To Weight or Not to Weight?

On the Relation Between Inverse-probability Weighting and Principal Stratification for Truncation by Death

Eric J. Tchetgen Tchetgen,[a,b] *M. Maria Glymour,*[c] *Ilya Shpitser,*[a] *and Jennifer Weuve*[d]

We thank the editor for the opportunity to respond to the commentary by Chaix et al.[1] In this rejoinder, we address the concerns voiced by Chaix et al[1] that inverse-probability weighting creates an immortal population, and that instead, an application of the principal stratification framework to our data would provide a more useful causal effect of smoking on cognitive decline. We distinguish between issues pertaining to statistical inference and those related to causal inference, and we show that, under a structural equation framework, the causal effect identified by inverse-probability weighting survivors, in fact, naturally incorporates principal-strata causal effects. Therefore, we formally establish relations between these 2 seemingly unrelated analytic frameworks.

We consider a simplified study design with only 3 occasions, as in the causal directed acyclic graph in Figure 1: a baseline $j = 0$ at which binary smoking status S is observed, and 2 follow-up contacts, with cognitive function $C(j)$ assessed at each $j = 1, 2$. All respondents participate at the first follow-up, but some die before the second follow-up, with $D = 0$ indicating survival. Survival is affected by S and $C(1)$, and death is the only source of attrition. The outcome of interest is change in cognitive function $\Delta C = C(2) - C(1)$. Throughout, we assume no measurement error, and we distinguish between an exogenous time-varying common cause of D and ΔC, which we indicate by L, and an endogenous common cause of D and ΔC, for example, $C(1)$. An exogenous common cause of D and ΔC is known not to be affected by S, whereas an endogenous common cause of the 2 variables may be an effect of S. To simplify further, we take $C(1)$ to be the only endogenous common cause of D and ΔC. The variable U encodes possible unmeasured common causes of cognitive function measurement $C(1)$ and change in cognitive function ΔC, the presence of which cannot be ruled out with certainty. For instance, there is evidence for genetic determinants of Alzheimer disease that suggests a genetic basis for an individual's cognitive function over time[2]; however, such genetic information was not available in our study. Although such a genetic component of U would not be affected by smoking behavior, U might also include an unknown epigenetic effect of smoking behavior on future cognitive function, in which case it would be an effect of smoking behavior. For simplicity and with no loss in generality, we further assume that all analyses are stratified by a set of time-constant confounders of S.

Issues of Statistical Inference

The first issue relates to statistical inference using inverse-probability weighting. Chaix et al[1] claim that we are creating an immortal population by applying inverse-probability weights to the survivors, or in their words, that, inverse-probability weighting

From the Departments of [a]Epidemiology, [b]Biostatistics, and [c]Society, Human Development and Health, Harvard School of Public Health, Boston, MA; and [d]Rush Institute for Healthy Aging and Department of Internal Medicine, Rush University Medical Center, Chicago, IL.

Editors' note: *Related articles appear on pages 119 and 129.*

Correspondence: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail: etchetge@hsph.harvard.edu.

ISSN: 1044-3983/12/2301-0132
DOI: 10.1097/EDE.0b013e31823b5081

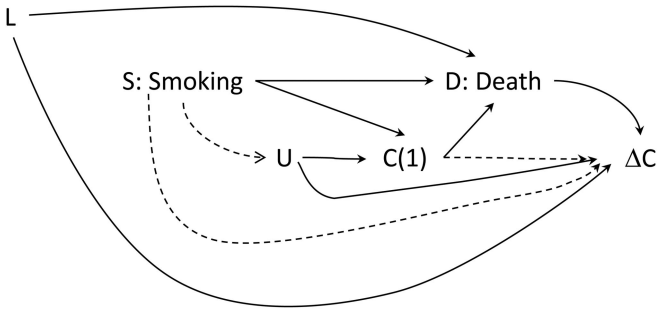

FIGURE 1. Causal directed acyclic graph representing the hypothesized causal structure linking smoking and cognitive change (Δ*C*). In this model, *L* represents an exogenous common cause of death and Δ*C*. An unmeasured variable *U* may be affected by smoking (this path is represented by a dotted arrow) and affects both cognitive function at time 1 (*C*(1) in the DAG) and Δ*C*. Similarly, *C*(1) is affected by smoking and may affect Δ*C* (represented by a dotted arrow). We define the absence of any of the 3 dotted arrows as the null hypothesis of no effect of smoking on Δ*C*. We show an arrow from death to Δ*C* because Δ*C* is undefined for dead people. Therefore, Δ*C* is affected by death, but this pathway is not of substantive interest in the research.

