

**Rejoinder by JM Robins, S Greenland, and FC Hu to Discussion of
“Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a
Repeated Binary Outcome”**

1. Introduction

We thank both the editor for organizing this discussion and the discussants for their comments. We are gratified that Wasserman, Rubin and Frangakis (hereafter RF), and Zeger and Liang (ZL) recognize the soundness and importance of our approach. We found the issues raised by Wasserman to be particularly challenging. Therefore, we have chosen to devote the largest part of our discussion to the two central questions he poses. First, what is the null paradox and how do the causal models of Section 8 succeed in circumventing it? Second, what sensitivity analysis methods are available for the models of our paper?

Causal Inference from Complex Longitudinal data – An Overview: Robins (1986, 1987) proposed a formal theory of causal inference that extended Neyman’s (1923) time-independent point treatment theory to longitudinal studies with direct and indirect effects and sequential, time-varying treatments and confounders. As noted by Wasserman, with time-independent treatments, the principal use of a formal theory has been to justify usual practice [although see Copas (1973), Rosenbaum and Rubin (1983), Robins (1988), Robins, Mark, and Newey (1992), and Rosenbaum (1995) for exceptions]; the real payoff comes when there are sequential or time-varying treatments. For example, in the context of a randomized trial with non-compliance, Robins (1989) characterized the dependence of bounds for the average causal affect (ACE) on three key assumptions – the exclusion restriction, the monotonicity assumption, and the randomization (ignorability) assumption – by conceptualizing the randomization and actual treatment indicators as sequential treatments. Angrist, Imbens and Rubin (1996) later studied the ACE among the compliers under these same assumptions.

The greatest benefit comes with time-varying treatments, since then standard methods can be biased even when the causal null is true and there are no unmeasured confounders. Indeed, contrary to what RF appear to suggest, prior to Robins (1986) the relevant causal estimands had not even been formally defined, much less had conditions for their identification or methods for their estimation been derived. In fact, as recently as 1991, Rubin (1991) analyzed the effect of a time-dependent treatment using methods only appropriate for time-independent treatments. In a final comment, Rubin (1991, p. 1232) acknowledged that his approach remained “less than fully satisfactory” and that the approach of Robins (1989) was “more realistic,” although “highly model-dependent, at least for some outcomes.” With highly complex longitudinal data, model dependence is unavoidable. However, the semiparametric causal models discussed in Sec. 8 of our paper and Sec. 2d below were developed to minimize the degree of model dependence.

Following Wasserman, suppose the outcome $Y = Y_{M+1}$ is measured only at the end of the study. We observe n i.i.d. copies of temporally ordered data $X, L_0, A_0, \dots, L_M, A_M, Y$, where the X are time-independent covariates, the L_m are time-varying covariates, the A_m are treatments. Let $\bar{A}_m = (A_0, \dots, A_m)$, let $V(\bar{a}) = (Y(\bar{a}), \bar{L}(\bar{a}))$ be the counterfactual outcomes under treatment regime $\bar{a} = (a_0, \dots, a_M)$, let $\bar{\mathcal{A}}$ denote the support of the random variable \bar{A} , and let $V(\cdot) = (Y(\cdot), \bar{L}(\cdot)) = \{V(\bar{a}); \bar{a} \in \bar{\mathcal{A}}\}$ be

the random function (i.e., stochastic process indexed by \bar{a}) of counterfactual outcomes. Our initial approach is to map our observational study to a hypothetical sequential randomized clinical trial (RCT), i.e., a trial in which the treatment A_m on day m is chosen at random with randomization probabilities $f(A_m | \bar{L}_m, \bar{A}_{m-1}, X)$ possibly depending on past measured covariate history (X, \bar{L}_m) and treatment history \bar{A}_{m-1} . Randomization guarantees that

$$V(\cdot) \prod A_m | \bar{L}_m, \bar{A}_{m-1}, X \quad (1)$$

which, in turn, implies

$$Y(\cdot) \prod A_m | \bar{L}_m, \bar{A}_{m-1}, X. \quad (2)$$

We refer to (1) and (2), respectively, as the assumption of no unmeasured confounders for all outcomes $V(\cdot)$ and for outcome $Y(\cdot)$ only.

