The International Journal of Biostatistics

Volume 6, *Issue* 2 2010 *Article* 11

CAUSAL INFERENCE

Selective Ignorability Assumptions in Causal Inference

Marshall M. Joffe, *University of Pennsylvania School of Medicine*
Wei Peter Yang, *University of Pennsylvania School of Medicine*
Harold I. Feldman, *University of Pennsylvania School of Medicine*

Recommended Citation:
Joffe, Marshall M.; Yang, Wei Peter; and Feldman, Harold I. (2010) "Selective Ignorability Assumptions in Causal Inference," *The International Journal of Biostatistics*: Vol. 6: Iss. 2, Article 11.
DOI: 10.2202/1557-4679.1199

Selective Ignorability Assumptions in Causal Inference

Marshall M. Joffe, Wei Peter Yang, and Harold I. Feldman

Abstract

Most attempts at causal inference in observational studies are based on assumptions that treatment assignment is ignorable. Such assumptions are usually made casually, largely because they justify the use of available statistical methods and not because they are truly believed. It will often be the case that it is plausible that conditional independence holds at least approximately for a subset but not all of the experience giving rise to one's data. Such selective ignorability assumptions may be used to derive valid causal inferences in conjunction with structural nested models. In this paper, we outline selective ignorability assumptions mathematically and sketch how they may be used along with otherwise standard G-estimation or likelihood-based methods to obtain inference on structural nested models. We also consider use of these assumptions in the presence of selective measurement error or missing data when the missingness is not at random. We motivate and illustrate our development by considering an analysis of an observational database to estimate the effect of erythropoietin use on mortality among hemodialysis patients.

KEYWORDS: causal inference, ignorability, end-stage renal disease, anemia

Author Notes: This work was supported by an unrestricted grant from Amgen.

1 Introduction

Most attempts at causal inference in observational studies are based on assumptions that treatment assignment is ignorable (Rosenbaum and Rubin 1983); ignorability involves conditional independence of treatment and the potential outcomes. Such assumptions are usually made casually, largely because they justify the use of available statistical methods and not because they are truly believed. It will often be the case that it is plausible that conditional independence holds at least approximately for a subset but not all of the experience giving rise to one's data. Such selective ignorability assumptions may be used to derive valid causal inferences in conjunction with structural nested models, whereas methods based on assuming full ignorability are biased in this setting. In this paper, we outline selective ignorability assumptions mathematically, discuss various variants of these assumptions, and sketch how these assumptions may be used along with otherwise standard G-estimation methods to obtain inference on structural nested models. We motivate and illustrate our development by considering an analysis of an observational database to estimate the effect of erythropoietin use on mortality among hemodialysis patients.

2 Potential outcomes, causal effects, and usual identifying assumptions

In this section, we outline a conceptual framework for causal effects based on potential outcomes. We do this first for scalar treatments which do not vary with time, then generalize to time-varying treatments.

2.1 Time-invariant treatments

The potential outcomes model (Neyman 1990;Rubin 1974) provides a useful framework for defining the effects of a treatment or exposure. For a scalar treatment, let A denote a subject's treatment, Y the outcome, and X measured pretreatment covariates. Let Y^a denote the outcome that would be seen in a subject were that subject to receive treatment level a. At the individual level, causal effects are contrasts of different potential outcomes Y^a and $Y^{a'}$, $a \neq a'$. At the group level, causal effects are contrasts of distributions of potential outcomes under different conditions for the same group; i.e., of $f(Y^a|\cdot)$ and $f(Y^{a'}|\cdot)$. These causal effects are not observable, since Y^a and $Y^{a'}$ are not simultaneously observable. Let

$\underline{Y} \equiv \{Y^a; a \in \boldsymbol{A}\}$ denote the set of all potential outcomes and $\boldsymbol{A}$ the set of all possible treatments.

The assumption of strongly ignorable treatment assignment (Rosenbaum and Rubin 1983) allows identification of causal effects from observed data on treatment, outcome, and covariates. The assumption may be written as

$$pr(A=a|X,\underline{Y}) = pr(A=a|X),$$
$$pr(A=a|X) > 0 \quad \forall a \in \boldsymbol{A}. \tag{1}$$

Variants of the assumption have also been proposed which still allow identification of treatment effects (Greenland and Robins 1986;Hernan and Robins 2009); for example, one might assume that

$$pr(A=a|X,Y^0) = pr(A=a|X). \tag{2}$$

Condition **(2)**, together with the assumption that $pr(A=0|X)>0$, is sufficient to identify the effect of treatment on the treated for a binary treatment, and, more generally, provides a basis for estimation in structural nested models. Alternatively, the conditional independence assumptions entailed by **(1)** and **(2)** may be written as $A \perp \underline{Y}|X$ or $A \perp Y^0|X$, respectively.

2.2 Time-varying treatments

For time-varying treatments, sequential versions of ignorability have been proposed. Let A_m denote the values of treatment provided at time m, L_m the value of covariates measured at m, Y_m the value of a repeated measures outcome at m, and T the value of a failure-time outcome. We include in L_m outcomes at m (i.e., Y_m and $T_m\{I(m-1 \le T_m < m)\}$ (the time of failures occuring between $m-1$ and m). We use overbars to denote history of a variable; thus, $\overline{A}_m \equiv \{A_0,...,A_m\}$ is the history of treatment through m, and $\overline{L}_m$ the history of covariates (and outcome) through m. We use the tilde symbol to denote future values of a variable; thus, $\tilde{Y}_m \equiv \{Y_{m+1}, Y_{m+2},...,Y_M\}$.

Potential outcomes and causal effects must now be indexed by treatment histories. Thus, we write the outcome that would be seen at m under treatment history $\overline{a}_M$ as $Y_m^{\overline{a}_M} = Y_m^{\overline{a}_{m-1}}$ and the vector of future outcomes at m as $\tilde{Y}_m^{\overline{a}_M}$. Denote

by $\{\bar{a}_m,0\}$ the treatment history $\bar{a}_m$ through *m* followed by receiving no treatment beyond that point in time. Thus, $Y_k^{\bar{A}_m,0}$,$k>m$ denotes the outcome that would be seen at *k* if someone followed his observed treatment history through *m* then received no treatment subsequently. Similarly, $\tilde{Y}_m^{\bar{A}_m,0}$ indicates the vector of potential outcomes beyond *m* that would have been seen had a subject received his or her observed treatment through *m* followed by no treatment subsequently.

One generalization of weak version of ignorability **(2)** to this setting is

$$pr(A_m=a_m|\bar{L}_m,\bar{A}_{m-1},\tilde{Y}_m^{\bar{A}_{m-1},0})=pr(A_m=a_m|\bar{L}_m,\bar{A}_{m-1}), \quad \textbf{(3)}$$

or, equivalently,

$$A_m \perp \tilde{Y}_m^{\bar{A}_{m-1},0}|\bar{L}_m,\bar{A}_{m-1}. \quad \textbf{(4)}$$

Robins et al. (Robins, Blevins, Ritter and Wulfsohn 1992) provides variant assumptions, some of which are more akin to strong ignorability. These sequential ignorability assumptions would be justified by a sequentially randomized trial. In such a trial, at each time *m*, decision-makers observe a subject's covariate levels and outcomes through that point in time, then assign a randomization probability to the subject based on that history. The probability of treatment may vary with current or previous covariates and treatment. The assumption has been termed "no unmeasured confounders" (Robins 1992;Robins et al. 1992), since there would be no confounding if enough confounding variables had been measured to render treatment at *m* and the potential outcomes conditionally independent.