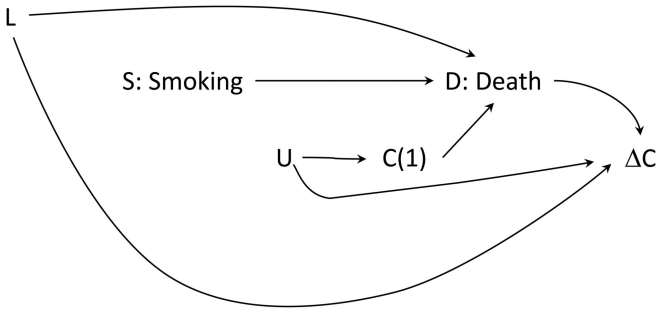

FIGURE 2. Causal directed acyclic graph representing no effect of smoking on Δ*C* and no endogenous common causes of death and Δ*C*. Under this DAG, *L* and *C*(1) d-separate smoking and Δ*C*. In other words, under this DAG, a conventional analysis predicting Δ*C* with smoking, *C*(1) and *L* in a regression model could be unbiased for the null even when estimated among survivors. This DAG has no arrow from smoking to *C*(1) because if smoking affects *C*(1) (regardless of whether *C*(1) affects Δ*C*), the conventional analysis conducted among survivors would be biased.

is "weighting up the dead." Thus, according to Chaix et al, "[i]n replacing dead participants by cloning the living, inverse-probability weighting generates a sample in which participants are not allowed to die." Although the "cloning" metaphor often used to describe inverse-probability weighting can be a helpful pedagogic tool, it should not be taken literally. The metaphor fails to convey the appropriate statistical interpretation of inverse-probability-weighted estimation, particularly in the context of attrition due to death. A more appropriate statistical interpretation is that inverse-probability weighting ensures the statistical independence between *D* and (measured) time-varying common causes (*C*(1), *L*) of *D* and Δ*C*. Most importantly, by virtue of weighting survivors by the inverse of their probability of surviving, inverse-probability weighting ensures such independence without altering the conditional density $f_1(\Delta C \mid C(1), L, U, S, D = 0)$ of Δ*C* given the past among survivors, and the joint conditional density $f_2(C(1), L, U \mid S)$ of covariates given smoking status. These 2 properties are key to our goal of making inferences about the effects of smoking on cognitive decline because, as we show later, all meaningful information about such effects is contained in f_1 and f_2.

Although Chaix et al argue it is "courageous" to develop an inverse-probability-weighting model to predict survival when only 20% of the sample remains at the end of follow-up, we think it is even more courageous to ignore selective survival in such a circumstance. A conventional analysis of survivors in such a situation—with regression adjustment for common causes of *D* and Δ*C*—could provide a valid test of the sharp null hypothesis of no effect of *S* on Δ*C*, as described in Figure 1 (omitting the dotted arrows), only if there were no endogenous common causes of *D* and Δ*C*, as in Figure 2. This scenario is inconsistent with prior empirical studies. If exposure and time-varying causes of Δ*C* are associated, as would likely be the case in an unweighted population of survivors, the time-varying covariates will bias associations between the exposure and Δ*C*.[3] Thus, ignoring the presence of strong selective survival is almost guaranteed to bias the observed association between *S* and Δ*C*, even under the sharp null hypothesis.

