In: **Computation, Causation, and Discovery**. Eds. C. Glymour and G. Cooper.
Menlo Park, CA, Cambridge, MA: AAAI Press/The MIT Press, 1999, pp. 349-405.

TESTING AND ESTIMATION OF DIRECT EFFECTS BY REPARAMETERIZING DIRECTED ACYCLIC GRAPHS WITH STRUCTURAL NESTED MODELS

James M. Robins

Departments of Epidemiology and Biostatistics, Harvard School of Public Health,
677 Huntington Avenue, Boston, MA 02115
email: robins@hsph.harvard.edu

Abstract: The standard way to parameterize the distributions represented by a directed acyclic graph is to insert a parametric family for the conditional distribution of each random variable given its parents. We show that when one's goal is to test for or estimate the direct effect of a treatment, this natural parameterization has serious deficiencies. Furthermore, in most settings, the no direct-effect null hypothesis does not entail any conditional independence restrictions and thus cannot be tested using computer programs designed to search for conditional independencies in the data. By reparameterizing the graph using direct effect structural nested models, these problems can be overcome. A direct-effect structural nested model is a causal model for the direct effect of a final brief blip of treatment on the outcome of interest.

1. Introduction

Consider a set of random variables $V = (X_1, \ldots, X_M)$ whose joint density $f(v)$ is represented by a Directed Acyclic Graph (DAG) G. If Pa_m represents the parents of X_m, then the density factorizes as

$$f(v) = \prod_{m=1}^{M} f(x_m|pa_m) \ . \tag{1.1}$$

In practice, in order to estimate $f(v)$ from independent realizations $V_i, i = 1, \ldots, n$, obtained on n study subjects, one often needs to assume some particular parametric form for each $f(x_m|pa_m)$. Thus one writes $f(v) = \prod\limits_{n=1}^{M} f(x_m|pa_m; \theta_m)$ where θ_m is an element of the parameter space $\boldsymbol{\theta}_m$, an open subset of a finite dimensional Euclidean space. For example, suppose the parent of X_2 is X_1. Then $p(x_2|pa_2; \theta_2)$ might be $N(\beta_0 + \beta_1 x_1, \sigma^2)$ so that $\theta_2 = (\beta_0, \beta_1, \sigma)$. In general, if one inserts a parametric family $f(x_m \mid pa_m; \theta_m)$ into the right hand side of each term of (1.1) and the θ_m are variation independent, we call this a standard parameterization of the DAG. This seems to be the usual way of using DAGS in practice. The parameters θ_m are said to be variation independent if the joint parameter space for $\theta = (\theta'_1, \ldots, \theta'_M)'$ is the product space $\boldsymbol{\theta}_1 \times \boldsymbol{\theta}_2 \times \ldots \times \boldsymbol{\theta}_M$.

As natural as it seems to parameterize a DAG in this way, there are problems with the standard parameterization when one's goal is to test for or estimate the direct effect of a treatment or control variable. This has been noted by Robins (1989, 1997). The next section gives a simple example which illustrates the problem. In our simple example, the goal is to test for and estimate the direct effect of an AIDS drug, prophylaxis therapy (A_P), on serum HIV RNA levels (Y) among AIDS patients, many of whom also receive a second active AIDS therapy, AZT (A_Z).

Briefly put, the problem is this: Suppose the DAG G represents treatments and covariates in a longitudinal study. Further suppose that the partial ordering of the variables in V entailed by the DAG G is consistent with the temporal ordering of the variables. Our goal is to test for a direct effect of A_P treatment on the outcome Y when A_Z is held fixed. In certain settings, such as a sequential randomized trial, the null hypothesis of "no direct treatment effect," although identifiable based on the observed data, cannot be tested simply by testing for the presence or absence of arrows in the DAG G as one might expect. These conditions, far from being pathological, are indeed likely to hold in most real examples. Fortunately, the direct effect null hypothesis can be tested by examining a particular integral called the "G-computation algorithm functional". The null is true if this integral satisfies a certain complex condition. Indeed, this complex condition does not entail any conditional independence restrictions on the joint distribution of the observed variables and thus cannot be tested using programs such as TETRAD [Spirtes, Glymour, and Scheines (SGS), 1993] that test for conditional independencies (Robins, 1986; Verma and Pearl, 1991; SGS, 1993). Furthermore, we prove in Theorem 2 that there is an additional complication. Specifically, common choices for the parametric families in a standard parameterization often lead to joint densities such that the integral can never satisfy the required condition; as a consequence, in large samples, the null hypothesis of no direct treatment effect, even when true, will be falsely rejected regardless of the data. These problems are exacerbated in high dimensional problems. In summary, currently available methods cannot be used to test for direct treatment effects.