3 Selective ignorability: motivation and formulation

Doubts about the appropriateness of ignorability assumptions (both simple and sequential) in practice has led to alternative approaches to inference. These alternatives include examining sensitivity of inference to departures from ignorability; methods for this have been proposed for both static (Greenland 1996;Rosenbaum 2002) and time-varying (Brumback, Hernan, Haneuse and Robins 2004;Robins, Rotnitzky and Scharfstein 2000) treatments. Other approaches include examining bounds for treatment effects (Manski 1990;Manski 1996) imposed nonparametrically or based on additional assumptions; these methods have been developed largely for static treatments. Usually, these approaches do not lead to

point identification of treatment effects, and, perhaps in part as a consequence, have not been widely adopted in practice.

We consider and develop an approach which allows careful tailoring of assumptions to the mechanisms underlying the treatment assignment process in one's data. The main idea is that it will often be plausible that conditional independence between potential outcomes and treatment holds in a subset of the experience under study but not in other subsets of that experience. We term such assumptions selective ignorability assumptions. We consider several ways in which selective ignorability might arise, motivated by an observational database used to estimate the effect of erythropoietin alpha (EPO) on mortality in hemodialysis patients. We then show how to formalize these ideas. In the next section, we show how these modified assumptions can lead to point identification and estimation of treatment effects, using structural nested models and G-estimation.

3.1 Departures from ignorability in the USRDS data

Hemodialysis is one of the main treatments for people with end-stage renal disease (ESRD). Subjects on hemodialysis tend to be anemic; consequently, EPO, a drug which stimulates red blood cell production, has become a central part of the treatment of anemia in hemodialysis. Questions have been raised about the safety of EPO, based both on randomized trials (Besarab et al. 1998;Drueke et al. 2006;Singh et al. 2006) and observational studies (Cotter et al. 2004).

The United States Renal Data System (USRDS) is a large database containing information on nearly all subjects in the United States on hemodialysis. We wanted to use this database to estimate the effect of EPO on mortality, but needed to account for the confounding role of time-invariant and time-varying predictors of EPO and mortality, especially hematocrit. We consider below various departures from ignorable treatment assignment in these or similar data.

3.1.1 Missing hematocrit information

Hematocrit is arguably the most important confounder of the effect of EPO in these data and is sometimes missing. Both current hematocrit and previous hematocrit are strong predictors of EPO use and of mortality (Yang, Joffe and Feldman 2009). Hematocrit is thus presumably a confounder of the effect of EPO. (There are other potential confounding variables to deal with as well; some of these are measured well, others not. The USRDS data has little information on other laboratory variables which vary over time; we presume as a working assumption that this missingness is not important. However, it does collect information on hospitalization and comorbidities as these vary over time; many of these variables are strong predictors

of EPO use as well as of mortality. Information on these variables is presumed to be reasonably complete.) Hematocrit is presumably the most important variable to control, because of its key role as an indication for EPO use and risk predictor.

Hematocrit is often missing in the USRDS. As a claims database, information on some variables is collected only when the medical provider reported it in order to justify a claim. For EPO, Medicare has had different rules regarding payment of claims for EPO. Hematocrit is generally recorded in the database only to justify a claim for EPO. Generally, the claim for EPO is submitted at the end of the month, and the rule is that the last hematocrit of the month is to be reported. Because EPO dose generally refers to the cumulative EPO dose over the period of reporting (most often a month), we might productively view the hematocrit in a given month as being affected by EPO use during that month. Our convention will generally be to record hematocrit measured in month m as occuring in month m+1 (in L_m), whereas EPO dose in that month will be recorded as A_m. Hematocrit information is missing in 10.6% of the eligible person-months; the proportion missing at least once in a 6 month period is higher. Thus, we have substantial missing information on the key time-varying confounding variable.

Because of the systematic nature of the missing information on hematocrit, the usual assumptions (e.g., missing at random (Little and Rubin 1987)) justifying the use of standard missing data methods do not apply. In particular, hematocrit is not missing at random, because the probability of missingness depends on the hematocrit. In this setting, if one believes that hematocrit (and other always measured covariates) are sufficient to render EPO use and the potential outcomes conditionally independent and so control for confounding of the effect of EPO, we might be willing to believe that this is true within the subset of the person-intervals in which complete confounder information is available. This is true especially if we believe that missingness of confounder information conveys no information about outcome conditional on the true confounder information. Proposition 1 formalizes this intuition.

Proposition 1: Let Δ_m be an indicator of whether complete information on confounders is available at m for a given subject. Suppose that treatment assignment is strongly ignorable, and that Δ_m may depend on A_m, $\overline{L}_m$, and $\overline{A}_{m-1}$ but not further on the potential outcomes $\tilde{Y}_m^{\overline{A}_{m-1},0}$; i.e., $pr(\Delta_m=1|A_m,\overline{L}_m,\overline{A}_{m-1},\tilde{Y}_m^{\overline{A}_{m-1},0}) = pr(\Delta_m=1|A_m,\overline{L}_m,\overline{A}_{m-1})$. Then, treatment assignment is strongly ignorable in the subset of the data in which complete confounder information is available; i.e.,

$$A_m \perp \tilde{Y}_m^{\overline{A}_{m-1},0} | \overline{L}_m,\overline{A}_{m-1},\Delta_m=1. \qquad \textbf{(5)}$$

We consider **(5)** a sort of selective ignorability assumption, since we assume ignorability holds in the subject of person-intervals in which complete confounder information is available. In general, if strong ignorability does not hold given measured covariates, (i.e., $pr(\Delta_m=1|A_m,\overline{L}_m,\overline{A}_{m-1},\tilde{Y}_m^{\overline{A}_{m-1},0})\neq pr(\Delta_m=1|A_m,\overline{L}_m,\overline{A}_{m-1})$), we will not have ignorability even in the subset of the data in which information on confounders is not missing (i.e., $\Delta_m=1$). We consider later how this proposition assists with estimation in the presence of missing confounder information.

3.1.2 Covariates adequate in a subset

uppose next that no covariates are missing, but that measured covariate information is insufficient to control for confounding in some identifiable subset of the data. That subset of the data might be defined by treatment level and/or covariate histories.

uppose, for example, that only moderate doses of EPO are explainable in terms of measured covariate histories. That is, doses at or close to zero or very high doses are poorly explained by the measured covariates alone, but intermediate dosing patterns are explainable in terms of measured covariate histories.

