar{A}(1) = \bar{a}(1), Z(2) = 1, \bar{V}(1) = \bar{v}(1)\} \times dF_{\bar{V}(1)}\{\bar{v}(1)\}, \quad (2.2)$$

Fig. 2. Directed acyclic graph expanded to two follow-up visits. The most general causal diagram is shown under the assumption of missing at random, where a variable may affect any future variable with the exception that $\{L(j+1), A(j+1), Y(j+1)\}$ are assumed not to affect $R(j+1)$ as indicated by the absence of arrows from $\{L(j+1), A(j+1), Y(j+1)\}$ to $R(j+1)$, $j = 0, 1$. Bold lines show that $A(0)$ can affect $Y(2)$ through both a direct pathway and an indirect pathway mediated by $\{Y(1), L(1)\}$. Under this DAG, no conditioning set is sufficient to remove selection bias in a conventional completers-only analysis.

where

$$\begin{aligned} dF_{\bar{Y}(1)}\{\bar{y}(1)\} &= dF_{L(1)|A(0)=a(0), Z(1)=1, L(0)=l(0), \bar{Y}(1)=\bar{y}(1)}\{l(1)\} \\ &\times dF_{Y(1)|A(0)=a(0), Z(1)=1, L(0)=l(0)}\{y(1)\} \times dF_{L(0)}\{l(0)\}. \end{aligned} \quad (2.3)$$

$F_{L(1)|A(0)=a(0), Z(1)=1, L(0)=l(0), \bar{Y}(1)=\bar{y}(1)}\{l(1)\}$ is the conditional distribution of $L(1)$ at $l(1)$; other distributions are similarly defined. By MAR, completers-only analyses to estimate expectations in (2.1) and (2.2) and to estimate the first two terms of (2.3) can be unbiased. Relations of $\bar{A}(j)$ with $Y(j+1)$ are contrasts $g\{\bar{a}(j)\} - g\{\bar{a}'(j)\}$, comparing exposure histories $\bar{a}(j)$ and $\bar{a}'(j)$. However, these contrasts' causal interpretation depends on assumptions about mortality, due to conditioning on $Z(j+1) = 1$. In Section 2.2, we show that without additional modeling assumptions, vital status would be treated as an intervention, with contrasts $g\{\bar{a}(j)\} - g\{\bar{a}'(j)\}$ interpreted as controlled direct effects of $\bar{A}(j)$ on $Y(j+1)$ setting $Z(j) = 1$. We thus propose a model to identify an alternative interpretation of $g\{\bar{a}(j)\} - g\{\bar{a}'(j)\}$ without treating vital status as an intervention. Supplementary material available at [Biostatistics](#) online shows $g\{\bar{a}(j)\}$ for general j .

2.2 Causal interpretation

To derive the causal interpretation of $g\{\bar{a}(j)\} - g\{\bar{a}'(j)\}$, we propose nonparametric structural equation models (Pearl, 2009) as in TGSW and extend the authors' exposition to repeated measures. These models satisfy the two core assumptions and those encoded by Figure 2 for $j = 0, \dots, J$:

$$L(j) = \begin{cases} g_{L(j)}\{\bar{A}(j-1), \bar{L}(j-1), \bar{Y}(j), \epsilon_{L(j)}\} & \text{if } Z(j) = 1 \\ \text{undefined} & \text{if } Z(j) = 0 \end{cases} \quad (2.4)$$

$$A(j) = \begin{cases} g_{A(j)}\{\bar{A}(j-1), \bar{Y}(j), \bar{L}(j), \epsilon_{A(j)}\} & \text{if } Z(j) = 1 \\ \text{undefined} & \text{if } Z(j) = 0 \end{cases} \quad (2.5)$$

$$Z(j+1) = \begin{cases} g_{Z(j+1)}\{\bar{A}(j), \bar{Y}(j), \bar{L}(j), \epsilon_{Z(j+1)}\} & \text{if } Z(j) = 1 \\ 0 & \text{if } Z(j) = 0 \end{cases} \quad (2.6)$$

$$Y(j+1) = \begin{cases} g_{Y(j+1)}\{\bar{A}(j), \bar{Y}(j), \bar{L}(j), \epsilon_{Y(j+1)}\} & \text{if } Z(j+1) = 1 \\ \text{undefined} & \text{if } Z(j+1) = 0. \end{cases} \quad (2.7)$$

Right-side variables of (2.4–2.7) are presumed causes of left-side variables; errors $\{\epsilon_X, \epsilon_{Y(0)}, \epsilon_{L(j)}, \epsilon_{A(j)}, \epsilon_{Z(j+1)}, \epsilon_{Y(j+1)}\}$ are independent exogenous causes. The equations explicate that $L(j)$, $A(j)$, and $Y(j+1)$ are only defined for survivors. By MAR, (2.4–2.6) and (2.7) equal the respective functions given $R(j) = 1$ and $R(j+1) = 1$; we therefore suppress notation for R . Like others, we define causes as interventions (Pearl, 2009). We encode an intervention setting $\bar{A} = \bar{a}$ by deleting arrows into $\bar{A}(1)$ in Figure 2, replacing $g_{A(j)}$ with $a(j)$ in (2.5), and replacing observables with counterfactuals in (2.4–2.7). Subscripts of $\bar{a}(j)$ denote counterfactuals of exposure. These changes result in a set of equations of counterfactual variables:

$$L_{\bar{a}(j-1)}(j) = \begin{cases} g_{L(j)}\{\bar{a}(j-1), \bar{L}_{\bar{a}(j-2)}(j-1), \bar{Y}_{\bar{a}(j-1)}(j), \epsilon_{L(j)}\} & \text{if } Z_{\bar{a}(j-1)}(j) = 1 \\ \text{undefined} & \text{if } Z_{\bar{a}(j-1)}(j) = 0 \end{cases} \quad (2.8)$$

$$A(j) = \begin{cases} a(j) & \text{if } Z_{\bar{a}(j-1)}(j) = 1 \\ \text{undefined} & \text{if } Z_{\bar{a}(j-1)}(j) = 0 \end{cases}$$

$$Z_{\bar{a}(j)}(j+1) = \begin{cases} g_{Z(j+1)}\{\bar{a}(j), \bar{Y}_{\bar{a}(j-1)}(j), \bar{L}_{\bar{a}(j-1)}(j), \epsilon_{Z(j+1)}\} & \text{if } Z_{\bar{a}(j-1)}(j) = 1 \\ 0 & \text{if } Z_{\bar{a}(j-1)}(j) = 0 \end{cases} \quad (2.9)$$

$$Y_{\bar{a}(j)}(j+1) = \begin{cases} g_{Y(j+1)}\{\bar{a}(j), \bar{Y}_{\bar{a}(j-1)}(j), \bar{L}_{\bar{a}(j-1)}(j), \epsilon_{Y(j+1)}\} & \text{if } Z_{\bar{a}(j)}(j+1) = 1 \\ \text{undefined} & \text{if } Z_{\bar{a}(j)}(j+1) = 0. \end{cases} \quad (2.10)$$

Error terms in (2.8), (2.9) and (2.10) are unchanged from (2.4), (2.6), and (2.7), respectively, because errors are presumed to be exogenous. The individual causal effect of $\bar{a}(j)$ on $Y(j+1)$ is the contrast $Y_{\bar{a}(j)}(j+1) - Y_{\bar{a}'(j)}(j+1)$, which is only defined in the principal stratum $\{Z_{\bar{a}(j)}(j+1) = Z_{\bar{a}'(j)}(j+1) = 1\}$ and equals

$$g_{Y(j+1)}\{\bar{a}(j), \bar{Y}_{\bar{a}(j-1)}(j), \bar{L}_{\bar{a}(j-1)}(j), \epsilon_{Y(j+1)}\} - g_{Y(j+1)}\{\bar{a}'(j), \bar{Y}_{\bar{a}'(j-1)}(j), \bar{L}_{\bar{a}'(j-1)}(j), \epsilon_{Y(j+1)}\}. \quad (2.11)$$

The model implies $Y_{\bar{a}(j)}(j+1) \perp Z_{\bar{a}'(j)}(j+1) \mid \{Z_{\bar{a}(j)}(j+1), \bar{Y}_{\bar{a}(j-1)}(j), \bar{L}_{\bar{a}(j-1)}(j)\}$, $\bar{a}(j) \neq \bar{a}'(j)$, or “cross-world” independence, where $A \perp B \mid C$ means “ A independent of B given C .” Thus, the average effect in principal stratum $\{Z_{\bar{a}(j)}(j+1) = Z_{\bar{a}'(j)}(j+1) = 1\}$ is an expectation of (2.11).

Deriving average PS effects of $\bar{A}(j)$ on $Y(j+1)$ requires counterfactuals of $\bar{A}(j)$ and components of $\{\bar{L}(j), \bar{Y}(j)\}$ that mediate the effect of $\bar{A}(j)$ on both $Y(j+1)$ and $Z(j+1)$. First consider the effect of $A(0)$ on $Y(1)$ in Figure 1, where no variable mediates the effect of $A(0)$ on both $Y(1)$ and $Z(1)$. Replacing observables with counterfactuals in (2.5–2.7) setting $A(0) = a(0)$ implies

$$Y_{a(0)}(1) \perp Z_{a'(0)}(1) \mid \{Z_{a(0)}(1), L(0)\} \quad \text{for } a(0) \neq a'(0). \quad (2.12)$$

An implication of (2.12) is that the PS direct effect, given $L(0)$, is

$$\begin{aligned} E\{Y_{a(0)}(1) - Y_{a'(0)}(1) \mid Z_{a(0)}(1) = Z_{a'(0)}(1) = 1, L(0)\} \\ = E[g_{Y(1)}\{a(0), L(0), \epsilon_{Y(1)}\} - g_{Y(1)}\{a'(0), L(0), \epsilon_{Y(1)}\}]. \end{aligned} \quad (2.13)$$

Since (2.4–2.7) imply (2.12), the g-formula (2.1) under (2.4–2.7) identifies the causal contrast

$$g\{a(0)\} - g\{a'(0)\} = \int E\{Y_{a(0)}(1) - Y_{a'(0)}(1) \mid Z_{a(0)}(1) = Z_{a'(0)}(1) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} dF_{\mathbf{L}(0)}\{\mathbf{I}(0)\}. \quad (2.14)$$

The effect (2.14) is a weighted average of (2.13) and is interpreted as the $\mathbf{L}(0)$ -standardized PS effect. Assumption (2.12) is crucial to identify a causal contrast that does not rely on counterfactuals of Z . Assumption (2.12) implies that the quantity $E\{Y(1) \mid A(0) = a(0), Z(1) = 1, \bar{\mathbf{L}}(0) = \bar{\mathbf{I}}(0)\}$, which can be estimated by observed data, can identify a mean within a principal stratum: $E\{Y_{a(0)}(1) \mid Z_{a(0)}(1) = Z_{a'(0)}(1) = 1, \bar{\mathbf{L}}(0) = \bar{\mathbf{I}}(0)\}$. Without (2.12), $g\{a(0)\} - g\{a'(0)\}$ would be interpreted as the controlled direct effect of $A(0)$ on $Y(1)$, setting $Z(1) = 1$: $E\{Y_{a(0),z(1)=1}(1) - Y_{a'(0),z(1)=1}(1)\}$. As in TGSW (and Supplementary material available at [Biostatistics](#) online), the interpretation is more complex if effects of $A(0)$ are common causes of $Z(1)$ and $Y(1)$.