In contrast, under a standard positivity assumption for survival,[4] inverse-probability weighting is well known to estimate the g-formula of Robins[5]:

$$g(s) = \sum_{c,u,l} E(\Delta C \mid S = s, D = 0, C(1) = c, U = u, L = l) \times f_2(C(1) = c, U = u, L = l \mid S = s) \quad (1)$$

$$= \sum_{c,l} E(\Delta C \mid S = s, D = 0, C(1) = c, L = l) \times f_2(C(1) = c, L = l \mid S = s) \quad (2)$$

$s = 0, 1.$

where the second representation follows from the conditional independence of *U* and *D* given (*S, L, C*(1)). Thus, *g*(*s*) is a statistical object that takes as input the distribution of the observed data, and generates a number for each value of smoking status. In particular, irrespective of its causal interpretation, *g*(*s*) is always well defined for both values of *s*, and its definition requires no reference to counterfactuals. Later in

the text, we address the causal interpretation of $g(s)$. For now, consider again the sharp null hypothesis. It is well known that the null implies that $g(s)$ is a constant, and therefore that $g(s)$ is guaranteed not to find an effect of S when there is none. In other words, $g(s)$ can be used to construct a valid test of the null of no effect of S on ΔC in the presence of dependent truncation due to death. This can be observed in Figure 1 under the sharp null (ie, omitting the dotted arrows from the diagram) in which S is independent of (d-separated from) U, and $C(1)$ is conditionally independent of ΔC given L, U, and $D = 0$.

The fact that a weighted analysis is guaranteed to give the correct answer under the sharp null for the causal structure in Figure 1 in itself constitutes compelling justification for epidemiologists to routinely present analyses inverse-probability weighted for attrition due to death, even if as supplemental analyses.

Issues of Causal Inference

We now turn to issues of causal inference. Chaix et al[1] question the interpretation of the inverse-probability-weighting estimator and thus the meaning of $g(s)$. We have established earlier that $g(s)$ recovers the sharp null when it holds, in which case the causal interpretation of $g(s)$ should be uncontroversial. Next, we consider the alternative hypothesis. In this case, by the argument given earlier, if a test statistic based on $g(s)$ rejects, we can safely conclude that such an effect is present at the usual type-1-error nominal level, provided no modeling error exists. The ability of such a statistical test to reject is intimately related to the causal interpretation one can assign to the contrast $g(1) - g(0)$, corresponding to the effect of smoking measured by the g-formula. In addition, a causal interpretation is indispensable for understanding the direction and magnitude of the effects of smoking, and for providing meaningful policy recommendations.

Next, we show that $g(s)$ can be given a meaningful causal interpretation in which principal-strata causal effects are naturally incorporated. Specifically, we establish that under a structural equation model, inverse-probability weighting for survival as implemented by Weuve et al[3] delivers inferences about a causal effect of S on ΔC that is a combination of an average principal-strata direct effect of S on ΔC and indirect effects of S on ΔC. For instance, in the simple case of a linear structural model for ΔC, the direct effect is defined for the principal stratum of individuals who are alive and for whom smoking has no causal effect on ($C(1)$, U, D). In this case, the indirect effects include some components that are defined for everyone in the population and others defined only for individuals in a principal stratum. The indirect effect combines the total direct causal effects of S on (measured and unmeasured) endogenous causes of ΔC with principal-strata direct effects of the latter variables on ΔC. Therefore, we formally establish that the g-formula (hence inverse-probability-weighted estimation) and principal stratification are in fact not disparate concepts, but are instead connected.

The exposition is framed around a structural equation theory of causal inference, described by Pearl.[6] Structural equations provide a nonparametric algebraic interpretation of the diagram of Figure 1 corresponding to 6 functions, one for each variable on the causal graph:

$$S = g_s(\varepsilon_S) \tag{3}$$

$$U = g_U(S, \varepsilon_U) \tag{4}$$

$$C(1) = g_{C(1)}(S, U, \varepsilon_{C(1)}) \tag{5}$$

$$L = g_L(\varepsilon_L) \tag{6}$$

$$D = g_D(S, C(1), L, \varepsilon_D) \tag{7}$$

$$\Delta C = \begin{cases} g_{\Delta C}(S, C(1), U, L, \varepsilon_{\Delta C}) & \text{if } D = 0 \\ \text{undefined} & \text{if } D = 1 \end{cases} \tag{8}$$

Each of the nonparametric functions $\{g_S, \ldots, g_{\Delta C}\}$ represents a causal mechanism that determines the value of the left variable, known as the output, from variables on the right, known as the inputs.[6] The errors (ε_S, ε_U, $\varepsilon_{C(1)}$, ε_L, ε_D, $\varepsilon_{\Delta C}$) stand for all factors not included on the graph that could possibly affect their corresponding outputs when all other inputs are held constant. To be consistent with the causal graph presented in Figure 1, it is required that these errors be mutually independent, but we allow their distribution to remain arbitrary. Lack of a causal effect of a given variable on an output is encoded by an absence of the variable from the right-hand side. For example, the absence of U from the arguments of g_D encodes the assumption that variations in U will leave D unchanged, as long as variables S, $C(1)$, and ε_D remain constant, which is consistent with the assumption that there are no unmeasured common causes of death and cognitive decline.