This was plausible in the USRDS data; subjects receiving very high or zero doses of EPO at *m* had hazards of mortality in the period shortly after *m* that were substantially higher than those with moderate doses. The increased hazards were present even after adjustment for covariates and were not much larger than could plausibly be explained by the effect of EPO.

There are other scenarios in which recorded information may be sufficient to control confounding in some subgroups of the data but not others. For example, suppose that covariate information is sufficient to control for confounding for subjects without certain disease conditions but insufficient for subjects with those conditions. This could arise if the severity of the condition varies and is associated with treatment decisions and independently associated with the outcome.

Joffe *et al*. (Joffe et al. 1998) provide an example in a different context where it was plausible that treatment initiation was ignorable in terms of measured covariates while discontinuation was not (i.e., ignorability held if $A_{m-1}=0$ but not if $A_{m-1}=1$).

3.1.3 Reverse causation

Ideas about selective ignorability may extend to dealing with "reverse causation" or "protopathic bias." Suppose that there is a latent condition that, if one remains untreated, will manifest itself within a given but fixed period of time. This latent

condition may influence treatment decisions and be associated with the outcome of interest. In our example, we might consider the latent condition impending death due to severe complications of ESRD. One might assume that that treatment decisions at *m* would be influenced by whether someone would die before $m+\delta$ if untreated, and, if so, by how close the impending death is. Later death, however, would not influence treatment decisions (beyond the association of prognosis with measured covariates). A version of this assumption has been considered in evaluating the effect of obesity (Robins 2008).

3.2 Selective ignorability: formulation

We now formalize the idea of selective ignorability. Let $s(\bar{a}_k,\bar{l}_k)$ be a user-specified binary function of its arguments. We say that treatment assignment is selectively ignorable if

$$A_m \perp \tilde{Y}_m^{\bar{A}_{m-1},0} | \bar{L}_m, \bar{A}_{m-1}, s(\bar{L}_m,\bar{A}_m)=1. \qquad \textbf{(6)}$$

In our first example, confounder information is adequate when hematocrit is measured, and so $s(\bar{L}_m,\bar{A}_m)=1$ if $\Delta_m=1$ (here, we include Δ_m in L_m). In our second example, where confounder information is adequate for intermediate EPO doses, $s(\bar{L}_m,\bar{A}_m)=1$ (0 otherwise) if $a_{\min}<A_m<a_{\max}$ for some user-specified values of $a_{\min}$ and $a_{\max}$. In the third example, $s(\bar{L}_m,\bar{A}_m)=1$ if $L_{m,1}=0$, where $L_{m,1}=1$ if a subject has a given disease state at *m*, 0 otherwise. In the last example, $s(\bar{L}_m,\bar{A}_m)=1$ if $A_{m-1}=0$.

To deal with "reverse causation," we consider assumptions of the form

$$A_m \perp T^{\bar{A}_{m-1},0} | \bar{L}_m, \bar{A}_{m-1}, s(\bar{L}_m,\bar{A}_m,T_m^{\bar{A}_{m-1},0})=1, T>m. \qquad \textbf{(7)}$$

Here, we would let $s(\bar{L}_m,\bar{A}_m,T^{\bar{A}_{m-1},0})=1$ if $T^{\bar{A}_{m-1},0}>m+\delta$, 0 otherwise. This states that treatment assignment at *m* is ignorable among the subset of subjects who would live at least δ more years. The function $s(\cdot)$ may be used to consider more complicated selective ignorability assumptions. For example, we could have $s(\bar{L}_m,\bar{A}_m,T^{\bar{A}_{m-1},0})=1$ if $T^{\bar{A}_{m-1},0}>m+\delta$ and $L_{m,1}=1$; this would say that treatment assignment at *m* is

ignorable for subjects who would live beyond $m+\delta$ and who do not have some measurable condition.

For outcomes measured at fixed time points, there are analogous assumptions allowing ignorability depending on some future potential outcomes. We might assume treatment assignment at *m* is ignorable for the subset of subject-intervals for whom (say) the next potential outcome is greater than some particular value (e.g., $Y_{m+1}^{\overline{A}_{m-1},0} < y$). Mathematically, we consider assumptions of the form

$$A_m \perp \tilde{Y}_m^{\overline{A}_{m-1},0} | \overline{L}_m, \overline{A}_{m-1}, s(\overline{L}_m, \overline{A}_m, \tilde{Y}_m^{\overline{A}_{m-1},0}) = 1. \quad \textbf{(8)}$$

Here, we would let $s(\overline{L}_m, \overline{A}_m, \tilde{Y}_m^{\overline{A}_{m-1},0}) = 1$ if $Y_{m+1}^{\overline{A}_{m-1},0} < y$, 0 otherwise. Inference under expanded selective ignorability conditions involving future potential outcomes $T^{\overline{A}_{m-1},0}$ or $\tilde{Y}_m^{\overline{A}_{m-1},0}$ will be more challenging than under selective ignorability conditions not involving future potential outcomes.

3.3 Selective mismeasurement of treatment

Finally, consider a setting in which the treatment of interest is measured well in some identifiable part of the data but measured poorly in another part. In the USRDS, EPO use while subjects are not in the hospital is recorded fairly reliably; however, EPO use while in the hospital is not reliably recorded. It is thus not clear what EPO use patterns are while in the hospital, and what are determinants of EPO dose; the available data do not provide a way to check these patterns. The dates and duration of hospitalization are reliably recorded in the database. Thus, one can identify the months in which EPO dose is recorded reliably, as well as the months in which EPO dose is not recorded reliably.

We formalize ideas here. Let *H* be a variable indicating the degree of accuracy with which *A* is measured; in the EPO example, *H* denotes the proportion of the month in which a subject was not hospitalized. *H*=1 indicates that *A* is measured accurately, whereas *H*=0 indicates that it is measured poorly or not at all. We consider generalizations of selective ignorability for this setting. First, we suppose that

$$A_m \perp \tilde{Y}_m^{\overline{A}_{m-1},0} | \overline{L}_m, \overline{A}_{m-1}, s(\overline{L}_m, \overline{A}_m, \tilde{Y}_m^{\overline{A}_{m-1},0}) = 1, H_m \quad \textbf{(9)}$$

We suppose here that H_m is not affected by A_m; we will reconsider this this in discussing the effect of EPO. Let A_m^* denote treatment measured at *m*. Since A_m is known precisely only if $H_m = 1$, **(9)** trivially implies

$$A_m^* \perp \tilde{Y}_m^{\bar{A}_{m-1},0} | \bar{L}_m, \bar{A}_{m-1}, s(\bar{L}_m, \bar{A}_m, \tilde{Y}_m^{\bar{A}_{m-1},0}) = 1, H_m = 1. \quad \textbf{(10)}$$

uppose further that a known part of A_m is measured accurately; denote this measurement by $A_m^\#$. In the EPO example, $A_m^\#$ might be the dose provided over the part of the month in which there is no hospitalization. Adapting **(9)**, we might suppose that the measured part of treatment is ignorable; i.e,

$$A_m^\# \perp \tilde{Y}_m^{\bar{A}_{m-1},0} | \bar{L}_m, \bar{A}_{m-1}, s(\bar{L}_m, \bar{A}_m, \tilde{Y}_m^{\bar{A}_{m-1},0}) = 1, H_m \quad \textbf{(11)}$$

Condition **(11)** is a sort of selective ignorability assumption, since it assumes that the selected part of treatment which is measured well is conditionally independent of the potential outcomes. **(11)** also implies **(10)**.