Suppose we are given assumption (2) plus (i) the consistency assumption $Y = Y(\bar{A})$ and (ii) the “positivity” assumption that, for each m , the conditional probability given (\bar{L}_m, X) of receiving each treatment a_m in the support of A_m is non-zero. Then the conditional mean $E[Y(\bar{a}) | x]$ is given by the g-computation algorithm formula (hereafter, g-formula)

$$b(\bar{a}, x) \equiv \int \cdots \int E[Y | \bar{\ell}_M, \bar{a}_M, x] \prod_{m=0}^M f(\ell_m | \bar{\ell}_{m-1}, \bar{a}_{m-1}, x) d\mu(\ell_m), \quad (3)$$

where μ is a dominating measure. Furthermore, the sharp null hypothesis

$$Y(\bar{a}) = Y(\bar{0}) \text{ w.p.1 for all } \bar{a} \in \bar{\mathcal{A}} \quad (4)$$

then implies the g-null mean hypothesis that $b(\bar{a}, x)$ is not a function of \bar{a} . It is a primary goal of an epidemiologist conducting an observational study to collect in L_m data on a sufficient number of covariates to try to make assumption (2) at least approximately true. However, since (2) cannot be guaranteed to hold even approximately and is not subject to empirical test, it is useful to investigate sensitivity to violations of (2) through a formal sensitivity analysis.

In order to choose an optimal treatment strategy, it is often necessary to consider dynamic treatment regimes. A dynamic regime is one in which a subject’s covariate history $\bar{\ell}_m$ through m determines the treatment a_m to be taken at m . Formally, a treatment regime g is a collection of $m + 1$ functions $g = (g_0, \dots, g_M)$ where $g_m : \bar{\mathcal{L}}_M \rightarrow \mathcal{A}_m$ maps the support $\bar{\mathcal{L}}_m$ of \bar{L}_m to the support \mathcal{A}_m of A_m . Let $V(g) = (Y(g), \bar{L}(g))$ be the counterfactual outcomes of a subject when following regime g . If, for each m , $g_m(\bar{\ell}_m)$ is a constant a_m not depending on $\bar{\ell}_m$, we say that regime g is non-dynamic and write $V(g)$ as $V(\bar{a})$. Under consistency and positivity assumptions, if (1) holds, $E(Y(g) | x)$ is given by the RHS of (3), with $\bar{a}_{m-1} \equiv g(\bar{\ell}_{m-1}) \equiv (g_0(\ell_0), \dots, g_{m-1}(\bar{\ell}_{m-1}))$ and $\bar{a}_M \equiv g(\bar{\ell}_M)$ being the treatment histories prescribed by regime g . If, as an example, small values of Y are healthier, the optimal regime g_{opt} (under squared error loss) minimizes $E[Y(g)]$. The regime g_{opt} will be dynamic when A_m is treatment with a prescription drug, since an optimal strategy must stop the drug when toxicity develops.

2. Reply to specific discussants

2a. Zeger-Liang (ZL)

Statistical models: We certainly agree with ZL that well-justified causal inferences draw on matters other than statistical analyses of epidemiologic data. But epidemiologic analyses are important and can be misleading unless confounding is controlled. As noted by Wasserman, the main message of our paper

was that standard methods of confounder control fail in the presence of time-varying confounders and that the g-formula provides correct adjustment. This is a completely model-free statement.

Causal models and extrapolations: We agree with ZL that tests of the causal null using the methods described in Secs. 1 through 6 require correct specification of parametric models. If, however, X , L_m , Y_m , and A_m are discrete with only a few levels, as in the MSCM data, we can obtain an asymptotic distribution-free test of the sharp null using the non-parametric g-null test described in Sec. 2d.3 below, completely alleviating ZL’s concerns.

We restricted attention to the contrasts $\theta_m^{(1)} - \theta_m^{(0)}$ for expositional purposes. We agree with ZL that other contrasts may be of interest. Most such contrasts can be expressed as $E[Y(g_1) - Y(g_2) | x]$ for some (possibly dynamic) regimes g_1 and g_2 . For example, ZL seem to be interested in a non-dynamic regime \bar{a} in which stress levels alternate, say on a daily or weekly basis. Methods for estimation of such contrasts are reviewed in Sec. 2d below. ZL also correctly point out that estimation of $\theta_m^{(1)}$ and $\theta_m^{(0)}$ requires extrapolation. This criticism is severely diminished if we take as our object of inference the parameter of a marginal structural model (MSM) for $E[Y(\bar{a}) | x]$, since the extrapolation involved in specifying a model for $E[Y(\bar{a}) | x]$ is no different from that involved in specifying a regression (GEE) model for $E[Y | X = x, \bar{A} = \bar{a}] \equiv E[Y(\bar{a}) | \bar{a}, x]$. The critical difference is that, because a MSM is a causal model, the parameter of a MSM is only identified under the untestable assumption of no unmeasured confounders.