The last equation makes explicit the fact that ΔC is observed only among survivors with ($D = 0$), with structural equation $g_{\Delta C}(S, C(1), U, L, \varepsilon_{\Delta C})$. As stated by Pearl,[6] the invariance of structural equations permits their use as a basis for modeling causal effects and counterfactuals. In fact, to emulate the intervention in which one sets $\{S = s\}$ for all individuals simply amounts to replacing the equation for S with $S = s$, producing the following set of modified equations:

$$S = s$$

$$U_s = g_U(s, \varepsilon_U)$$

$$C_s(1) = g_{C(1)}(s, U_s, \varepsilon_{C(1)})$$

$$L_s = g_L(\varepsilon_L)$$

$$D_s = g_D(s, C(1), L_s, \varepsilon_D)$$

$$\Delta C_s = \begin{cases} g_{\Delta C}(s, C_s(1), U_s, L_s, \varepsilon_{\Delta C}) & \text{if } D_s = 0 \\ \text{undefined} & \text{if } D_s = 1 \end{cases}$$

with $(U_s, C_s(1), L_s, D_s, \Delta C_s)$ denoting the counterfactual outcomes had smoking status been set to s (possibly contrary to fact). We emphasize that although the model specifies a structural equation for death, survival is not manipulable, and, together with ΔC, should be understood as part of the outcome produced by the system of equations. Structural equations are particularly helpful to clarify the difficulty with interpreting the effect of smoking when truncation by death is present. Specifically, we note that the individual effect of smoking is recovered by taking the contrast $\Delta C_1 - \Delta C_0$, which clearly is defined only for individuals in the principal stratum $\{D_0 = D_1 = 0\}$ and is equal to the following:

$$g_{\Delta C}(S = 1, C_1(1), U_1, L, \varepsilon_{\Delta C}) - g_{\Delta C}(S = 0, C_0(1), U_0, L, \varepsilon_{\Delta C})$$

Consider the submodel M_1 in which equations (4), (5), and (8) are linear.

$$S = g_S(\varepsilon_S) \tag{9}$$

$$U = \theta_S S + \varepsilon_U \tag{10}$$

$$C(1) = \alpha_S S + \alpha_U U + \varepsilon_{C(1)} \tag{11}$$

$$L = g_L(\varepsilon_L) \tag{12}$$

$$D = g_D(S, C(1), L, \varepsilon_D) \tag{13}$$

$$\Delta C = \begin{cases} \beta_S S + \beta_{C(1)} C(1) + \beta_U U + \beta_L L + \varepsilon_{\Delta C} & \text{if } D = 0 \\ \text{undefined} & \text{if } D = 1 \end{cases} \tag{14}$$

We note that β_S formally encodes the principal-strata causal effect of S among all individuals who survived irrespective of smoking status under an intervention in which we hold $(C(1), U)$ fixed, ie, among individuals with $\{D_{0,c,u} = D_{1,c,u} = 0\}$, where $D_{s,c,u}$ is the counterfactual outcome obtained by replacing equations (3)–(5) with the following:

$$S = s$$

$$U = u$$

$$C(1) = c$$

In other words, β_S is a principal-strata controlled direct effect. Similarly, $\beta_{C(1)}$ is the principal-strata causal effect of $C(1)$ on ΔC if one could intervene to set $C(1)$ to $c + 1$ versus c, while holding (S, U) fixed at (s, u), among individuals with $\{D_{s,c,u} = D_{s,c+1,u} = 0\}$; similar interpretations hold for β_U and β_L.