We propose a new class of tests, the direct-effect g-tests, of the hypothesis of no direct effect of a (possibly time-varying sequential) treatment A_P on an outcome Y when a second treatment A_Z is held fixed. A direct-effect g-test reweights the g-null test (Robins, 1986) of conditional independence of A_P and Y by the inverse of the probability of receiving treatment A_Z. When the direct-effect g-test rejects, it is of interest to estimate the magnitude of the effect of A_P on Y when treatment A_Z is held fixed at a specified value a_Z. In order to do so, we introduce the class of direct-effect structural nested models (SNMs). A direct-effect SNM models the effect of a final brief blip of treatment with A_P at time t on Y when A_Z is set to a_Z. Direct-effect SNMs offer a unified approach to testing for and estimation of direct effects. This unified approach utilizes a class of robust semiparametric tests and estimators, the direct-effect g-tests and estimators, for the parameter ψ of a direct-effect SNM. These tests and estimators are semiparametric in the sense that they only require that one specify a parametric model for the probability of treatment with A_P and A_Z given their parents on the DAG.

In contrast to this approach, in the graphical modelling literature, the usual approach to estimation is to *(i)* specify a fully parametric model for the distribution of the DAG and *(ii)* then estimate parameters of interest either by maximum likelihood or Bayesian methods. In this spirit, we describe a complete reparameterization of the distribution of the DAG with a direct-effect structural nested model for the effect of A_P on Y with A_Z set to a_Z as one component. A second component of the reparameterization is a marginal structural model (Robins, 1998, 1999) for the effect of A_Z on Y when A_P is withheld. This reparameterization allows a fully parametric likelihood or Bayesian approach to testing for and estimation of direct effects. We compare the strengths and weaknesses of our semiparametric approach with this fully parametric approach. We ultimately recommend a "mixed" approach that combines the best aspects of both.

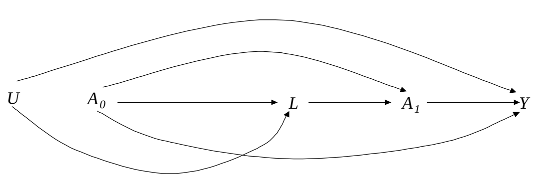
We study two different direct-effect structural nested models, direct-effect structural nested mean models (SNMMs) and direct-effect structural nested distribution models (SNDMs). The SNMMs model the direct effect of a treatment A_P on the mean of the outcome Y while the SNDMs model the direct effect of A_P on the entire distribution of Y. SNDMs are only appropriate if the outcome Y is continuous, while SNMMs can be used to analyze both discrete and continuous responses. Our direct-effect SNMs differ from the standard SNMs discussed in Robins (1989, 1993, 1995, 1997) and Robins and Wasserman (1997). We have introduced direct-effect SNMs because, as shown by Robins and Wasserman (1997), the standard SNDMs are not adequate to test for and/or estimate direct effects except in special cases. Specifically, in Section 12, we show that the standard SNDMs are adequate to test for and estimate the direct effect of A_P when A_Z is held fixed only if the magnitude of the effect of A_Z on Y is not modified by pre-treatment covariates.

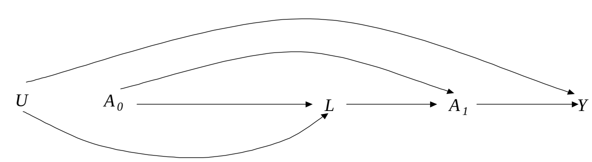
Finally, we note that the direct-effect SNMs introduced in this paper differ from the direct-effect SNMs introduced in Appendix 3 of Robins (1997). We recommend that an analyst use the direct-effect SNMs of this paper rather than those of Robins (1997), because the direct-effect SNMs introduced here, in contrast to those of Robins (1997), admit robust semiparametric direct-effect g-tests and estimates. However, the direct-effect SNMs introduced here are intrinsically asymmetric, in that, in contrast to the models in Robins (1997), a single model cannot be used to simultaneously test for a direct effect of A_P fixing A_Z and for a direct effect of A_Z fixing A_P.

1.1. An Example

To illustrate the problem we are concerned with, consider the following generic example of a sequential randomized clinical trial depicted by DAG 1a in which data have been collected on variables (A_0, A_1, L, Y) on each of n AIDS patients. The continuous variable A_0 represents the dose in milligrams of prophylaxis therapy for PCP (AIDS pneumonia) received by AIDS patients at time t_0; the continuous variable A_1 represents the dose in milligrams of AZT received at time t_1; the dichotomous variable L records whether a patient developed PCP in the interval from t_0 to t_1; the continuous variable Y represents a subject's HIV-viral load measured at end-of-follow-up; and the hidden (unmeasured) variable U denotes a patient's underlying immune function at the beginning of the study. U is a measure of a patient's underlying health status.