GEE: Proposition (6) is a condition for the validity of GEEs as a method of inference. This proposition was originally stated and proved in Hu’s (1993) Sc.D. thesis and discussed in Fitzmaurice, Laird, and Rotnitzky (1993). It was also discovered by Pepe and Anderson (1994).

2b: Rubin-Frangakis (RF)

G-formula: Robins (1986, Secs. 1-3) first derived the g-formula (3) for the law of $Y(g)$ by mapping an observational study to a randomized trial missing data on the assigned treatment regimes g , without reference to counterfactual outcomes or to conditional independence statements like (1) and (2) above. His derivation based on counterfactuals came afterwards.

RF are incorrect in their assertion that assumption (1) of our paper [which is the analog of assumption (2) in this rejoinder] necessarily implies ignorability of the assignment mechanism in the sense of Rubin (1976, 1978). Specifically, Robins (1997, pp. 81-83) describes a substantively plausible data-generating mechanism which depends on further unobserved variables U_m , and for which assumption (2) of no unmeasured confounders for $Y(\cdot)$ is true but assumption (1) of no unmeasured confounders for all outcomes $V(\cdot)$ is false, once the U_m have been integrated out. If we specify a finite dimensional variation-independent parameterization for the underlying data-generating mechanism, then the treatment assignment mechanism is not ignorable. Furthermore, the identifiability results of Robins (1986, 1987) are not immediate consequences of results in Rosenbaum and Rubin (1983). For example, Robins (1986, 1987) shows that assumption (2) is sufficient for $E[Y(g) | x]$ to be given by the g-formula (3) for g non-dynamic but (1) is required when g is dynamic, and Robins (1987) shows that when (1) is false but (2) is true, the integral $b(\bar{a}, x)$ has a causal interpretation as $E[Y(\bar{a}) | x]$, yet none of the terms in the integrand on the RHS of (3) have any causal interpretation.

Time vs. clustering: RF appear to agree with the assumptions and analysis of our Sec. 7. However, they seem to prefer to add an intermediate step for clarity, which we eliminated for brevity. Specifically, RF’s C_i can be identified with the unobserved covariate mentioned in the third paragraph of this section (where we have used i rather u for units). Our covariate (random) effects are then $\sigma_{1i} = \beta_{01}C_i$ and $\sigma_{2i} = \beta_{02}C_i$ where β_{01} and β_{02} are as in Eqs. (25)-(26). RF remark that assumption (1) of our paper can be tested if we can assume treatment (stress) has no effect on a subset of the population. We do

not understand the relevance of this remark, since there was no subset about which we had such prior knowledge.

Potential confounders and sensitivity analysis: RF outline an approach to sensitivity analysis but specific details are lacking. In Sec. 2d below, we provide sensitivity analysis algorithms that are based on classes of non-parametric (just) identified (NPI) models introduced in Robins et al. (1999).

2c: Neuhaus: Neuhaus begins his discussion by asserting that the effect $E[Y(\bar{a}) - Y(\bar{0}) | x]$ of maternal stress history $\bar{a} = \bar{a}_M$ on the presence of childhood illness $Y = Y_{M+1}$ on day $M + 1$ can be estimated from cross-sectional data by propensity score matching. This assertion is incorrect, because (i), by (3), identifiability of $E[Y(\bar{a}) - Y(\bar{0}) | x]$, in general, requires longitudinal data $(\bar{L}_M, \bar{A}_M, Y)$, and (ii) even were such longitudinal data available, $E[Y(\bar{a}) - Y(\bar{0}) | x]$ cannot be estimated by propensity score matching methods. Neuhaus also incorrectly states that the parametric g-formula estimator used in our paper is an adaption of the methods of Robins (1992) based on g-estimation SNMs. On the contrary, SNMs were developed to overcome inadequacies of the parametric g-formula estimator.