The above approach is problematic inasmuch as expressions **(10)** and **(11)** condition on previous true treatment history $\bar{A}_{m-1}$ even though this may not be measured accurately, and so cannot in general provide a direct basis for inference using measured variables. An analogue to **(10)** and **(11)** using only measured variables is

$$A_m^\# \perp \tilde{Y}_m^{\bar{A}_{m-1},0} | \bar{L}_m, \bar{A}_{m-1}^*, s(\bar{L}_m, \bar{A}_m^*, \tilde{Y}_m^{\bar{A}_{m-1},0}) = 1. \quad \textbf{(12)}$$

Here, we incorporate the conditions for measuring A_m accurately (i.e., H_m) into L_m. Further, A_m^* now refers to all aspects of measured treatment history, while $A_m^\#$ refers to a specific measured component of that history (e.g., the dose provided in the part of the month a subject is not hospitalized (in **(11)**), or the entire measured dose for the month (in **(10)**)). In **(12)**, $A_m^\#$ is an instrument for the effect of treatment starting at *m* (i.e., $\tilde{A}_{m-1}$). Full consideration of conditions under which **(12)** will be reasonable is beyond the scope of this paper; to this end, one might productively use directed acyclic graphs (Pearl 2000), including separate nodes for true and measured values of treatment variables.

4 Inference under selective ignorability

Nonparametric identification of treatment effects from longitudinal data requires a nonselective ignorability (Robins et al. 2000). In the presence of selective ignorability, some modeling assumptions will be required. We will consider inference for structural nested models, as these will allow inference based solely on the modeling assumptions about causal effects and the selective ignorability assumptions; inference can be based on simple modifications of G-estimation. In contrast, other methods for data on repeated treatments and covariates, including the G-formula and marginal structural models, require specification of the degree of departures from ignorability, and the methods sketched by Robins (Robins et al. 2000) might be used.

In this section, we first provide a brief review of structural nested distribution models. We then consider inference for those models under selective ignorability, using G-estimation, a semiparametric method. Because G-estimation can be problematic for estimating multiparameter causal models for failure-times, we also consider parametric likelihood-based inference for this setting.

4.1 Structural nested distribution models

tructural nested models provide a useful way to parametrize treatment effects; the presentation in this subsection largely follows that of Robins (Robins et al. 2000). We consider here structural nested distribution models, for which we assume that given $\overline{L}_{k-1},\overline{A}_{k-1}$, $\tilde{Y}_k$ has a continuous multivariate distribution with probability 1. Define the multivariate blip-down function $\gamma_k(\tilde{y}_k,\bar{l}_k,\bar{a}_k) \equiv \{\gamma_{k,k+1}(\tilde{y}_k,\bar{l}_k,\bar{a}_k),...,\gamma_{k,K}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)\}$ as the unique solution to

$$F_{\tilde{Y}_k^{\bar{a}_{k-1},0}|\bar{l}_k,\bar{a}_k}\{\gamma_k(\tilde{y}_k,\bar{l}_k,\bar{a}_k)\} = F_{\tilde{Y}_k^{\bar{a}_k,0}|\bar{l}_k,\bar{a}_k}(\tilde{y}_k) \quad \textbf{(13)}$$

satisfying

$$\gamma_{k,m}(\tilde{y}_k|\bar{l}_k,\bar{a}_k) \text{ is a function of } \bar{y}_K \text{ only through } \bar{y}_m \quad \textbf{(14)}$$

for $k=0,...,K-1$. Define $z_{k:m} \equiv \{z_k,...,z_m\}$. Then,

$$\gamma_{k,k+1}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=F^{-1}_{Y_{k+1}^{\bar{a}_{k-1},0}|\bar{l}_k,\bar{a}_k}\circ F_{Y_{k+1}^{\bar{a}_k,0}|\bar{l}_k,\bar{a}_k}(y_{k+1})$$

and

$$\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=F^{-1}_{Y_m^{\bar{a}_{k-1},0}|\bar{l}_k,\bar{a}_k,Y_{k+1:m-1}^{\bar{a}_{k-1},0}=\gamma_k^{k+1:m-1}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)}\circ F_{Y_m^{\bar{a}_k,0}|\bar{l}_k,\bar{a}_k,Y_{k+1:m-1}^{\bar{a}_k,0}=(y_{k+1},\ldots,y_{m-1})}(y_m)$$

The blip-down functions $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)$ map percentiles of the distribution of $Y_m^{\bar{a}_k,0}$, the potential outcomes that would be seen had a subject received treatment history $\bar{a}_k$ through *k* and level 0 of treatment thereafter, into percentiles of the distribution of $Y_m^{\bar{a}_{k-1},0}$, the potential outcomes that would be seen had the subject received treatment history $\bar{a}_{k-1}$ through *k*-1 and level 0 of treatment thereafter. Thus, the blip-down functions simulate removal of the effect of treatment at *k* after having removed the effects of all subsequent treatments.

For inference, we will need to parametrize the blip-down functions. Write $\gamma_k(\tilde{Y}_k,\bar{l}_k,\bar{a}_k)\in\{\gamma_k(\tilde{Y}_k,\bar{l}_k,\bar{a}_k;\Psi);\Psi\in\boldsymbol{\Psi}\}$. We consider some simple parametrizations of blip-down functions. Suppose that *K*=1 (i.e., a study with a scalar treatment which does not vary over the course of the study). One simple blip-down function is $\gamma_{0,1}(y_1,l_0,a_0)=y_1-a_0\Psi$. This is a simple location shift model in which the effect of treatment at baseline is proportional to the value of treatment (for an interval-scaled treatment), the effect of treatment does not vary with covariates L_0, and the shift is constant for all values of y_1. The model is implied by, but does not imply, a rank-preserving model in which the effect of a binary treatment is the same for each subject; this model has been used frequently in the literature with scalar treatments (Rosenbaum 2002). The model can easily be extended to allow for modification of the effect of treatment by baseline covariates: e.g., $\gamma_{0,1}(y_1,l_0,a_0;\Psi)=y_1-a_0\Psi_1-a_0l_0\Psi_2$, or to allow the effect of treatment to be greater in the upper or lower ranges of the outcome; e.g., $\gamma_{0,1}(y_1,l_0,a_0;\Psi)=y_1\exp(-a_0\Psi_2)-a_0\Psi_1$.