The prophylaxis therapy dose A_0 was assigned at random to subjects at time t_0 so, by design, $A_0 \perp\!\!\!\perp U$. AZT treatment A_1 was randomly assigned at time t_1 with randomization probabilities that depend on the observed past (A_0, L), so, by design $U \perp\!\!\!\perp A_1 \mid A_0, L$. For simplicity, we shall assume that no other unmeasured common causes (confounders) exist. That is, each arrow in DAG 1a represents the direct causal effect of a parent on its child, as in Pearl and Verma (1991) or SGS (1993). Note DAG 1a is not complete because of three missing arrows: the arrows from U to A_0 and A_1 and the arrow from L to Y. The arrows from U are missing by design. The missing arrow from L to Y represents a priori biological knowledge that L has no effect on HIV viral load Y. (The missing arrow from L to Y is not essential to what follows and is assumed to simplify the exposition.) Hence, by the Markov properties of a DAG, we know that $L \coprod Y \mid A_0, A_1, U$. It is known that A_0 causes PCP, so $A_0 \not\!\perp\!\!\!\perp L$. Also it is known that the unmeasured variable U has a direct effect on L and Y, i.e., U causes both PCP and an elevated HIV RNA. Finally, we suppose it is known that AZT has a direct effect on the outcome Y. That is, there is an arrow from A_1 to Y and so $A_1 \not\!\perp\!\!\!\perp Y \mid L, A_0, U$.





DAG 1a

DAG 1b

1.2. Representing the Direct Effect Null Hypothesis

Suppose the trial data has been collected in order to test the null hypothesis that prophylaxis therapy A_0 has no direct effect on viral load Y. This "no direct effect null" hypothesis is the hypothesis that the arrow from A_0 to Y in DAG 1a is missing, which would imply that the true causal graph generating the data was DAG 1b. The alternative to this null hypothesis is that the true causal graph generating the data is graph 1a.

Following Pearl and Verma (1991) and SGS (1993), we assume the joint distribution of $W = (V, U)$ is faithful to the true graph where $V = (A_0, A_1, L, Y)$. That is, if B, C, and D are distinct (possibly empty) subsets of the variables in (V, U), then B is independent of C given D if and only if B is d-separated from C given D on the true causal graph generating the data. It follows that the "no direct effect null hypothesis" of DAG 1b is true if and only if $A_0 \coprod Y \mid A_1, L, U$. Indeed, since we have assumed no arrows from L to Y, $A_0 \coprod Y \mid A_1, L, U$ *is* equivalent to the hypothesis $A_0 \coprod Y \mid A_1, U$. The question is: can we still characterize the null hypothesis even of U is not observed. The answer is yes, according to the following Theorem.

Theorem 1.1: Suppose the distribution of $W = (V, U)$ is faithful to either DAG 1a or 1b, one of which generated the data. Then, the direct effect null hypothesis $A_0 \perp\!\!\!\perp Y \mid A_1, U$ holds (i.e., DAG 1b generated the data) if and only if

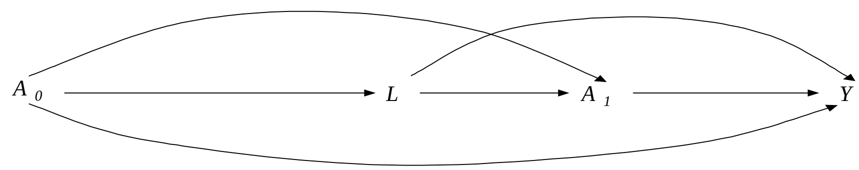
$$\sum_{\ell=0}^{1} f(y \mid a_0, a_1, \ell) \, f(\ell \mid a_0) \text{ does not depend on } a_0. \tag{1.2}$$

Thus, even though U is unobserved, we can still tell if the null holds by checking (1.2) which only involve the observables.

Remark: Even without imposing faithfulness, $A_0 \coprod Y \mid A_1, U$ implies (1.2), although the converse is no longer true. Note $\sum_{\ell=0}^{1} f(y \mid a_0, a_1, \ell) \, f(\ell \mid a_0)$ is the marginal density of Y in the manipulated subgraph (SGS, 1993) of the complete DAG (2) for the observed data $V = (A_0, A_1, L, Y)$ *in* which arrows into A_0 and A_1 have been removed and (A_0, A_1) set to (a_0, a_1). Robins (1986) refers to this marginal density as the g-computational algorithm formula for the effect of (A_0, A_1) on Y. Because U has no arrows into A_0 or A_1, Theorem 1.1 is a special case of Theorem F.3 of Robins (1986). See also Pearl and Robins (1995).