In his second paragraph, Neuhaus is interested in the estimation of the counterfactual probability $pr[Y_m(\bar{a}) = 1 | Y_{m-1}(\bar{a}) = 0, \bar{Y}_{m-2}(\bar{a})]$ that a child becomes ill on day m given not ill on day $m - 1$ and past illness history $\bar{Y}_{m-2}(\bar{a})$ under regimes \bar{a} with various stress patterns, where $Y_m(\bar{a})$ is a time-dependent counterfactual outcome. Under assumption (1) of our paper, Neuhaus' interest can be easily accommodated either (i) by a parametric estimator of the g-functional for the joint law of $\bar{Y}_M(\bar{a})$ or, better yet, (ii) by an inverse-probability-of-treatment-weighted (IPTW) estimator of a discrete time marginal structural hazard model such as *logit* $pr[Y_m(\bar{a}) = 1 | Y_{m-1}(\bar{a}) = 0, \bar{Y}_{m-2}(\bar{a})] = \eta_{1m}^* + \eta_2^* a_m + \eta_3^* a_{m-1} + \eta_4^* a_{m-2} + \eta_5^* a_{m-1} a_{m-2} + \eta_6^* Y_{m-2}(\bar{a})$. If (1) of our paper only holds conditional on an unmeasured subject-specific covariate C_i , IPTW estimation can still be carried out under functional form and stationarity assumptions on $f[A_k | \bar{L}_k, \bar{A}_{k-1}, C, X]$, similar to those discussed in Sec. 7 of our paper.

2d: Wasserman:

2d.1: The null paradox: In our paper, we estimated $b(\bar{a}, x)$ of (3) by a parametric g-formula estimator in which we substituted parametric models with variation-independent parameters for the density and expectations in (4), estimated these parameters by maximum likelihood, and plugged the estimates into (4). However, Robins (1986, 1997) and Robins and Wasserman (1997) show that this approach leads to the following null paradox. Suppose, (i) the sharp null (4) holds, (ii) treatment affects the L_m (i.e., (4) is false with \bar{L} replacing Y), and (iii), as will nearly always be the case, pre-treatment unmeasured health status, say U_0 , is a causal determinant of both L_m and Y . Then, with probability going to 1, the parametric g-functional estimator of $b(\bar{a}, x)$ will depend on \bar{a} , and thus the null hypothesis (4) will be falsely rejected for most choices of the parametric models. The explanation for this paradox is that if (i)-(iii) hold, then, under the true data-generating mechanism, both integrands $f(\ell_m | \bar{\ell}_{m-1}, \bar{a}_{m-1}, x)$ and $E[Y | \bar{\ell}_M, \bar{a}_M, x]$ will depend on the regime \bar{a} but the integral $b(\bar{a}, x)$ will not. Such delicate cancellation of the dependence on \bar{a} is not possible for most parametric models; that is, for most models there is no combination of parameter values that makes $b(\bar{a}, x)$ independent of \bar{a} . Below we consider methods that avoid the null paradox.

2d.2: Marginal structural models and IPTW semiparametric estimators: A marginal structural model (MSM) for $E[Y(\bar{a}) | x]$ specifies

$$E[Y(\bar{a}) | x] = d(\bar{a}, x; \eta^*) \tag{5}$$

where $d(\bar{a}, x; \eta)$ is a known function, η^* is an unknown parameter vector, and $\eta = (\eta'_1, \eta'_2)'$ is coded so that $\eta_2 = 0$ if and only if $d(\bar{a}, x, \eta)$ does not depend on \bar{a} . The sharp null (4) implies $\eta_2^* = 0$, thus avoiding the null paradox. We call (5) a MSM because it causally models the marginal distribution of the $Y(\bar{a})$ (rather than, say, the correlation of $Y(\bar{a})$ and $Y(\bar{a}^*)$ for $\bar{a} \neq \bar{a}^*$). In contrast, a regression association (model) such as a GEE model specifies a parametric form for $E[Y | \bar{A} = \bar{a}, X = x] = E[Y(\bar{a}) | \bar{a}, x]$. Note that $E[Y(\bar{a}) | \bar{a}, x]$ may depend on \bar{a} even if (4) is true, i.e., association is not causation.