The interpretation of $g\{\bar{a}(1)\} - g\{\bar{a}'(1)\}$, the effect of $\bar{A}(1)$ on $Y(2)$, is complicated by the fact that $\{Y(1), \mathbf{L}(1)\}$ mediates the effect of $A(0)$ on both $Y(2)$ and $Z(2)$ and is itself truncated by death. To see this, we replace observables in (2.4–2.7) with counterfactuals of $\{\bar{A}(j), Y(j), \mathbf{L}(j)\}$ for $j = 1$ setting $\{\bar{A}(1), Y(1), \mathbf{L}(1)\} = \{\bar{a}(1), y(1), \mathbf{I}(1)\}$. The resulting expressions imply:

$$Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \perp Z_{\bar{a}'(1),y'(1),\mathbf{I}'(1)}(2) \mid \{Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2), \mathbf{L}(0)\} \\ \text{for } \bar{a}(1) \neq \bar{a}'(1) \text{ or } \{y(1), \mathbf{I}(1)\} \neq \{y'(1), \mathbf{I}'(1)\}. \quad (2.15)$$

That is, the “cross-world” independence assumption at follow-up visit 2 is a function of counterfactual variables of mediators of the effect of $A(0)$ on both $Y(2)$ and $Z(2)$. From (2.15), the PS controlled direct effect, given $\mathbf{L}(0) = \mathbf{I}(0)$, setting $\{\bar{A}(1), Y(1), \mathbf{L}(1)\} = \{\bar{a}(1), y(1), \mathbf{I}(1)\}$ is

$$E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) - Y_{\bar{a}'(1),y'(1),\mathbf{I}'(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}'(1),y'(1),\mathbf{I}'(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \\ = E[g_{Y(2)}\{\bar{a}(1), y(1), \bar{\mathbf{I}}(1), \epsilon_{Y(2)}\} - g_{Y(2)}\{\bar{a}'(1), y'(1), \bar{\mathbf{I}}(1), \epsilon_{Y(2)}\}]. \quad (2.16)$$

Equation (2.16) is a *controlled* direct effect because $\{Y(1), \mathbf{L}(1)\}$ is set to the same value, $\{y(1), \mathbf{I}(1)\}$, for both counterfactual outcomes. Therefore, under (2.4–2.7), and hence (2.15),

$$E\{Y(2) \mid \bar{A}(1) = \bar{a}(1), Z(2) = 1, Y(1) = y(1), \bar{\mathbf{L}}(1) = \bar{\mathbf{I}}(1)\} = E[g_{Y(2)}\{\bar{a}(1), y(1), \bar{\mathbf{I}}(1), \epsilon_{Y(2)}\}] \\ = \int g_{Y(2)}\{\bar{a}(1), y(1), \bar{\mathbf{I}}(1), \epsilon\} dF_{\epsilon_{Y(2)}}(\epsilon) = E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \\ = E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}'(1),y'(1),\mathbf{I}'(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \\ = E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}(1),y'(1),\mathbf{I}'(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\}, \quad (2.17)$$

where $F_{\epsilon_{Y(2)}}(\epsilon)$ is the distribution of $\epsilon_{Y(2)}$ evaluated at ϵ . That is, (2.4–2.7) implies (2.15), which implies that the mean of observable $Y(2)$ conditioned on $Z(2) = 1$ and past observables, which can be estimated by the data, equals the mean of a counterfactual variable within a principal stratum. Equations (2.4–2.7) encode additional independence assumptions that ensure that the g-formula identifies a function of PS controlled direct effects at follow-up visit 2. Namely, for $j = 1$, setting $\{A(0), Y(1)\} = \{a(0), y(1)\}$ in (2.4) implies

$$\mathbf{L}_{a(0),y(1)}(1) \perp Z_{a'(0),y'(1)}(1) \mid \{Z_{a(0),y(1)}(1), \mathbf{L}(0)\} \quad \text{for } a(0) \neq a'(0) \text{ or } y(1) \neq y'(1). \quad (2.18)$$