By the d-separation criterion applied to DAG 1a and 1b, we see that if either of DAGs 1a or 1b generated the data, then the joint distribution of V is represented by the complete DAG 2 without missing arrows. The additional restriction (1.2) that distinguishes the no direct effect hypothesis of DAG 1b from DAG 1a is not representable by removing arrows from DAG 2. This is an important observation because a common way of testing the direct effect null hypothesis is to test for the absence of an arrow from A_0 to Y and to test $A_0 \coprod Y \mid L, A_1$; we call this the "naive test." This test is incorrect. Specifically, if the no direct effect hypothesis of DAG 1b is correct and the distribution of W is faithful, then $A_0 \coprod Y \mid L, A_1$ will be false, and the naive test will falsely reject the no direct effect null with probability converging to one in large samples.

Thus, testing the null hypothesis of no direct effect of prophylaxis therapy A_0 cannot be accomplished by testing for the presence or absence of arrows on DAG 2. This is because the arrows of the marginal DAG 2 do not have a causal interpretation (even though the arrows on the underlying causal DAG do have a causal interpretation). Indeed, by applying the d-separation criteria to DAG 1b, we discover that DAG 1b does not entail any conditional independence restrictions among the observed variables V. Thus, the direct effect null hypothesis of DAG 1b cannot be distinguished from the alternative DAG 1a by a search program such as TETRAD that uses conditional independence relations among the observed variables V to distinguish between underlying causal graphs (Robins, 1986; Verma and Pearl, 1991; SGS, 1993, page 193). One solution to this problem is to test (1.2) directly. With standard parameterizations, this approach will also fail, as the next section shows. In Sec. 2, we will derive an appropriate alternative test of (1.2), the direct-effect g-test.



DAG 2

1.3. The Problem With Standard Parameterizations

We saw that to test the null hypothesis, it does not suffice to test whether the arrow in DAG 2 from A_0 to Y is broken. Rather, we need to test the Eq. (1.2). We now show that such a test will falsely reject if one uses a standard parameterization. To test the null hypothesis (1.2), the standard approach is to first specify parametric models for the conditional distribution of each parent given its children in the complete DAG 2 representing the observed data. Hence let $\{f(y \mid a_0, a_1, \ell; \theta); \theta \in \Theta \subset R^q\}$ and $\{f(\ell \mid a_0; \gamma); \gamma \in \Gamma \subset R^p\}$ denote parametric models for the unknown densities $f(y \mid a_0, a_1, \ell)$ and $f(\ell \mid a_0)$. Of course, we cannot guarantee these models are correctly specified. We say the model $f(y \mid a_0, a_1, \ell; \theta)$ is correctly specified if there exists $\theta_0 \in \Theta$ such that $f(y \mid a_0, a_1, \ell; \theta_0)$ is equal to the true (but unknown) density $f(y \mid a_0, a_1, \ell)$ generating the data. Results in this Section require the concept of linear faithfulness. We say that the distribution of W is linearly faithful to the true causal graph in generating the data, if for any disjoint (possibly empty) subsets B, C, and D of the variables in B, B is d-separated from C given D on the graph if and only if the partial correlation matrix $r_{BC.D}$ between B and C given D is the zero matrix. If W is jointly normal, linear faithfulness and faithfulness are equivalent. For W non-normal, neither implies the other. However, the argument that the distribution of V should be linearly faithful to the generating causal DAG is essentially identical to the argument that the distribution should be faithful to the causal DAG given by SGS (1993) and Pearl and Verma (1991).

To see why standard parameterizations may not work, consider a specific example. Recall that Y is continuous and that L is binary. Commonly used models in these cases are normal linear regression models and logistic regression models. Thus suppose that we adopt the following models:

$$Y | a_0, a_1, \ell; \theta, \sigma^2 \sim N(\theta_0 + \theta_1 a_0 + \theta_2 \ell + \theta_3 a_1, \sigma^2) \tag{1.3}$$

and

$$f(\ell = 1|a_0;\gamma) = expit(\gamma_0 + \gamma_1 a_0) \tag{1.4}$$

where $expit(b) = e^b/(1+e^b)$ and $N(\mu, \sigma^2)$ denotes a Normal distribution with mean μ and variance σ^2. We will now prove the following results.

Lemma 1.1: If the no direct effect null hypothesis represented by DAG 1b is true and the distribution of (V, U) is either faithful or linearly faithful to DAG 1b, then model (1.3) and/or model (1.4) is guaranteed to be misspecified; that is, the set of distributions $\mathcal{F}_{par}$ for V satisfying (1.3)-(1.4) is disjoint from the set $\mathcal{F}_{mar}$ of distributions for (V, U) that are marginals of distributions for W that are either faithful or linearly faithful to DAG 1b.

Since model (1.3) and/or (1.4) are guaranteed to be misspecified under the no direct effect null hypothesis, one might expect that tests of the null assuming (1.3)-(1.4) will perform poorly. This expectation is borne out by the following theorem.

Theorem 1.2: Suppose *(i)* the data analyst tests the no direct effect null hypothesis using the parametric models (1.3)-(1.4) fit by the method of maximum likelihood, *(ii)* the no direct effect null hypothesis represented by DAG 1b is true, *(iii)* the distribution of (V, U) is linearly faithful to DAG 1b. Then, with probability converging to 1, the no direct effect null hypothesis (1.2) will be falsely rejected.