We consider next models for the effect of a vector treatment which may vary over time. A simple, one-parameter model, is $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi$; that is, treatment a_k at *k* increases the outcome at *m* by $a_k\Psi$ units. One can allow the effect of treatment to increase or decrease with time between treatment *k* and outcome *m*;

e.g., $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi_1-a_k\exp\{(m-k)\Psi_2\}$ or the more restrictive $\gamma_{k,m}(\underline{y}_{k+1},\bar{l}_k,\bar{a}_k)=y_m-a_k/(m-k)\Psi_1$. One can also allow the effect of treatment at *k* to vary with past covariate and treatment histories; e.g., $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi_1-a_k l_k\Psi_2$. Realistic models for the effect of EPO will require that the effect of EPO vary with time-varying covariates (especially hematocrit), as well as allowing nonlinear dose-response functions. A simple model incorporating all of this is

$$\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi_1-a_k l_k\Psi_2-a_k^2 l_k\Psi_3. \quad \textbf{(17)}$$

Define recursively for $k=K,K-1,...,0$

$$u_k(\bar{y}_K,\bar{l}_K,\bar{a}_K)=\gamma_k[\{y_{k+1},u_{k+1}(\bar{y}_{K+1},\bar{l}_K,\bar{a}_K)\},\bar{l}_m,\bar{a}_m] \quad \textbf{(18)}$$

with $u_{K-1}(\bar{y}_{K+1},\bar{l}_K,\bar{a}_K)\equiv\gamma_K(\underline{y}_{K+1},\bar{l}_K,\bar{a}_K)$. Define $U_k=u_k(\bar{Y}_{K+1},\bar{L}_K,\bar{A}_K)$; U_k is a random vector of the same dimension as $\underline{Y}_{k+1}$. Further, define , and define $U_{k,m}$, $m>k$ as the component of U_k corresponding to the blipped-down Y_m. Robins et al. (Robins et al. 2000) show that

$$pr(U_k>\tilde{y}_k|\bar{l}_k,\bar{a}_k)=pr(\tilde{Y}_k^{\bar{a}_{k-1},0}>\tilde{y}_k|\bar{l}_k,\bar{a}_k) \quad \textbf{(19)}$$

Define $U_k(\Psi)$ as the value of U_k we would calculate from the data using a presumed value of Ψ in the blip-down functions $\gamma_k(\cdot)$. Thus, we can recursively create a random variable $U_k(\Psi)$ from observable data and a presumed model for treatment effects which, if our presumed model is correct, has the same conditional distribution at *k* (given treatment and covariate histories through *k*) as the potential outcome that would have been seen among these subjects had they received level 0 of treatment from *k* on. This fact will be useful in estimation. As an example, for model **(17)**, we have

$$U_{k,m}(\psi)=Y_m-(\sum_k^m a_k)\psi_1-(\sum_k^m a_k l_k)\psi_2-(\sum_k^m a_k^2)\psi_3. \quad \textbf{(20)}$$

DOI: 10.2202/1557-4679.1199

4.2 G-estimation and testing

4.2.1 Standard G-estimation

tandard G-estimation may be viewed as being based on inverting tests of sequential ignorability under an assumed model for treatment effects (Robins 1992;Robins et al. 1992). Sequential ignorability **(4)**, together with an assumed model for treatment effect, implies

$$A_m \perp U_m(\psi) | \overline{L}_m, \overline{A}_{m-1} \qquad \textbf{(21)}$$

if $\psi = \psi_0$, where ψ_0 indicates the true value of ψ. For G-estimation, one inverts an appropriate test of **(21)** (with the same number of degrees of freedom as the dimension of ψ).

The problem may be formulated in terms of estimating functions. Appropriate estimating functions will depend on the nature of the causal model parametrized by ψ; the theory of semiparametric inference may be used to derive (asymptotically) the class of all estimating functions as well as the most efficient member of the class (Robins 1992;Robins et al. 1992;Tsiatis 2006). For **(20)**, we suggest the functions

$$\begin{aligned} \textstyle\sum_{i,m} (A_{i,m} - p_{i,m}^{[1]}) \epsilon_{i,m}(\psi) &= 0 \\ \textstyle\sum_{i,m} (A_{i,m} - p_{i,m}^{[1]}) L_{i,m} \epsilon_{i,m}(\psi) &= 0 \\ \textstyle\sum_{i,m} (A_{i,m}^2 - p_{i,m}^{[2]}) \epsilon_{i,m}(\psi) &= 0, \end{aligned} \qquad \textbf{(22)}$$

where $p_{i,m}^{[j]} \equiv E(A_{i,m}^j | \overline{L}_m, \overline{A}_{m-1})$, $\epsilon_{i,m}(\psi) \equiv 1/(M-m) \sum_{k=m+1}^{M} \epsilon_{i,m,k}(\psi)$, $\epsilon_{i,m,k}(\psi) \equiv U_{i,m,k}(\psi) - \mu_{i,m,k}(\psi)$, and $\mu_{i,m,k}(\psi) \equiv E\{U_{i,m,k}(\psi) | \overline{L}_m, \overline{A}_{m-1}\}$. Both $p_{i,m}^{[j]}$ and $\mu_{i,m,k}(\psi)$ are unknown parameters which must be estimated from the data. Estimates based on **(22)** are doubly robust (Bang and Robins 2005), in that estimates of ψ are consistent if either the model for $p_{i,m}^{[j]}$ or the model for $\mu_{i,m,k}(\psi)$ is correct. These estimators are (locally) efficient if the treatment and outcome are scalars (i.e., not repeated measures) and $\epsilon_{i,m,k}(\psi)$ is independent and identically distributed with a normal distribution.

4.2.2 Modifications for selective ignorability

Under selective ignorability, we substitute $U_m(\psi)$ for $\tilde{Y}_m^{\bar{A}_{m-1},0}$ in **(12)**, yielding

$$A_m^{\#} \perp U_m(\psi) | \bar{L}_m, \bar{A}_{m-1}^*, s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\} = 1. \quad \textbf{(23)}$$

We would again like to base inference on testing the conditional independence in **(23)**, and would like to use estimating functions like **(22)**. Several modifications are required.

First, the sum in the estimating equations must now be over all subject-intervals i,m for which $s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\} = 1$, as the conditional independence in **(23)** applies only to that subset. Implications of this vary according to whether $s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\}$ is a function of $U_m(\psi)$. If $s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\}$ is not a function of $U_m(\psi)$, then inference may proceed as in standard G-estimation, using or summing over only the subset of subject-intervals for which $s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\} = 1$ in **(22)**. This approach is valid, because the conditional independence in **(23)** implies that the expectation of the estimating functions in **(22)** (or other appropriate estimating functions) is 0 for the true value of ψ.

If $s\{\bar{L}_m, \bar{A}_m^*, U_m(\psi)\}$ is a function of $U_m(\psi)$, a couple of modifications are necessary. First, in **(22)**, the probability of treatment in each interval i,m corresponding to condition **(23)** will now be a function of ψ, since that probability may depend on future putative potential outcomes $U_m(\psi)$. We will thus need to substitute $p_{i,m}^{[j]}(\psi) \equiv E[A_{i,m}^j | \bar{L}_m, \bar{A}_{m-1}, s\{\bar{L}_m, \bar{A}_{m-1}, U_m(\psi)\} = 1]$ for $p_{i,m}^{[j]}$ and $\mu_{i,m,k}^*(\psi) \equiv E[U_{i,m,k}(\psi) | \bar{L}_m, \bar{A}_{m-1}, s\{\bar{L}_m, \bar{A}_m, U_{i,m}(\psi)\} = 1]$ for $\mu_{i,m,k}(\psi)$.