When the treatment process is not ancillary, a random measure of statistical non-ancillarity is $W = \prod_{m=0}^M f[A_m | \bar{A}_{m-1}, \bar{L}_m, X] / \prod_{m=0}^M f[A_m | \bar{A}_{m-1}, X]$. Suppose W is known and consider the IPTW estimator $\hat{\eta}$ solving $0 = \sum_{i=1}^n U_i(W_i, \eta)$ where i indexes the n study subjects, $U(W, \eta) \equiv W^{-1} [\partial d(\bar{A}, X; \eta) / \partial \eta] \hat{\Sigma}^{-1}(\bar{A}, X) [Y - d(A, X; \eta)]$, and $\hat{\Sigma}(\bar{A}, X)$ is an estimated working covariance matrix. The IPTW estimator $\hat{\eta}$ generalizes ideas in Kalbfleisch and Lawless (1988) and Flanders and Greenland (1991). The consistency of $\hat{\eta}$ for η^* of (5) under (2) is a corollary of the following lemma.

Lemma 1: The g-formula $b(\bar{a}, x)$ is the unique function satisfying $E[q(\bar{A}, X) [Y - b(\bar{A}, X)] / W] = 0$ for all functions $q(\bar{A}, X)$ for which the expectation exists.

The $\tilde{\eta}$ solving $\sum_{i=1}^n U_i(1, \eta)$ is the usual GEE estimator. It will generally be inconsistent for η^* unless $W = 1$ w.p.1. Lemma 1 implies that, when W is replaced by an estimate \widehat{W} , $\hat{\eta}$ remains consistent if the assignment mechanism $f(A_m | \bar{A}_{m-1}, \bar{L}_m, X)$ is either known (as in a sequential RCT) or a correct model $f(A_m | \bar{A}_{m-1}, \bar{L}_m, X; \phi)$ is specified and the parameter ϕ is consistently estimated. Misspecification of the model for $f(A_m | \bar{A}_{m-1}, X)$ does not affect consistency (Robins, 1999).

2d.3.: Sensitivity analysis for MSMs: Suppose the $Y(\bar{a})$ are non-negative. The assumption (2) of no unmeasured confounders implies that the contrast $\ln E[Y(\bar{a}) | x, \bar{\ell}_m, \bar{a}_{m-1}, a_m] - \ln E[Y(\bar{a}) | x, \bar{\ell}_m, \bar{a}_{m-1}, a_m^\dagger]$ between the logarithm of the conditional counterfactual mean among subjects with $A_m = a_m$ and that among subjects with $A_m = a_m^\dagger$ takes the value zero for all $\bar{a}, x, \bar{\ell}_m, a_m^\dagger$. Hence a natural approach to modelling the magnitude of confounding due to unmeasured factors is to specify a model $q(x, \bar{\ell}_m, \bar{a}, \alpha_m^\dagger; \alpha^*)$ for the above contrast, where α^* is an unknown parameter and $q(x, \bar{\ell}_m, \bar{a}, \alpha_m^\dagger; \alpha)$ is a known function [such as $\alpha(a_m - a_m^\dagger)$] that is zero if $a_m^\dagger = a_m$ or $\alpha = 0$. The magnitude and sign of α^* indicates the degree and direction of confounding due to unmeasured factors.

Robins et al. (1999) prove that α^* is non-parametrically unidentified. Hence, rather than trying to estimate α^* from the data, we vary α in a sensitivity analysis as follows. Define the α -corrected version $Y^*(\alpha)$ of Y to be $Y^*(\alpha) = Y \prod_{m=0}^M Q_m(\alpha)$ where $Q_m(\alpha) = \exp[-q(X, \bar{L}_m, \bar{A}, 1 - A_m; \alpha)] \{1 - f(A_m | \bar{L}_m, \bar{A}_{m-1}, X)\}$ when A_m is a dichotomous (0, 1) variable, and $Q_m(\alpha) \equiv \int \exp\{-q(X, L_m, \bar{A}, a_m; \alpha)\} dF(a_m | \bar{L}_m, \bar{A}_{m-1}, X)$ more generally. Robins et al. (1999) show that the IPTW estimator $\hat{\eta} = \hat{\eta}(\alpha)$, computed by replacing Y by $Y^*(\alpha)$ in $U(W, \eta)$, remains consistent for the parameter η^* of model (5) when $\alpha = \alpha^*$, provided that our model $f(A_k | \bar{L}_k, \bar{A}_{k-1}, X; \phi)$ is correctly specified. Final conclusions depend upon the values of α^* considered plausible by relevant experts. Because the functional form of q is not identified, several sensitivity analyses using different functional forms should be reported. The consistency of $\hat{\eta}(\alpha^*)$ follows from the fact that Lemma 1 remains true when we replace the g-formula $b(\bar{a}, x)$ by $c(\bar{a}, x) \equiv E[Y(\bar{a}) | x]$ and Y by $Y^*(\alpha^*)$. When the mean of Y may assume any positive value (e.g., Y is Poisson), any value of α may be chosen. However, when Y is Bernoulli, the magnitude of α is somewhat limited by the requirement that probabilities not exceed one.