Theorem 1.2 implies that if we use models (1.3)-(1.4), then in large samples, we will reject the no direct effect null hypothesis, even when true, for nearly all data sets (i.e., with probability approaching 1). That is, by specifying models (1.3)-(1.4), we will have essentially rejected the no direct effect null hypothesis, when true, even before seeing the data!

Proof of Theorem 1.2 and Lemma 1.1: The following Proof of Theorem 2 also proves Lemma 1. Note Eqs. (1.2) implies that

$$b(a_0, a_1) \equiv \sum_{\ell=0}^{1} E[Y \mid \ell, a_0, a_1] f(\ell \mid a_0) \tag{1.5}$$

does not depend on a_0. Now, under model (1.3)-(1.4), the maximum likelihood estimator of $b(a_0, a_1)$ is $b\left(a_0, a_1; \widehat{\theta}, \widehat{\gamma}\right) = \widehat{\theta}_0 + \widehat{\theta}_1 a_0 + \widehat{\theta}_3 a_1 + \left\{\widehat{\theta}_2 e^{\widehat{\gamma}_0 + \widehat{\gamma}_1 a_0}\right\} / \left\{1 + e^{\widehat{\gamma}_0 + \widehat{\gamma}_1 a_0}\right\}$ where the maximum likelihood estimators $\widehat{\theta}, \widehat{\gamma}$ satisfy the normal and logistic score equations $\sum_{i=1}^{n}\left(Y_i - \widehat{\theta} Z_i\right) Z_i = 0$ and $0 = \sum_{i=1}^{n}\left\{L_i - expit\left(\widehat{\gamma}_0 + \widehat{\gamma}_1 A_{0i}\right)\right\}(1 A_{0i})'$ where $Z_i = (1, A_{0i}, L_i, A_{1i})'$, $\theta' = (\theta_0, \theta_1, \theta_2, \theta_3)$, and $\gamma' = (\gamma_0, \gamma_1)$. Further, the probability limits θ^* and γ^* of $\widehat{\theta}$ and $\widehat{\gamma}$ satisfy $E\left[\left\{Y_i - \theta^* Z_i\right\} Z_i\right] = 0$ and $E\left[\left\{L_i - expit\left(\gamma_0^* + \gamma_1^* A_{0i}\right)\right\}(1, A_{0i})'\right] = 0$, where the expectations are with respect to the true distribution generating the data regardless of whether models (1.3)-(1.4) are correctly specified. The MLE $b\left(a_0, a_1; \widehat{\theta}, \widehat{\gamma}\right)$ converges in probability to $b(a_0, a_1; \theta^*, \gamma^*)$. It follows that an analyst using models (1.3)-(1.4) fit by maximum likelihood will reject (1.2) with probability approaching 1 as $n \to \infty$ if $b(a_0, a_1; \theta^*, \gamma^*)$ depends on a_0. We now prove such a dependence by contradiction.

It is clear that $b(a_0, a_1; \theta^*, \gamma^*)$ does not depend on a_0 if and only if either *(i)* $\theta_1^* = \theta_2^* = 0$, or *(ii)* $\theta_1^* = \gamma_1^* = 0$. However, it follows from standard least squares theory that *(i)* is true if and only if the partial correlations between Y and A_0 and between Y and L given A_1 are both zero. But this contradicts the assumption that the distribution of (V, U) is linearly faithful to DAG 1b since both Y and L and Y and A_0 are not d-separated conditional on A_1. Similarly, if *(ii)* is true, then $\gamma_1^* = 0$. But an easy calculation shows that $\gamma_1^* = 0$ if and only if $cov(L, A_0) = 0$. However, $cov(L, A_0) = 0$ contradicts the linear faithfulness assumption since L and A_0 are not d-separated on DAG 1b. The argument in this last paragraph also proves Lemma 1.1.

Remark: One might conjecture that the problem could be solved by adding a small number of interaction terms to the model. However, using reasoning like that above, one can show that this is not the case.

2. The Direct Effect g-Null Test

An appropriate approach to testing the direct-effect null hypothesis is based on the following theorem which is a special case of Theorem 5.1 below.

Theorem 2.1: Direct Effect g-Null Theorem: Eq. (1.2) is true if and only if for any function $t(\cdot, \cdot)$

$$E\left[t(Y, A_1) / f(A_1 \mid L, A_0) \mid A_0\right] \ \textit{does not depend on } A_0 \ \textit{w.p.}1 \qquad (2.1)$$

whenever the expectation is finite. Here w.p.1 means "with probability 1."