In addition, the sums in **(22)** will, in general, be discontinuous in ψ, both because the sums involved will be over different subject-intervals i,m for different values of ψ, and because $p_{i,m}^{[j]}(\psi)$ and $\mu_{i,m,k}^*(\psi)$ are discontinuous in ψ. This, in turn, will have two consequences. First, efficient algorithms for finding the root of multivariate estimating equations, such as Newton-Raphson, will no longer work. The higher dimensional the causal parameter, the more problematic this will become. In addition, the causal model indexed by ψ will almost inevitably be misspecified when the treatment is continuous-valued and/or there are interactions between the treatment and a high-dimensional covariate. Under such inevitable model specification, there may be no values of ψ consistent with the data. We have found

similar problems for structural nested failure-time models (SNFTMs), where, as here, artificial censoring requires using different parts of the data for testing different hypotheses ψ; it is likely that the part of the data used for testing one value of ψ are most consistent with a value ψ^1 of the parameter, whereas the part of the data used for testing another value of ψ are most consistent with a different value (ψ^2) of the parameter. We explore these issues in greater detail elsewhere (Joffe, Yang and Feldman 2009).

4.2.3 Missing data and measurement error

Inference is complicated in the presence of missing data and measurement error. Above, we considered situations in which some version of selective ignorability was justified on the basis of missing data on covariates (e.g., hematocrit) or error in the measurement of treatment. Even when selective ignorability applies, inference on treatment effects is more limited under these circumstances.

Inference based on testing the conditional independence in **(23)** depends on being able to calculate all the quantities in **(23)** from observable data. In particular, we have supposed in previous sections that $A_m^{\#}$ is available for all subjects. However, $U_m(\psi)$ may not be available if treatment or covariates are mismeasured on not measured at all. If ψ=0, then $U_m(\psi)=\tilde{Y}_m$, an observable quantity. Thus, validity tests of the null hypothesis based on testing the conditional independence of $A_m^{\#}$ and $U_m(\psi)$ does not depend on correct measurement of A_k or L_k, $k>m$.

Tests of non-null hypotheses ψ will, in general, depend on correct measurement of any variables used to calculate $U_m(\psi)$ from observable data and ψ. Thus, if there are no covariates in the causal model for the effect of treatment (e.g., if we assume $\psi_2=0$ in **(20)**), we can obtain consistent estimates of ψ if treatment is measured accurately and selective ignorability holds. In our EPO example, we require models that allow the effect of EPO to vary with hematocrit, as it is implausible that the same dose of EPO has the same effect for subjects with normal hematocrit and for subjects who are severely anemic. Notably, to the extent that information on hematocrit is missing only in months in which there is no EPO, we can still compute $U_m(\psi)$ under model **(17)** ($\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi_1-a_k l_k\Psi_2-a_k^2 l_k\Psi_3$), as, whenever L_k is missing, $A_k=0$ and so $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)$ can be computed. Here, the choice of the baseline value of treatment to be a zero dose (i.e., $A_k=0$ represents a zero dose) is consequential; choice of a different dose as the baseline or reference

level would prevent computing $U_m(\psi)$ under this missingness scheme and causal model. If the model allows the effect of EPO at m to vary also with prior hematocrit (e.g, by modifying **(17)** to $\gamma_{k,m}(\tilde{y}_k,\bar{l}_k,\bar{a}_k)=y_m-a_k\Psi_1-a_k l_{k-1}\Psi_2-a_k^2\Psi_3$), we can no longer compute $U_m(\psi)$ under the given pattern of missingness for hematocrit.

In contrast with mismeasurment of covariates, in the presence of error in measurement of treatment, selective ignorability assumptions cannot yield consistent estimates. In this setting, it is worth examining the potential benefits and disadvantages of inference based on selective ignorability assumptions compared with inference under standard ignorability assumptions. Inference in **(22)** under modified assumptions **(23)** is based on components $\sum_{i,m}(A_{i,m}^{\#\ j}-p_{i,m}^{[j]})g(\bar{L}_{i,m})\epsilon_{i,m}(\psi)=0$ (here $g(\bar{L}_{i,m})=1$ for the first and third components equations in **(22)** and $g(\bar{L}_{i,m})=L_{i,m}$ for the second component equation). $A_{i,k}$ may contribute to the estimating function directly in two terms: through $A_{i,m}^{\#}$ and through $\epsilon_{i,m}(\psi)$, which is a function of $U_m(\psi)$, which is, in turn, a function of $A_{i,k}, k\geq m$. Thus, error in $A_{i,k}$ can influence estimation through both of those terms. Under selective ignorability, the sum may be only over intervals i,m in which $A_{i,m}^{\#}$ is known accurately; thus, error in $A_{i,k}$ can influence estimation only through $\epsilon_{i,m}(\psi)$. We would thus expect that estimates based on selective ignorability assumptions will, in general, be both less biased and less sensitive to assumptions about the missing or mismeasured values of $A_{i,k}$.

imilar arguments would suggest that, in models where the effect of treatment is modified by sometimes mismeasured covariates, selective ignorability assumptions (under which the sum is only over subject-intervals i,m in which $g(\bar{L}_{i,m})$ is known accurately) would lead to less influence of and bias from mismeasurement of the covariate.

We formalize these ideas below. Let $\psi_{jk}(\theta)$ denote an estimator of ψ, and let $\tilde{\psi}_{jk}(\theta)$ denote its large-sample expectation. We use j to index the estimating function used, k to index the ignorability assumptions made, and θ to index the method used for imputing the missing or mismeasured values of treatment and/or covariates. The large-sample bias of an estimator is $\tilde{\psi}_{jk}(\theta)-\psi_0$. Let Θ denote the range of values we are prepared to entertain for θ. We conjecture that the maximum bias $sup_\Theta\{|\tilde{\psi}_{jk}(\theta)-\psi_0|\}$ is smaller for estimators k making appropriate selective ignorability assumptions. Similarly, the variability of an estimator is represented by $sup_\Theta\{|\tilde{\psi}_{jk}(\theta)-\tilde{\psi}_{jk}(\theta')|\}$, where θ and θ' are two different values, and we conjecture that this quantity will be smaller for estimators k making appropriate selective

ignorability assumptions. In a Bayesian formulation, we presume a prior distribution for θ, $f_\theta(\theta;\lambda)$ indexed by λ. Here, the bias is $E_\lambda\{\psi_{jk}(\theta)-\psi_0\}$, the variance is $Var_\lambda\{\psi_{jk}(\theta)\}$ (a measure of sensitivity of inference to assumptions about θ), and the mean-squared error is $E_\lambda[\{\psi_{jk}(\theta)-\psi_0\}^2]$. One can also formulate minimax or Bayes versions of risk. Careful attention is needed for formulation of these measures for vector ψ and θ. We hope to evaluate these conjectures about selective ignorability assumptions being able to reduce sensitivity and maximum bias or risk through simulation or theoretical study.