2d.4.: Structural nested models: Given \bar{a} , let $(\bar{a}_m, 0)$ be the treatment history that agrees with

\bar{a} through m and is 0 thereafter. A structural nested mean model (SNMM) for the effect of a final blip of treatment of magnitude a_m among subjects with past history $(\bar{\ell}_m, \bar{a}_m, x)$ specifies that

$$\Phi \{E [Y (\bar{a}_m, 0) | \bar{\ell}_m, \bar{a}_m, x]\} = \Phi \{E [Y (\bar{a}_{m-1}, 0) | \bar{\ell}_m, \bar{a}_m, x]\} + \gamma (\bar{\ell}_m, \bar{a}_m, x, \psi^*)$$

where $\Phi(u) = u$ for an additive SNMM and $\Phi(u) = \ln u$ for a multiplicative SNMM, $\gamma(\bar{\ell}_m, \bar{a}_m, x, \psi)$ is a known function satisfying $\gamma(\bar{\ell}_m, \bar{a}_m, x, \psi) = 0$ if $a_m = 0$ or $\psi = 0$, and ψ^* is an unknown parameter. The sharp null (4) implies $\psi^* = 0$, thus avoiding the null paradox. In fact, under assumption (1), $\psi^* = 0$ is equivalent to the g-null mean hypothesis $E(Y(g) | x) = E(Y(\bar{0}) | x)$ for all g . To estimate ψ^* , define $\Phi \{H_m(\psi)\} = \Phi(Y) - \sum_{k=m}^M \gamma(\bar{L}_k, \bar{A}_k, X; \psi)$ and let $t_m(\bar{A}_m, \bar{L}_m, X)$ be user-supplied functions of the dimension of ψ . Then, under (2), Robins (1994) proves that $E[H_m(\psi^*) | \bar{L}_m, \bar{A}_m, X] = E[H_m(\psi^*) | \bar{L}_m, \bar{A}_{m-1}, X]$, which implies that the ‘‘g-estimator’’ $\hat{\psi}$ solving $\sum_i U_i(\psi) = 0$ will be consistent for ψ^* where $U(\psi) = \sum_{m=0}^M H_m(\psi) \{t_m(\bar{A}_m, \bar{L}_m, X) - \int t_m(\bar{A}_m, \bar{L}_m, X) dF[A_m | \bar{L}_m, \bar{A}_{m-1}, X]\}$. The estimator $\hat{\psi}$ generalizes estimators of Rosenbaum (1984) and Robins, Mark, and Newey (1992) to time-dependent treatments.

The test of $\psi^* = 0$ [and thus of the sharp null (4)] based on $\sum_i U_i(0)$ and its estimated variance is referred to as a g-null test (Robins, 1986, 1997). In an observational study, if (X, A_m, L_m) are all discrete with but a few levels, the nonparametric g-null test that replaces $dF(A_m | \bar{L}_m, \bar{A}_{m-1}, X)$ by its non-parametric (empirical) estimate is an asymptotically distribution free (ADF) test of the sharp null, even if M is large or the ‘‘positivity’’ assumption is false. Specifying a parsimonious parametric model $f(A_m | \bar{L}_m, \bar{A}_{m-1}, X; \phi)$ will increase power, but sacrifice the ADF property. G-null tests are the optimal solution to the null paradox. In contrast, a test of $\eta^* = 0$ of MSM (5) based on $\sum_i U_i(\widehat{W}, 0)$ and an estimate of its variance can never be ADF when M is large or the positivity assumption is false. Finally, a Monte-Carlo estimate of the treatment effect $E[Y(g) - Y(\bar{0}) | x]$ based on an additive SNMM is $J^{-1} \sum_{j=1}^J \sum_{m=0}^M \gamma(\bar{\ell}_{m,j}, \bar{a}_{m,j}, x; \hat{\psi})$, where the $\ell_{m,j}$ are drawn recursively from an estimate of $f(\ell_m | \bar{\ell}_{(m-1),j}, \bar{a}_{(m-1),j}, x)$ and $\bar{a}_{m,j} = g(\bar{\ell}_{m,j}, x)$. Estimation of $E[Y(g) - Y(\bar{0}) | x]$ based on multiplicative SNMMs or for MSMs with g dynamic is considered in Robins (1997, 1999) and Robins et al. (1999).