Under the structural linear model, it is easy to verify that the causal effect of S on ΔC for individuals in the principal stratum $\{D_0 = D_1 = 0\}$ is constant and equal to the following:

$$\underbrace{\beta_S}_{\substack{\text{direct effect} \\ S \to \Delta C}} + \underbrace{\beta_{C(1)}\alpha_S}_{\substack{\text{indirect effect} \\ S \to C(1) \to \Delta C}} + \underbrace{\beta_{C(1)}\alpha_U\theta_S}_{\substack{\text{indirect effect} \\ S \to U \to C(1) \to \Delta C}} + \underbrace{\beta_U\theta_S}_{\substack{\text{indirect effect} \\ S \to U \to \Delta C}}$$

As stated above, β_S encodes a principal-strata controlled direct effect of smoking, whereas $\beta_{C(1)}\alpha_S$ is the product of the principal-strata controlled direct effect of $C(1)$ on ΔC (setting $S = s$, $L = l$, $U = u$) and the controlled direct effect of S on $C(1)$ (setting $U = u$), to produce an indirect effect of smoking through the pathway $S \to C(1) \to \Delta C$. Similarly, $\beta_{C(1)}\alpha_U\theta_S$ encodes an indirect effect of S through the pathway $S \to U \to C(1) \to \Delta C$, and finally $\beta_U\theta_S$ is an indirect effect of S through the pathway $S \to U \to \Delta C$. The equation in the above display states that the total causal effect of smoking is the sum of the effects through these various pathways. Although the parameters β_S, $\beta_{C(1)}$, β_U, and β_L have only a causal interpretation within principal strata, the parameters α_S, θ_S, α_S, and α_U are well defined as causal contrasts for all individuals. Model M_1 assumes that S has a constant effect on U and $C(1)$ in all individuals, in which case the decomposition in the above display is the average principal strata effect

$$E\{\Delta C_1 - \Delta C_0 | D_0 = D_1 = 0\}$$

The homogeneity assumption may not be realistic, as the $S \to C(1)$ pathway may be activated only in a subset of the population, and similarly the pathway $S \to U \to C(1)$ may be activated only for a subset of individuals. A generalization of M_1 accommodates such heterogeneity. Consider the model M_2 corresponding to the system of equations (3)-(7) together with equation (14). In this model, the causal effect of S on ΔC for individuals in the principal stratum $\{D_0 = D_1 = 0\}$ is as follows:

$$\beta_S + \beta_{C(1)}\,(C_1(1) - C_0(1)) + \beta_U\,(U_1 - U_0)$$

so that β_S may be interpreted as the principal-strata direct effect of smoking among individuals with

$$\{D_0 = D_1 = 0,\, C_1(1) = C_0(1),\, U_1 = U_0\}$$

In the Appendix, we prove the following result.

Result 1: Under the structural equation model M_2, we have that the causal contrast identified by the g-formula is

$$g(1) - g(0) = \beta_S + \beta_{C(1)}E\{C_1(1) - C_0(1)\} + \beta_U E\{U_1 - U_0\}$$

The result establishes that $g(s)$ correctly identifies a combination of causal effects of S on ΔC under the linear structural equation: a principal-strata direct effect of smoking, with indirect effects, which themselves combine principal-strata direct effects of endogenous variables, with the controlled direct effect of smoking on the latter variables. In the Appendix, we relax the linearity assumption of equation (14), and we provide a general nonparametric decomposition of the causal effect of S on ΔC identified by $g(s)$. The result also implies that, under the submodel M_1 of M_2, the g-formula, and therefore inverse-probability-weighted estimation, correctly identifies the average principal-strata effect of smoking to be

$$\beta_S + \beta_{C(1)}\alpha_S + \beta_{C(1)}\alpha_U\theta_S + \beta_U\theta_S$$

which is a combination of a direct principal-strata effect of smoking β_S, and indirect effects of smoking mediated through $(U,\, C(1))$.