Independence assumption (2.18) implies that the distribution of $\mathbf{L}(1)$ conditioned on $Z(1) = 1$ and past observables can be interpreted as a distribution within principal strata. Define

$$\begin{aligned} dF_{\tilde{\nu}_{\bar{a}(0)}(1)}\{\tilde{\mathbf{v}}(1)\} &= dF_{\mathbf{L}_{a(0),y(1)}(1)|Z_{a(0),y(1)}(1)=1,\mathbf{L}(0)=\mathbf{I}(0)}\{\mathbf{I}(1)\} \\ &\times dF_{Y_{a(0)}(1)|Z_{a(0)}(1)=1,\mathbf{L}(0)=\mathbf{I}(0)}\{y(1)\} \times dF_{\mathbf{L}(0)}\{\mathbf{I}(0)\}, \end{aligned} \quad (2.19)$$

where the use of counterfactuals follows from (2.4–2.7) and implied assumptions (2.12), (2.15), and (2.18). Replacing the expectation in g-formula (2.2) with a counterfactual one from (2.17), and replacing (2.3) with (2.19) then plugging into (2.2) yields the identified causal contrast,

$$\begin{aligned} g\{\bar{a}(1)\} - g\{\bar{a}'(1)\} &= \int E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times dF_{\tilde{\nu}_{a(0)}(1)}\{\tilde{\mathbf{v}}(1)\} \\ &- \int E\{Y_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\}, \end{aligned}$$

a difference in integrals of mean counterfactuals within principal strata. Integrals are taken over differing counterfactual distributions; $F_{\tilde{\nu}_{a(0)}(1)}\{\tilde{\mathbf{v}}(1)\} \neq F_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\}$, in general, for $a(0) \neq a'(0)$. To derive the causal interpretation, we add (and subtract) strategic terms to obtain

$$\begin{aligned} &\int E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\} \\ &- \int E\{Y_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\} \\ &+ \int E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times [dF_{\tilde{\nu}_{a(0)}(1)}\{\tilde{\mathbf{v}}(1)\} - dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\}] \\ &- \int E\{Y_{\bar{a}(1),y'(1),\mathbf{I}'(1)}(2) \mid Z_{\bar{a}(1),y'(1),\mathbf{I}'(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \times [dF_{\tilde{\nu}_{a(0)}(1)}\{\tilde{\mathbf{v}}(1)\} - dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\}], \end{aligned} \quad (2.20)$$

where $\{y'(1), \mathbf{I}'(1)\}$ is a fixed reference value of $\{Y(1), \mathbf{L}(1)\}$. Rearranging the first three terms in (2.20), noting that the last term equals 0, and applying (2.17) leads to

$$\begin{aligned} g\{\bar{a}(1)\} - g\{\bar{a}'(1)\} &= \int E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) - Y_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}'(1),y(1),\mathbf{I}(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \\ &\times dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\} \\ &+ \int E\{Y_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) - Y_{\bar{a}(1),y'(1),\mathbf{I}'(1)}(2) \mid Z_{\bar{a}(1),y(1),\mathbf{I}(1)}(2) = Z_{\bar{a}(1),y'(1),\mathbf{I}'(1)}(2) = 1, \mathbf{L}(0) = \mathbf{I}(0)\} \\ &\times [dF_{\tilde{\nu}_{a(0)}(1)}\{\tilde{\mathbf{v}}(1)\} - dF_{\tilde{\nu}_{a'(0)}(1)}\{\tilde{\mathbf{v}}(1)\}] = \text{Direct Effect}(1) + \text{Indirect Effect}(1). \end{aligned} \quad (2.21)$$

Direct Effect(1) in (2.21) is the PS-controlled direct effect (2.16) averaged over the distribution of baseline confounders $\mathbf{L}(0)$ and PS distributions of $\{Y_{a'(0)}(1), \mathbf{L}_{a'(0),y(1)}(1)\}$. The Indirect Effect(1) is so named because it is a function of the causal effect of $A(0)$ on $\{Y(1), \mathbf{L}(1)\}$ and the causal effect of $\{Y(1), \mathbf{L}(1)\}$ on $Y(2)$. Specifically, Indirect Effect(1) is the difference in average PS effects of $\{Y(1), \mathbf{L}(1)\}$

on $Y(2)$ setting $\bar{A}(1)$ to $\bar{a}(1)$ versus $\bar{a}'(1)$, comparing distributions of $\{Y_{a(0)}(1), \mathbf{L}_{a(0),y(1)}(1)\}$ with those of $\{Y_{a'(0)}(1), \mathbf{L}_{a'(0),y(1)}(1)\}$. See supplementary material available at [Biostatistics](#) online for general visit j .

2.3 Estimation

As described in [Robins \(1998\)](#), the g-formulas (2.1) and (2.2) can be estimated using IPW estimation. Herein, we describe an IPW estimator for $g\{\bar{a}(j)\}$ and use it to derive a doubly robust AIPW estimator. To do so, we assume positivity for mortality, exposure, and dropout, i.e. any possible mortality time, pre-death exposure history, and pre-death dropout time can occur ([Robins and Rotnitzky, 1992](#)). To perform IPW estimation, one specifies the functional form of the g-formula, rather than the conditional distributions; that is, the left-hand sides of (2.1) and (2.2), rather than the right-hand sides. In AIPW estimation, the functional forms of the right-hand sides of (2.1) and (2.2) are additionally specified. To be concrete, we consider a specific $g\{\bar{a}(j)\}$:

$$g\{\bar{a}(j), \boldsymbol{\beta}\} = \beta_0 + \beta_1 j + \beta_2 \text{cum}\{\bar{a}(j)\}, \quad (2.22)$$

where $\text{cum}\{\bar{a}(j)\} = \sum_{k=0}^j a(k)$, the cumulative exposure, and $g\{\bar{a}(j), \boldsymbol{\beta}\} - g\{\bar{a}'(j), \boldsymbol{\beta}\} = \beta_2 [\text{cum}\{\bar{a}(j)\} - \text{cum}\{\bar{a}'(j)\}]$.