Proof: By Fubini's theorem, the expectation in (2.1) can be written

$$\int \left\{ \int t(y, a_1) \left[\sum_{\ell=0}^{1} f(y \mid \ell, a_1, A_0) f(\ell \mid A_0) \right] dy \right\} da_1 \ .$$

Comparing the expression in brackets with Eq. (1.2) and recalling that $t(y, a_1)$ is arbitrary proves the theorem.

Theorems 2.1 and 1.1 together then have the following corollary.

Corollary 2.1: Under the suppositions of Theorem 1.1, the direct effect null hypothesis

$$A_0 \coprod Y \mid A_1, U \quad \textit{i.e.,}\ f(y \mid a_0, a_1, u) = f(y \mid a_1, u) \qquad (2.2)$$

is true if and only if Eq. (2.1) is true.

Remark: The theorem and corollary remain true if we restrict attention to the set of functions $t(y, a_1)$ that factor as

$$t(y, a_1) = t_1(y)\, t_2(a_1) \ . \qquad (2.3)$$

Indeed, they remain true if we, in addition, require $t_2(a_1)$ to be a density, i.e.,

$$1 = \int t_2(a_1)\, da_1 \ . \qquad (2.4)$$

Because of the fundamental importance of Corollary 2.1, we shall reprove the Corollary with $t(y, a_1)$ satisfying (2.3)-(2.4) using a proof that should be particularly helpful to readers familiar with testing hypotheses about DAGs using d-separation and conditional independence.

Alternative Proof of Corollary 2.1: Let DAG 1a* and DAG 1b* be "manipulated subgraphs" of DAGs 1a and 1b in which all arrows into A_1 have been removed and equip DAG 1a* with the density $f^*(v, u) = f(u) f(a_0) f(\ell \mid a_0, u) t_2(a_1) f(y \mid a_0, a_1, u)$ where $t_2(a_1)$ is the new marginal density for the now parentless variable A_1 and the remaining factors in $f^*(v, u)$ are the same as in the density $f(v, u)$ of the true causal DAG 1a. Suppose, for the moment, that DAG 1a* or 1b* generated the data with density $f^*(v, u)$. If the null hypothesis (2.2) is true, then there is no arrow from A_0 to Y and DAG 1b* generated the data. By d-separation applied to DAG 1b*, the null hypothesis (2.2) now implies the conditional independence restriction that $A_0 \coprod Y$ under $f^*(v, u)$. Given faithfulness, the converse also is true. Now it is well-known that $A_0 \coprod Y$ under $f^*(v, u)$ if and only if for all functions $t_1(y)$

$$E^*[t_1(Y) \mid A_0] \ \textit{does not depend on } A_0 \ \textit{w.p.}1 \tag{2.5}$$

where $E^*(\bullet)$ denotes expectations with respect to $f^*(v, u)$ and $E(\bullet)$ denotes expectations with respect to the true density $f(v, u)$. Since the direct-effect null hypothesis (2.2) is the same under $f^*(v, u)$ as under $f(v, u)$, Corollary (2.1) will be proved if we can establish the following.

Lemma 2.1: Eq. (2.5) is true if and only if Eq. (2.1) is true for $t(y, a_1)$ satisfying (2.3)-(2.4).

Proof of Lemma: Writing out (2.5) explicitly, we obtain

$$\sum_{\ell=0}^{1} \iint t_1(y) f(y \mid a_1, A_0, \ell) t_2(a_1) f(\ell \mid A_0) \, dy \, da_1 \ \textit{does not depend on } A_0 \textit{w.p.}1. \tag{2.6}$$

Now, upon multiplying the integrand in (2.6) by $1 = f(a_1 \mid \ell, A_0) / f(a_1 \mid \ell, A_0)$, we may rewrite (2.5) as

$$\sum_{\ell=0}^{1} \iint \{t_1(y) t_2(a_1) / f(a_1 \mid \ell_1, A_0)\} f(y \mid a_1, A_0, \ell_0) f(a_1 \mid \ell, A_0) f(\ell \mid A_0) \, dy \, da_1 \tag{2.7}$$

which is precisely (2.1).

The direct effect g-null theorem implies that any test of the conditional independence of Y and A_0 (linear in Y) that one would have used to test the direct effect null hypothesis had A_1 been parentless can still be used to test the null hypothesis when A_1 has parents L and A_0 provided, when calculating the test statistic, Y is replaced by $\mathcal{W} = Y t_2(A_1) / f(A_1 \mid L, A_0)$ or, even more generally, by $t(A_1, Y) / f(A_1 \mid L, A_0)$ where $t_2(A_1)$ and $t(A_1, Y)$ are arbitrary functions chosen by the data analyst. This prescription can be followed in a sequential randomized trial where $f(A_1 \mid L, A_0)$ is known by design. In observational studies, $f(A_1 \mid L, A_0)$ will have to be estimated from the data in a preliminary step. We suspect that, after having estimated $f(A_1 \mid L, A_0)$ in a preliminary step, causal search programs such as TETRAD can be modified to allow the analyst to test for direct effects.