An additional issue involving measurement error arises when $s\{\bar{L}_m,\bar{A}^*_m,U_m(\psi)\}$ is a function of $U_m(\psi)$. The subset $s\{\bar{L}_m,\bar{A}^*_m,U_m(\psi)\}=1$ in which ignorability is presumed to hold will not be identified correctly in the presence of measurement error.

4.3 Parametric likelihood-based inference

In attempting to estimate the effect of EPO on mortality using the USRDS data, we encountered difficulties in using G-estimation, presumably because of issues arising from artificial censoring (Joffe and Brensinger 2001;Joffe et al. 2009). These issues include difficulty in finding appropriate solutions to the estimating equations, and uncertainty about the correct function to optimize in the presence of misspecification of the causal model. The motivations for basing inference on selective ignorability assumptions remain, though. Thus, we consider likelihood-based methods as an alternative for inference under selective ignorability. We again do this for a repeated-measures continuous outcome; with some modification, the methods presented and issues raised here will apply to failure-time outcomes as well. Our discussion below presumes that there are no missing data (which is involved in several of the presumed mechanisms leading to selective ignorability in the USRDS data).

For likelihood-based inference, we need to formulate a model or likelihood for the data. In this, we follow and modify the corresponding formulation for failure-time outcomes (Robins 1992;Robins et al. 1992). We express the likelihood as follows:

$$f_{\bar{Y}_K,\bar{L}_K,\bar{A}_K}(\bar{y}_K,\bar{l}_K,\bar{a}_K)=\frac{\Delta U_0}{\Delta \bar{Y}_k} f(U_0,\bar{L}_K,\bar{A}_K), \qquad \textbf{(24)}$$

where $\frac{\Delta U_0}{\Delta \bar{Y}_k} \equiv \prod_{k=1}^{K} \frac{\partial U_{0,k}}{\partial Y_k}$, $U_0 \equiv U_0(\psi_0)$, and ψ_0 represents the true value of ψ. The joint density $f(U_0, \bar{L}_K, \bar{A}_K)$ factorizes further as

$$f(U_0, \bar{L}_K, \bar{A}_K; \alpha) = f(U_0; \alpha_1) \prod_{m=0}^{K-1} \{ f(L_m | U_0, \bar{L}_{m-1}, \bar{A}_{m-1}; \alpha_2) f(A_m | U_0, \bar{L}_m, \bar{A}_{m-1}; \alpha_3) \}, \quad \textbf{(25)}$$

where α is a parameter indexing the given densities. Under ignorable treatment assignment, $f(A_m | U_0, \bar{L}_m, \bar{A}_{m-1}) = f(A_m | \bar{L}_m, \bar{A}_{m-1})$, and the density $f(A_m | \bar{L}_m, \bar{A}_{m-1})$ contains no information about ψ. Likelihood-based inference proceeds by maximizing the remaining parts of the likelihood.

For inference under selective ignorability, we need to formulate models for the data which allow A_m to depend on U_0 for the subset of the data in which $s\{\bar{L}_m, \bar{A}^*_m, U_m(\psi)\} = 0$ but restrict A_m to be independent of U_0 when $s\{\bar{L}_m, \bar{A}^*_m, U_m(\psi)\} = 1$. For a simple example, suppose that $s\{\bar{L}_m, \bar{A}^*_m, U_m(\psi)\} = I(L_m > l)$, and, for a binary A_m, that

$$\text{logit}\{p(A_m = 1 | \bar{L}_m, \bar{A}_{m-1}, U_0)\} = \beta_0 + L_m \beta_1 + A_{m-1} \beta_2 + I(L_m > l) U_{0,m+1}(\psi) \beta_3. \quad \textbf{(26)}$$

In principle, likelihood-based inference presents considerable challenges. First, correct specification of models will be difficult. One must correctly specify models for the treatment process and the covariate process; in many applications, the covariate process will be of very high dimension, complicating correct specification. Further complicating correct specification is the dependence of the treatment and covariates processes on the unmeasured potential outcomes. In principle, G-estimation and our modifications thereof circumvent these problems by not requiring correct specification of the association of the covariates or treatment and potential outcomes to obtain consistent estimation. It may be possible to mitigate bias due to misspecification through the use of targeted maximum likelihood methods (van der Laan and Rubin 2006).

In addition, the form of dependence of A_m on $U_0(\psi)$ will convey information about ψ. This will be undesirable and may lead to bias in estimation of ψ if the form of the dependence is chosen based on convenience rather than knowledge. Less information may be conveyed about ψ through more flexible specifications of this

dependence (e.g., replacing the linear dependence on $U_{0,m+1}(\psi)$ in **(26)** by a third-order polynomial expression).

Computational difficulties may arise in attempting to maximize the likelihood, due to the complicated nature of the likelihood. These difficulties will likely be magnified in likelihood-based inference for SNFTMs, where the outcome is a censored failure-time. Here, the likelihood for censored observations involves integration, and so computing the likelihood may require numerical integration. An alternative approach would involve Markov Chain Monte Carlo estimation in a Bayesian framework, which would, in turn, require specification of a prior distribution for a very high-dimensional parameter.

5 Discussion

We have proposed to use semiparametric or parametric methods for estimation of treatment effects under selective ignorability assumptions. Our approaches requires the investigator to make judgments about which subsets of the data it is plausible that there is conditional independence between treatment and potential outcomes. However, they do not require making judgment about the extent of nonignorability in those subsets. This contrasts with a sensitivity analysis approach, which would require such judgments, which may be difficult to make (Robins 2002) and are not as natural as assumptions about conditional independence (Pearl 2000). Unlike the sensitivity analysis approach, our approaches produce point identification of treatment effects, which may make conclusions easier to present in applied journals. In observational data, these assumptions are, at best, approximations to the truth and should be made on the basis of subject-matter knowledge. We have attempted to show how such assumptions arise naturally in consideration of the USRDS data.

We have outlined methods for inference under selective ignorability for structural nested distribution models, and, to some extent, SNFTMs. Both types of models map distributions of the observed outcome to the distribution of a baseline potential outcome that would have been seen had a baseline treatment been given instead. Structural nested mean models (SNMMs) are a less restrictive class of models for the joint causal effects of treatments provided at different times. Parameters in these models may also be estimated in conjunction with selective ignorability assumptions. Here, we can weaken the identifying assumptions (e.g., in **(23)**) to selective conditional uncorrelateness of $A_m^{\#}$ and $U_m(\psi)$ (a variable whose interpretation is different in SNMMs than SNDMs). A further important difference with these models is that the selection function $s(\cdot)$ cannot be a function of the future potential outcomes, but only of observed prior treatment and covariates $\overline{A}_m$ and $\overline{L}_m$.