2.3.1 IPW estimator. Inverse probability of exposure weighted estimation can consistently estimate (2.22) in longitudinal studies because it can handle collider-stratification like that shown in Figure 2 ([Bryan and others, 2004](#); [Robins and others, 2000](#)). To estimate (2.22) with death and dropout, weights are also used to ensure that Z and R are statistically independent of their common causes of Y .

To specify the weights, let $\mathcal{F}(k) = \{\bar{A}(k-1), \bar{R}(k)=1, \bar{Z}(k)=1, \bar{Y}(k), \bar{\mathbf{L}}(k)\}$, $k=0, \dots, J$, be pre- $A(k)$ history and let $f_{A(k)} = f\{A(k) | \mathcal{F}(k)\}$, where $f\{\cdot | \cdot\}$ is a conditional probability mass or density function. Similarly, let $p_{Z(k+1)} = \Pr\{Z(k+1)=1 | A(k), \mathcal{F}(k)\}$, and $p_{R(k+1)} = \Pr\{R(k+1)=1 | Z(k+1)=1, A(k), \mathcal{F}(k)\}$. Each participant has time j weight,

$$\begin{aligned} SW(j) &= \prod_{k=0}^j f\{A(k) | \bar{A}(k-1), \bar{R}(k)\bar{Z}(k)=1\} f_{A(k)}^{-1} \\ &\quad \times \Pr\{Z(k+1)=1 | \bar{A}(k), \bar{R}(k)\bar{Z}(k)=1\} p_{Z(k+1)}^{-1} \\ &\quad \times \Pr\{R(k+1)=1 | \bar{A}(k), \bar{R}(k)\bar{Z}(k+1)=1\} p_{R(k+1)}^{-1}, \end{aligned}$$

$j=0, \dots, J$, where $A(-1)=0$. The SW are called stabilized weights due to the numerator terms that improve efficiency ([Robins, 1998](#)). The IPW estimator for $\boldsymbol{\beta}$ solves the estimating equation,

$$0 = \mathbf{U}_{\text{IPW}} = \sum_{i=1}^N \mathbf{U}_{i,\text{IPW}} = \sum_{i=1}^N \frac{\delta^t \boldsymbol{\mu}_i}{\delta \boldsymbol{\beta}} \mathbf{R}_i \mathbf{Z}_i \mathbf{S} \mathbf{W}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i), \quad (2.23)$$

where $\mathbf{Y}_i = \{Y_i(1), \dots, Y_i(J+1)\}^t$, $\boldsymbol{\mu}_i = \{g(A_i(0), \boldsymbol{\beta}), \dots, g(\bar{A}_i(J), \boldsymbol{\beta})\}^t$, and $\mathbf{R}_i \mathbf{Z}_i \mathbf{S} \mathbf{W}_i$ is a $J \times J$ diagonal matrix of $R_i(j+1)Z_i(j+1)SW_i(j)$. An independence correlation avoids bias from any dependence of $R(j+1)$ or $Z(j+1)$ on $\bar{Y}(j)$ ([Tchetgen Tchetgen and others, 2012b](#)). SW is unknown, thus an estimator, $\hat{SW}$, is plugged into (2.23). $\hat{\boldsymbol{\beta}}_{\text{IPW}}$ solves (2.23) and is consistent due to m -estimation and positivity for exposure, mortality, and dropout with correct SW .

Table 1. Results from 5000 simulations of sample size 1000

Method	SW correct?	$\bar{m}$ correct?	$\hat{\beta}_2$	Bias	SE	ESE
IPW	Yes	—	1.997	−0.003	0.306	0.338
	No	—	1.279	−0.721	0.245	0.224
AIPW	Yes	Yes	2.006	0.006	0.125	0.149
		No	1.998	−0.002	0.285	0.315
	No	Yes	1.994	−0.006	0.115	0.113
		No	1.293	−0.707	0.226	0.219

The true average causal effect is 2. SE denotes the average standard error, ESE denotes empirical standard error; SW denotes stabilized weight; and $\bar{m} = \{m_{Y(1)}, m_{Y(2)}\}$ denotes outcome generating models.

2.3.2 Augmented inverse-probability weighted (AIPW) estimator. IPW estimation is inefficient if SW is highly variable (Robins, 1999) and inconsistent if SW is incorrect. AIPW estimation can address both problems. The AIPW estimator is calculated by projecting U_{IPW} onto the nuisance tangent space, which here is the Hilbert space spanned by mean-0 functions of $A(j)$, $Z(j+1)$, and $R(j+1)$, conditioned on the past equipped with the covariance operator. The estimating equation, U_{AIPW} , is U_{IPW} minus its projection and is located in the orthogonal complement of the nuisance tangent space. A technical treatment of Hilbert spaces and projections for missing data or causal inference is in van der Laan and Robins (2003) and Tsiatis (2006). In the Supplementary Material available at *Biostatistics* online, we derive the AIPW estimator and its variance, and we provide a proof of double robustness.