As an example, suppose, by design, $A_0 \sim N(\pi_1, 1)$ and $A_1 \mid A_0, L \sim N(\pi_2(A_0, L), 1)$. If we define $U = \{f(A_1 \mid L, A_0)\}^{-1} t(Y, A_1)\{A_0 - E(A_0)\}$ where $t(\bullet,\bullet)$ is chosen by the data analyst, then, by the direct effect g-null theorem, $\sum_{i=1}^{n} U_i$ is a sum of independent and identically distributed random variables that have mean zero under the direct effect null (1.2). Therefore, provided U_i has a finite variance, $\chi \equiv \sum_i U_i / \left\{\sum_i U_i^2\right\}^{\frac{1}{2}}$ is asymptotically distributed $N(0,1)$ under the direct effect null. Thus the test that rejects when $|\chi| > 1.96$ is an asymptotically .05-level test of this null hypothesis. As discussed later, the power of the test will depend on the choice of the function $t(y, a_1)$. As noted above, a simple alternative to this test is simply to test whether $\mathcal{W}$ and A_0 are correlated using standard software.

Remark: Note if A_1 is discrete, we can choose $t_2(A_1) \equiv 1$. However, if A_1 is continuous, as in our example, $\mathcal{W}$ may not have a finite expectation or variance if $t_2(A_1) \equiv 1$ [since $E(\mathcal{W}) = \int t_2(a_1) q(a_1) da_1$ with $q(a_1) \equiv \sum_{\ell=0}^{1} \iint y \; f[y \mid \ell, a_0, a_1] f(\ell \mid a_0) f(a_0) dyda_0$]; in that case, $t_2(A_1)$ needs to be chosen to downweight the tails of $\mathcal{W}$'s distribution so that the expectation and variance of $\mathcal{W}$ will be finite.

There are some important caveats an analyst needs to be aware of. First, the direct effect g-null hypothesis (1.2) does not imply that $\mathcal{W} \equiv Y t_2(A_1) / f(A_1 \mid L, A_0)$ is independent of A_0. If it were, then for any function $g(\bullet)$, $E[g(\mathcal{W}) \mid A_0]$ would not depend on A_0 which will not be true in general, because $f(A_1 \mid L, A_0)$ may be "bound up" in the possibly non-linear function $g(\bullet)$. There is also a lack of symmetry which is not seen with conditional independence. Specifically, the direct effect g-null hypothesis (1.2) does imply that $\mathcal{W}$ and A_0 are uncorrelated. However, it does not imply that Y and $\mathcal{W}^* \equiv A_0 t_2(A_1) / f(A_1 \mid L, A_0)$ are uncorrelated, since $E(Y) E(\mathcal{W}^*) = E(Y)\left[\int t_2(a_1) da_1\right] E(A_0)$ while $E[Y\mathcal{W}^*] = E[\mathcal{W}A_0] = E[\mathcal{W}] E[A_0] = \left\{\sum_{\ell=0}^{1} \int E[Y \mid A_0, a_1, \ell] f(\ell \mid A_0) t_2(a_1) da_1\right\} E(A_0)$ where, under the direct effect g-null hypothesis (1.2), the expression in set braces is a constant independent of A_0.

In observational studies, the densities $f(a_1 \mid a_0, \ell)$ and $f(a_0)$ will not be known and thus $f(a_1 \mid a_0, \ell)$ and $E(A_0)$ will need to be estimated from the observed data $V_i, i = 1, \ldots, n$. As an example, suppose we assume $f(a_0)$ and $f(a_1 \mid \ell, a_0)$ lie in parametric families $\left\{f\left(a_0; \alpha^{(0)}\right); \alpha^{(0)} \in \boldsymbol{\alpha}^{(0)}\right\}$ and $\left\{f\left(a_1 \mid \ell, a_0; \alpha^{(1)}\right); \alpha^{(1)} \in \boldsymbol{\alpha}^{(1)}\right\}$. The maximum likelihood estimator $\widehat{\alpha} = \left(\widehat{\alpha}^{(0)}, \widehat{\alpha}^{(1)}\right)'$ maximizes the likelihood $\mathcal{L}(\alpha) = \prod_{i=1}^{n} f\left(A_{0i}; \alpha^{(0)}\right) f\left(A_{1i} \mid L_i, A_{0i}; \alpha^{(1)}\right)$. Now define $\widehat{\chi} \equiv \sum_i U_i(\widehat{\alpha}) / \widehat{\nu}^{\frac{1}{2}}$ where $\widehat{U} \equiv U(\widehat{\alpha}) = t(Y, A_1)\left\{A_0 - E_{\widehat{\alpha}^{(0)}}[A_0]\right\} / f\left(A_1 \mid L, A_0; \widehat{\alpha}^{(1)}\right)$ and the correction term $\widehat{corr} = \widehat{\Gamma}\widehat{I}^{-1}\widehat{\Gamma}'$ in the estimated variance $\widehat{\nu} \equiv \sum_i \widehat{U}_i^2 - \widehat{corr}$ accounts for the estimation of the unknown α. Here $\widehat{I}$ is the 2×2 matrix of second partial derivatives of $\log \mathcal{L}(\alpha)$ evaluated at $\widehat{\alpha}$ and $\widehat{\Gamma} = \sum_i \partial U_i(\widehat{\alpha}) / \partial \alpha$. $\widehat{\chi}$ remains asymptotically $N(0,1)$ under the direct effect null hypothesis (1.2) provided the models for $f(a_0)$ and $f(a_1 \mid \ell, a_0)$ are correctly specified. Similarly, the p-value outputted by an off-the-shelf software program testing the independence of $\mathcal{W}$ and A_0 will be greater than the true p-value when $f(A_1 \mid L, A_0)$ has been estimated (i.e., the test is conservative), although an appropriate p-value can be simply computed. [In this simple example, it would have been appropriate and perhaps simpler to have estimated $E(A_0)$ by $\sum_i A_{0i}/n$].