In conclusion, instead of "weighting up the dead," inverse-probability-weighted analyses are an accessible and flexible approach for estimating the causal effects of smoking on cognitive change in the context of dependent truncation by death. Under an assumption that there are no unmeasured common causes of mortality and cognitive decline, we have characterized the causal effect that is identified by inverse-probability weighting survivors using a structural equation framework, and we have shown that this causal effect naturally incorporates principal-strata effects. Although the identifying assumption of "no unmeasured common cause of mortality and cognitive decline" cannot be confirmed with certainty empirically, methodology is currently available for conducting sensitivity analyses to assess the impact on inferences of a violation of this assumption in inverse-probability-weighted analyses.[7,8]

APPENDIX

Lemma

Consider the system of nonparametric structural equations (3)–(8). Then, we have that the causal contrast identified by the g-formula is as follows:

$$g(1) - g(0)$$

$$= \int E\{\Delta C_{s=1,w} - \Delta C_{s=0,w} | D_{s=1,w} = D_{s=0,w} = 0\} dF_{W_{s=0}}(w)$$

$$+ \int E\{\Delta C_{s=1,w} - \Delta C_{s=1,0} | D_{s=1,w} = D_{s=1,0} = 0\} \{dF_{W_{s=1}}(w) - dF_{W_{s=0}}(w)\}$$

$$= DE + IE$$

where $W = (C(1),\, L,\, U)$, $W_s = (C_s(1),\, L_s,\, U_s)$, with distribution $F_{Ws}(w)$ evaluated at w, $s = 0, 1$; DE is an average of principal-strata average direct effect and IE is the integral of an indirect effect of S on ΔC through W.

Proof of Lemma

Let $F_{\varepsilon_{\Delta C}}(\varepsilon)$ denote the distribution function of $\varepsilon_{\Delta C}$ evaluated at ε. Below, we use w^* to indicate a fixed baseline value of W. Evaluating the g-formula under the model gives

$$g(1) - g(0) = \left[\int\left\{\int g_{\Delta C}(s = 1, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\} dF_{W_{s=1}}(w)\right] - \left\{\int\int g_{\Delta C}(s = 0, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\} dF_{W_{s=0}}(w)$$

$$= \int\left\{\int g_{\Delta C}(s = 1, w, \varepsilon)\, dF_{\varepsilon_{\Delta c}}(\varepsilon)\right\} dF_{W_{s=0}}(w) - \left\{\int\int g_{\Delta C}(s = 0, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\} dF_{W_{s=0}}(w)$$

$$+ \left[\int\left\{\int g_{\Delta C}(s = 1, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\}\{dF_{W_{s=1}}(w) - dF_{W_{s=0}}(w)\}\right] - \underbrace{\left[\int\left\{\int g_{\Delta C}(s = 1, 0, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\}\{dF_{W_{s=1}}(w) - dF_{W_{s=0}}(w)\}\right]}_{= 0}$$

$$= \int\left[\begin{array}{l}\{\int g_{\Delta C}(s = 1, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\} \\ -\{\int g_{\Delta C}(s = 0, w, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\end{array}\right] dF_{W_{s=0}}(w)$$

$$+ \int\left[\left\{\int g_{\Delta C}(s = 1, w, \varepsilon) - g_{\Delta C}(s = 1, w^*, \varepsilon)\, dF_{\varepsilon_{\Delta C}}(\varepsilon)\right\}\right] \{dF_{W_{s=1}}(w) - dF_{W_{s=0}}(w)\}$$

by the independence assumptions encoded in the structural equations, which gives the result.

Proof of Result 1

The result is obtained by applying the Lemma with $g_{\Delta C}$ given by equation (14).

ACKNOWLEDGMENTS

We thank Jamie Robins, Andrea Rotnitzky, Tyler VanderWeele, and Miguel Hernan for very helpful discussions.

REFERENCES

1. Chaix B, Evans D, Merlo J, Suzuki E. Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology.* 2012;23:129–131.
2. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39:17–23.
3. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012;23.
4. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell N, Dietz K, Farewell V, eds. *AIDS Epidemiology—Methodological Issues*. Boston: Birkhäuser; 1992:297–331.
5. Robins JM. Causal inference from complex longitudinal data. Latent variable modeling and applications to causality. In: Berkane M, ed. *Lecture Notes in Statistics (120)*. New York: Springer Verlag; 1997:69–117.
6. Pearl J. *Causality: Models, Reasoning, and Inference.* 2nd ed. New York: Cambridge University Press; 2009.
7. Robins JM, Rotnitzky A, Scharfstein D. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology: The Environment and Clinical Trials. IMA.* Vol 116. New York: Springer-Verlag; 1999:1–92.
8. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for non-ignorable drop-out using semi-parametric non-response models [Comments and Rejoinder]. *J Am Statist Assoc*. 1999;94:1096–1120.