3. SIMULATION STUDY

We simulated 5000 datasets with $N = 1000$ and two follow-up visits. $L(0) = \{X, Y(0)\}$ were $N(0, 1)$, and binary $A(0)$ was assigned using $\Pr\{A(0) = 1 \mid L(0)\}$. $Z(1)$ was $\Pr\{Z(1) = 1 \mid L(0), a(0)\}$, and $Y(1)$ was $N[m_{Y(1)}\{L(0), a(0), \bar{Z}(1) = 1\}, 1]$. Binary $A(1)$ was assigned using $\Pr\{A(1) = 1 \mid \bar{L}(1), A(0), \bar{Z}(1) = 1\}$, where $\bar{L}(1) = \{X, \bar{Y}(1)\}$. $R(1)$ was $\Pr\{R(1) = 1 \mid L(0), A(0), \bar{Z}(1) = 1\}$, and $Z(2)$ was $\Pr\{Z(2) = 1 \mid \bar{L}(1), \bar{A}(1), \bar{Z}(1) = 1\}$. $Y(2)$ was $N[m_{Y(2)}\{\bar{L}(1), \bar{a}(1), \bar{Z}(2) = 1\}, 1]$, and $R(2)$ was $\Pr\{R(2) = 1 \mid \bar{L}(1), \bar{A}(1), \bar{Z}(2) = 1, \bar{R}(1) = 1\}$. The true data-generating models in Table 1 of Supplementary material available at *Biostatistics* online result in $g\{a(0) = 0\} = 0$, $g\{a(0) = 1\} = 2$, $g\{\bar{a}(1) = (0, 0)\} = -2$, $g\{\bar{a}(1) = (1, 0)\} = g\{\bar{a}(1) = (0, 1)\} = 0$, and $g\{\bar{a}(1) = (1, 1)\} = 2$. Thus, $g\{\bar{a}(j), \beta\}$ in (2.22) has $\beta = \{0, -2, 2\}$; that is, the average causal effect, β_2 , is 2.

Table 1 shows estimates under both true and false models in Table 1 of Supplementary material available at *Biostatistics* online. IPW estimates were biased when false models were used to estimate SW , whereas AIPW estimates were biased only when both SW and outcome means $\bar{m} = \{m_{Y(1)}, m_{Y(2)}\}$ were incorrect. When SW was correctly modeled, AIPW with correct $\bar{m}$ was more efficient than IPW.

4. DATA ANALYSIS: THE INCHIANTI STUDY

We aim to assess the effect of 25(OH)D on gait speed, a measure of physical function, in InChianti. Low 25(OH)D relates to frailty (Shardell and others, 2012), of which slow gait speed is a part. Data here include 905 older adults. Follow-up visits were planned at 3 and 6 years post baseline; by year 6, 163 adults died, 288 dropped out, and 454 had complete data.

Sufficient serum 25(OH)D, according to the IoM (≥ 20 ng/mL) or Endocrine Society (≥ 30 ng/mL), is not an intervention per se, but it indicates self-selected vitamin D exposure. Thus, participants “intervene”

Table 2. *Estimates and standard errors for cumulative effect of sufficient vitamin D on gait speed (m/s)*

Sufficient Vitamin D	Method	$\hat{\beta}_2$	SE
≥ 20 ng/mL	IPW	0.056	0.022
	AIPW	0.021	0.044
	GEE	0.022	0.012
	W-IEE	0.027	0.012
	MW-GEE	0.035	0.014
≥ 30 ng/mL	IPW	0.055	0.021
	AIPW	0.0003	0.046
	GEE	0.008	0.012
	W-IEE	0.012	0.012
	MW-GEE	0.023	0.014

IPW denotes inverse-probability weighting; AIPW denotes augmented inverse-probability weighting; GEE denotes generalized estimating equation; W-IEE denotes weighted independence estimating equation; MW-GEE denotes multiply-weighted GEE.

on themselves. In this study, Y is 4 m gait speed in meters per second and binary A is the indicator of sufficient 25(OH)D. Serum 25(OH)D measures recent vitamin D exposure and can be affected by factors assessed at the same visit. Therefore, L , time-varying covariates, includes lifestyle factors and comorbidities; X is sex and baseline age.

We estimated (2.22) using IPW and AIPW. We also estimated associations using generalized estimating equations (GEE) (Liang and Zeger, 1986), weighted independence estimating equations (W-IEE) partly conditioned on being alive with weights for dropout only (Kurland and Heagerty, 2005; Shardell and Miller, 2008; Shardell and others, 2010), and multiply-weighted GEE (MW-GEE) with weights for deaths and dropout (Rajan and Leurgans, 2010). We conditioned on covariates in GEE, W-IEE, and MW-GEE; logistic regression estimated all weights.

Table 2 shows $\hat{\beta}_2$ from (2.22) and covariate-adjusted analogs. When assessing 25(OH)D ≥ 20 ng/mL, we found that the AIPW estimate was the smallest (0.021 m/s), and the IPW estimate was the largest (0.056 m/s); both of which are interpreted at the first follow-up visit as PS direct effects of 25(OH)D on gait speed averaged over age, sex, and baseline lifestyle factors and comorbidities; and at the second follow-up as PS-controlled direct effects averaged over age, sex, and lifestyle factors and comorbidities measured at baseline and the first follow-up visit plus the difference of PS direct effects of lifestyle factors and comorbidities on gait speed averaged over the distributions of lifestyle factors, comorbidities, and previous gait speed at high versus low 25(OH)D. Other estimates are covariate-adjusted non-causal associations of 25(OH)D with gait speed conditioned on being alive and observed (GEE), conditioned on being alive (W-IEE), and for an immortal cohort (MW-GEE). When 25(OH)D ≥ 30 ng/mL was assessed, we found again that the AIPW estimate was the smallest (0.0003 m/s), and the IPW estimate was the largest (0.055 m/s). Overall conclusions from this analysis about the causal effect of 25(OH)D on gait speed are consistent across IoM and Endocrine Society guidelines. Not only is the IPW point estimate larger than the AIPW point estimate, the 95% confidence intervals for β_2 ($\hat{\beta}_2 \pm 1.96$ SE) estimated using IPW exclude 0, whereas the 95% confidence intervals estimated using AIPW include 0. Thus, we found some evidence that sufficient 25(OH)D causes faster gait speed, regardless of 25(OH)D guideline, but the results are not robust to model misspecification. A caveat is that a causal interpretation depends on assumptions from Sections 2.1 and 2.2.