2d.5: Sensitivity analysis for SNMMs: For SNMMs, there are two distinct sensitivity analysis methodologies. The first is closely analogous to the MSM methodology described above.

Method 1. We specify a model $q_m(\bar{\ell}_m, \bar{a}_m, x; \alpha^*)$ for the contrast $\Phi \{E [Y (\bar{a}_{m-1}, 0) | \bar{\ell}_m, \bar{a}_m, x]\} - \Phi \{E [Y (\bar{a}_{m-1}, 0) | \bar{\ell}_m, \bar{a}_{m-1}, a_m = 0, x]\}$ between the (Φ -transformed) conditional counterfactual mean among subjects with $A_m = a_m$ and that among subjects with $A_m = 0$ with a common past $(x, \bar{\ell}_m, \bar{a}_{m-1})$. Here, α^* is a parameter and $q_m(\bar{\ell}_m, \bar{a}_m, x; \alpha)$ is a known function that is 0 if $\alpha = 0$ or $a_m = 0$. The parameter α^* is non-parametrically unidentified and encodes the degree and direction of confounding due to unmeasured factors. As before, we vary α in a sensitivity analysis. Robins et al. (1999) show that $\hat{\psi} = \hat{\psi}(\alpha)$ remains consistent for ψ^* when $\alpha = \alpha^*$ if we replace $H_m(\psi)$ by $H_m^*(\psi, \alpha) = \Phi^{-1} \{ \Phi [H_m(\psi)] - q(\bar{L}_m, \bar{A}_m, X; \alpha) \}$ in $U(\psi)$.

Method 2 - Instrumental variables: Suppose $A_m = (A_{pm}, A_{dm})$ where A_{pm} is the prescribed dose of a medicine and A_{dm} is the actual consumed dose. Since we often have good measures of why doctors prescribe a given dose but poor measures of why patients comply, it would often be reasonable to assume (2) was false but the weaker assumption $Y(\cdot) \perp\!\!\!\perp A_{pm} | \bar{L}_m, \bar{A}_{m-1}$ was true. In that case, we cannot use MSMs or the g-formula to identify the effect of the joint treatment \bar{A} on Y . However, our

g-estimate $\hat{\psi}$ remains consistent, provided we use functions $t_m(\bar{A}_m, \bar{L}_m, X) = t_m(A_{pm}, \bar{A}_{m-1}, \bar{L}_m)$ that only depend on A_m through A_{pm} . If $\gamma(\bar{\ell}_m, \bar{a}_m, x, \psi^*)$ is a function of a_m only through actual dose a_{dm} , we say the A_{pm} are “instrumental variables” for the A_{dm} (Robins, 1992).

Additional References

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91:444-455.

Copas, J.B. (1973). Randomization models for matched and unmatched 2×2 tables. *Biometrika*, 60:467-476.

Fitzmaurice, G.M., Laird, N., and Rotnitzky, A.G. (1993). Reply to comments on “Regression models for discrete longitudinal responses”. *Statistical Science*, 8:306-309.

Flanders, W.D., and Greenland, S. (1991). Analytic Methods for Two Stage Case-control Studies and Other Stratified Designs. *Statistics in Medicine*, 10:739-747.

Kalbfleisch, J.D., and Lawless, J.F. (1988). Likelihood Analysis of Multi-stage Models for Disease Incidence and Mortality. *Statistics in Medicine*, 7:149-160.

Robins, J.M. (1988). Confidence intervals for causal parameters. *Statistics in Medicine*, 7:773-785.

Robins, J.M., Mark SD, and Newey WK. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48:479-495.

Robins, J.M. and Wasserman L. (1997). Estimation of Effects of Sequential Treatments by Reparameterizing Directed Acyclic Graphs. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Providence, Rhode Island, August 1-3, 1997. Dan Geiger and Prakash Shenoy (Eds.), Morgan Kaufmann, San Francisco, pp. 409-420.

Rosenbaum, P.R. (1995). **Observational Studies**. New York: Springer-Verlag.

Rubin, D.B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47:1213-1234.