For analysis of the USRDS data, which involve failure-time outcomes and repeated treatments, extensions of these mean models to discretized failure-time outcomes are needed. Various options have been proposed; none of these involve artificial censoring, and so the problems encountered with SNFTMs may not apply. The approach using a log link (Robins et al. 2000;Page, Hernan and Robins 2009) in principle allows semiparametric estimation; however, it may require that the outcome be rare, and so is problematic in the USRDS data. The logit link does not have this limitation. However, semiparametric estimation is not possible, even for a scalar treatment and a binary outcome (Robins et al. 2000). Thus, estimation of parameters in nuisance models is required (Robins and Rotnitzky 2004;Vansteelandt and Goetghebeur 2003). These may be substantially more involved for sequential treatments and discretized failure-time outcomes. Thus, extension of structural nested models and associated estimation methods to discretized failure-time outcomes potentially poses problems for dealing with the USRDS data.

An alternative approach, easily implemented for failure-time outcomes, is an intention-to-treat analysis of observational data (Hernan et al. 2008). Here, one simply looks at the association between treatment A_m and subsequent outcome, summarizing across the treatment times *m*. Implementation is simple and easily allows for inference under selective ignorability assumptions not involving future potential outcomes (i.e., that $s(\cdot)$ is not a function of $\tilde{Y}_m^{\bar{A}_{m-1},0}$). A drawback of this approach is that it does not attempt to estimate the effect of treatment levels actually received, as is true of intention-to-treat analyses of randomized trials. We have used this approach, both in conjunction with and not in conjunction with selective ignorability assumptions in analyzing the USRDS data (Yang et al. 2009).

In summary, selective ignorability assumptions can, in principle, provide a basis for inference about the effects of a time-varying treatment under more reasonable assumptions than are typically made, and allow careful tailoring of one's analysis to assumptions that the analyst is prepared to make on the basis of knowledge of one's data. When dealing with failure-time outcomes, implementation is complicated by problems involved with estimation of structural nested models for failure-times.

DOI: 10.2202/1557-4679.1199

Reference list

Bang, H., and Robins, J. (2005), "Doubly robust estimation in missing data and causal inference models," *Biometrics*, 61, 962-972.

Besarab, A., Bolton, W. K., Browne, J. K., Egrie, J. C., Nissenson, A. R., Okamoto, D. M., Schwab, S. J., and Goodkin, D. A. (1998), "The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin," *New England Journal of Medicine*, 339, 584-590.

Brumback, B. A., Hernan, M. A., Haneuse, S. J. P. A., and Robins, J. M. (2004), "Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures," *Statistics in Medicine*, 23, 749-767.

Cotter, D. J., Stefanik, K., Zhang, Y., Thamer, M., Scharfstein, D., and Kaufman, J. (2004), "Hematocrit was not validated as a surrogate end point for survival among epoetin-treated hemodialysis patients," *Journal of Clinical Epidemiology*, 57, 1086-1095.

Drueke, T. B., Locatelli, F., Clyne, N., Eckardt, K.-U., Macdougall, I. C., Tsakiris, D., Burger, H.-U., and Scherhag, A. (2006), "Normalization of hemoglobin level in patients with chronic kidney disease and anemia," *New England Journal of Medicine*, 355, 2071-2084.

Greenland, S. (1996), "Basic methods for sensitivity analysis of biases," *International Journal of Epidemiology*, 25, 1107-1116.

Greenland, S., and Robins, J. M. (1986), "Identifiability, exchangeability, and epidemiological confounding," *International Journal of Epidemiology*, 15, 412-418.

Hernan, M. A., Alonso, A., Logan, R., Grodstein, F., Michels, K. B., Willett, W. C., Manson, J. E., and Robins, J. M. (2008), "Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease," *Epidemiology*, 19, 766-779.

Hernan, M. A., and Robins, J. M. (2009), *Causal Inference* Chapman and Hall/CRC.

Joffe, M. M., and Brensinger, C. (2001), "Administrative and Artificial Censoring in Censored Regression Models," *Statistics in Medicine*, 20, 2287-2304.

Joffe, M. M., Hoover, D. R., Jacobson, L. P., Kingsley, L., Chmiel, J. S., Visscher, B. R., and Robins, J. M. (1998), "Estimating the effect of zidovudine on Kaposi's sarcoma using a rank preserving structural failure-time model," *Statistics in Medicine*, 17, 1073-1102.

Joffe, M. M., Yang, W., and Feldman, H. I. (2009), G-estimation and artificial censoring: problems, challenges, and applications," *manuscript*.

Little, R. J., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data* New York: Wiley.

Manski, C. F. (1996), "Monotone treatment response," *Econometrica*, 65, 1311-1334.

----- (1990), "Nonparametric bounds on treatment effects," *American Economic Review Papers and Proceedings*, 80, 319-323.

Neyman, J. (1990), "On the application of probability theory to agricultural experiments. Essay on principles. Translated by D.M. Dabrowska and edited by T. P. Speed," *Statistical Science*, 5, 465-472.

Page, J. H., Hernan, M. A., and Robins, J. M. (2009), "Structural nested cumulative failure time models," *manuscript*.

Pearl, J. (2000), *Causality: models, reasoning, and inference* Cambridge University Press.

Robins, J. (1992), "Estimation of the time-dependent accelerated failure time model in the presence of confounding factors," *Biometrika*, 79, 321-334.

Robins, J. M. (2008), "Causal models for estimating the effects of weight gain on mortality," *International Journal of Obesity*, 32, S15-S41.

----- (2002), "Comment on "Covariance adjustment in randomized experiments and observational studies" by Paul R. Rosenbaum," *Statistical Science*, 17, 309-321.

Robins, J. M., Blevins, D., Ritter, G., and Wulfsohn, M. (1992), "G-estimation of the effect of prophylaxis therapy for pneumocystic carinii pneumonia on the survival of AIDS patients," *Epidemiology*, 3, 319-336.

Robins, J. M., and Rotnitzky, A. (2004), "Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome using structural mean models," *Biometrika*, 91, 763-783.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (2000), "Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models," in *Statistical Models in Epidemiology*, eds. E. Halloran and D. Berry, New York: Springer-Verlag.

Rosenbaum, P. R. (2002), *Observational Studies* New York: Springer-Verlag.

Rosenbaum, P. R., and Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41-55.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688-701.

ingh, A. K., Szczech, L., Tang, K. L., Barnhart, H., Sapp, S., Wolfson, M., and Reddan, D. (2006), "Correction of anemia with epoetin alfa in chronic kidney disease," *New England Journal of Medicine*, 355, 2085-2098.

Tsiatis, A. A. (2006), *Semiparametric Theory and Missing Data* New York: pringer.

van der Laan, M. J., and Rubin, D. (2006), "Targeted maximum likelihood learning," *International Journal of Biostatistics*, 2, 11.

Vansteelandt, S., and Goetghebeur (2003), "Causal inference with generalized structural mean models," *Journal of the Royal Statistical Society, Series B*, 65, 817-835.

Yang, W., Joffe, M. M., and Feldman, H. I. (2009), "Exploring the effect of erythopoeitin on mortality using USRDS data," *manuscript*.