Causal inference with dichotomized continuous exposures is controversial, because counterfactuals of such exposures are vague to the extent that a hypothetical intervention is unspecified (Hernán, 2005;

Robins and Greenland, 2000). Van der Laan *and others* (2005) and Hernán and VanderWeele (2011) showed that the average effect of dichotomized exposures, such as 25(OH)D, is the effect averaged over the population distribution of continuous 25(OH)D, which may not be transportable across populations.

5. DISCUSSION

We described IPW and doubly robust AIPW estimators of causal effects in longitudinal studies with truncation by death. Extending the exposition by TGSW, we showed that the identified causal contrast is a function of weighted averages of PS effects. The IPW approach can be performed using off-the-shelf software, which removes a barrier to applying causal inference to aging research; however, the causal contrast has the same limitations as SACE due to its dependence on indirect PS effects. In the context herein, the causal contrast at the second follow-up visit is challenging to interpret, a PS-controlled direct effect plus the difference of integrated indirect PS effects, due to factors that mediate the effect of exposure on both outcome and mortality. TGSW also considered these mediators for a single follow-up visit. In Supplementary material available at *Biostatistics* online, we include these mediators for all follow-up visits of a longitudinal study.

A causal contrast that handles truncation by death ideally satisfies the criteria being easily interpreted, easily estimated, philosophically sound, and scientifically interesting. A practical challenge is that no proposed contrast satisfies all four criteria. Counterfactuals of exposure A and vital status Z , $Y_{\bar{a}\bar{z}}$, can lead to contrasts that are easily interpreted as controlled direct effect of A on Y setting $Z = 1$, otherwise known as the “death blocking” causal effect (Joffe, 2011), and can be estimated using IPW; however, such counterfactuals are ill-defined unless intervention A can prevent/delay death. SACE avoids specifying counterfactuals of vital status via either monotonicity or “cross-world” independence assumptions (Hayden *and others*, 2005; Eggleston *and others*, 2007) and is interpretable, but it is more challenging to estimate in longitudinal studies of time-varying exposures without additional assumptions (Tchetgen Tchetgen, 2013). The causal contrast proposed herein avoids counterfactuals of vital status via “cross-world” independence assumptions and is easily estimable, but at the cost of interpretability. Lastly, scientific interest of both the “death blocking” contrast and SACE (and by extension, the contrast proposed herein) has been questioned (Pearl, 2011; Joffe, 2011); thus an objective for the biostatistics community on truncation by death is to define a causal contrast that satisfies this elusive criterion. Until then, causal contrasts that can be estimated using off-the-shelf statistical software can help raise awareness of truncation by death among practitioners and are an improvement over the usual practices of ignoring selective survival or treating mortality as synonymous with missingness.

Regarding doubly robustness estimation, recently proposed methods have shown improved efficiency over AIPW when outcome regression models are incorrect (Rotnitzky *and others*, 2012; Tan, 2010; Tsiatis *and others*, 2011; van der Laan, 2010, and references therein). Adapting these methods to handle truncation by death is a priority for future research.

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

FUNDING

This work was supported by the National Institutes of Health (K25AG034216, R01AG041202).

Conflict of Interest: None declared.

REFERENCES

- BANG, H. AND ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–972.
- BRYAN, J., YU, Z. AND VAN DER LAAN, M. J. (2004). Analysis of longitudinal marginal structural models. *Biostatistics* **5**, 361–380.
- EGLESTON, B. L., SCHARFSTEIN, D. O., FREEMAN, E. E. AND WEST, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8**, 526–545.
- FERRUCCI, L., BANDINELLI, S., BENVENUTI, E., DI IORIO, A., MACCHI, C., HARRIS, T. B. AND GURALNIK, J. M. (2000). Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *Journal of the American Geriatrics Society* **48**, 1618–1625.
- FRANGAKIS, C. E. AND RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- GLYMOUR, M. M., MUJAHID, M., WU, Q., WHITE, K. AND TCHETGEN TCHETGEN E. J. (2010). Neighborhood disadvantage and self-assessed health, disability, and depressive symptoms: Longitudinal results from the health and retirement study. *Annals of Epidemiology* **20**, 856–861.
- HAYDEN, D., PAULER, D. K. AND SCHOENFELD, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61**, 305–310.
- HERNÁN, M. A. (2005). Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite? *American Journal of Epidemiology* **162**, 618–620.
- HERNÁN, M. A., HERNÁNDEZ-DÍAZ, E. F. AND ROBINS, J. M. (2004). A structural approach to selection bias. *Epidemiology* **15**, 615–625.
- HERNÁN, M. A. AND VANDERWEELE, T. J. (2011). Compound treatments and transportability of causal inference. *Epidemiology* **22**, 368–377.
- HIV-CAUSAL COLLABORATION (2012). Impact of antiretroviral therapy on tuberculosis incidence among HIV-positive patients in high-income countries. *Clinical Infectious Diseases* **54**, 1364–1372.
- HOLICK, M. F., BINKLEY, N. C., BISCHOFF-FERRARI, H. A., GORDON, C. M., HANLEY, D. A., HEANEY, R. P., MURAD, M. H. AND WEAVER, C. M. (2011). Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* **96**, 1911–1930.
- HOLICK, M. F., BINKLEY, N. C., BISCHOFF-FERRARI, H. A., GORDON, C. M., HANLEY, D. A., HEANEY, R. P., MURAD, M. H. AND WEAVER, C. M. (2012). Guidelines for preventing and treating vitamin D deficiency and insufficiency revisited. *Journal of Clinical Endocrinology and Metabolism* **97**, 1153–1158.
- INSTITUTE OF MEDICINE (2011). *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: The National Academies Press.
- JEMAI, Y., ROTNITZKY, A., SHEPHERD, B. E. AND GILBERT, P. B. (2007). Semiparametric estimation of treatment effects given baseline covariates on an outcome measured after a post-randomization event. *Journal of the Royal Statistical Society: Series B* **69**, 879–901.
- JOFFE, M. (2011). Principal stratification and attribution prohibition: good ideas taken too far. *The International Journal of Biostatistics* **7**, article 35.
- KURLAND, B. F. AND HEAGERTY, P. J. (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6**, 241–258.
- LEE, L., DANIELS, M. J., AND SARGENT, D. J. (2010). Causal effects of treatment for informative missing data due to progression/death. *Journal of the American Statistical Association* **105**, 912–929.
- LIANG, K. Y. AND ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd edition. New York: Cambridge.
- PEARL, J. (2011). Principal stratification – a goal or a tool? *The International Journal of Biostatistics* **7**, article 20.
- RAJAN, K. B. AND LEURGANS, S. E. (2010). Joint modeling of missing data due to non-participation and death in longitudinal aging studies. *Statistics in Medicine* **29**, 2260–2268.
- ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. *Mathematical Modelling*, **7**, 1393–1512.
- ROBINS, J. M. (1998). Marginal structural models. In *1997 Proceedings of the American Statistical Association Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association, pp. 1–10.
- ROBINS, J. M. (1999). Marginal structural models versus structural nested models as tools for causal inference. In Halloran, M. E. and Berry, D (editors). *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York: Springer, pp. 95–133.
- ROBINS, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference. In *1999 Proceedings of the American Statistical Association Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association, pp. 6–10.
- ROBINS, J. M. AND GREENLAND, S. (2000). Comment on “Causal inference without counterfactuals.” *Journal of the American Statistical Association* **95**, 431–435.
- ROBINS, J. M., HERNÁN, M. A. AND BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.
- ROBINS, J. M. AND ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N., Dietz, K. and Farewell, V. (editors). *AIDS Epidemiology—Methodological Issues* Boston: Birkhäuser, pp. 297–331.
- ROSEN, C. J. and others (2012). IOM committee members respond to Endocrine Society vitamin D guideline. *Journal of Clinical Endocrinology and Metabolism* **97**, 1146–1152.
- ROTNITZKY, A., LEI, Q., SUED, M. AND ROBINS, J. M. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika* **99**, 439–456.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- RUBIN, D. B. (2006). Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death. *Statistical Science* **21**, 299–309.
- SHARDELL, M., D’ADAMO, C., ALLEY, D. E., MILLER, R. R., HICKS, G. E., MILANESCHI, Y., SEMBA, R. D., CHERUBINI, A., BANDINELLI, S. AND FERRUCCI, L. (2012). Serum 25-hydroxyvitamin D, transitions between frailty states, and mortality in older adults: the Invecchiare in Chianti Study. *Journal of the American Geriatrics Society* **60**, 256–264.
- SHARDELL, M., HICKS, G. E., MILLER, R. R., AND MAGAZINER, J. (2010). Semiparametric regression models for repeated measures of mortal cohorts with non-monotone missing outcomes and time-dependent covariates. *Statistics in Medicine* **29**, 2282–2296.
- SHARDELL, M. AND MILLER, R. R. (2008). Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine* **27**, 1008–1025.
- SHEPHERD, B. E., GILBERT, P. B., JEMAI, Y. AND 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**, 332–342.
- TAN, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97**, 661–682.
- TCHETGEN TCHETGEN, E. J. (2013). Identification and estimation of survivor average causal effects. *Harvard University Biostatistics Working Paper Series*, working paper 167.

- TCHETGEN TCHETGEN, E. J., GLYMOUR, M. M., SHPITSER, I. AND WEUVE J. (2012a). To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology* **23**, 132–137.
- TCHETGEN TCHETGEN, E. J., GLYMOUR, M. M., WEUVE J. AND ROBINS J. (2012b). Specifying the correlation structure in inverse-probability- weighting estimation for repeated measures. *Epidemiology* **23**, 644–646.
- TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.
- TSIATIS, A. A., DAVIDIAN, M. AND CAO, W. (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics* **67**, 536–545.
- VAN DER LAAN, M. J (2010). Targeted maximum likelihood based causal inference: part I. *International Journal of Biostatistics* **6**, article 2.
- VAN DER LAAN, M. J., HAIGHT, T. J. AND TAGER, I. B. (2005). Response to “Hypothetical interventions to define causal effects.” *American Journal of Epidemiology* **162**, 621–622.
- VAN DER LAAN, M. J. AND ROBINS, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.

[Received January 09, 2013; revised May 23, 2014; accepted for publication May 31, 2014]