In contrast to the disturbing results summarized in Lemma 1.1, any parametric models for A_0 and $A_1 \mid A_0, L$ are compatible with the direct effect null hypothesis (1.2). That is, there exist joint distributions for V under which these parametric models are correctly specified and the direct effect null hypothesis (1.2) holds.

We now have valid tests for the no direct effect null hypothesis of no effect of prophylaxis therapy A_0 on Y when A_1 is fixed (set) to any particular value a_1; but ultimately we want more. In particular, we would like to estimate the size of the direct effect. To discuss this, we first need to generalize the simple example and then precisely define the direct effect. We do this in sections 3 - 5. Then we introduce direct effect structural nested models which provide a unified approach to estimation of and testing for direct effects while avoiding both the problems of standardly parameterized DAGs and of search programs such as TETRAD that rely on conditional independences.

3. The G-computation Algorithm Formula

Let G be a directed acyclic graph with a vertex set of random variables $V = (V_1, \ldots, V_M)$ with associated distribution function $F(v)$ and density function $f(v)$ with respect to the dominating measure μ. Here μ is the product measure of Lebesgue and counting measure corresponding to the continuous and discrete components of V. By the defining Markov property of DAGs, the density of V can be factored $\prod_{j=1}^{M} f(v_j \mid pa_j)$ where pa_j are realizations of parents Pa_j of V_j on G. Our results will not require that G has any missing arrows; that is, G may be taken to be complete.

We assume V is partitioned into disjoint sets A, L, and Y where the univariate outcome variable Y is the variable in V with the latest temporal occurrence, $A = \{A_0, \ldots, A_K\}$ are temporally ordered treatments or control variables given at times $t_0, \ldots, t_K$ and $L = \{L_0, L_1, \ldots, L_K\}$ are other response variables. The response variables in L_m are temporally subsequent to A_{m-1} and prior to A_m. As will be clear below, both A_m and L_m may consist of more than one variable. Now for any variable Z, let $\mathcal{Z}$ be the support (i.e., the possible realizations) of Z. For any $z_0, \ldots, z_m$, define $\overline{z}_m = (z_0, \ldots, z_m)$. By convention $\overline{z}_{-1} \equiv z_{-1} \equiv 0$. Now define a treatment regime or plan g to be a collection of $K+1$ functions $g = \{g_0, \ldots, g_K\}$ where $g_m : \overline{\mathcal{L}}_m \rightarrow \mathcal{A}_m$ maps outcome histories $\overline{\ell}_m \in \overline{\mathcal{L}}_m$ into a treatment $g_m\left(\overline{\ell}_m\right) \in \mathcal{A}_m$. If $g_m\left(\overline{\ell}_m\right)$ is a constant, say a_m^*, not depending on $\overline{\ell}_m$ for each m, we say regime g is non-dynamic and write $g = \overline{a}^*$, $\overline{a}^* \equiv (a_0^*, \ldots, a_K^*)$. Otherwise, g is dynamic. We let $\mathcal{G}$ be the set of all regimes g.

Associated with each regime g is the "manipulated" graph G_g and distribution $F_g(v)$ with density $f_g(v)$ (SGS, 1993). Given the regime $g = (g_0, g_1, \ldots, g_K)$ and the joint density

$$f(v) = f(\ell_0) f(a_0|\ell_0) f(\ell_1|a_0, \ell_0) f(a_1|\ell_1, a_0, \ell_0) \cdots f\left(y \mid \overline{\ell}_K, \overline{a}_K\right), \tag{3.1}$$

$f_g(v)$ is the density $f(v)$ except that in the factorization (3.1), $f(a_0|\ell_0)$ is replaced by a degenerate distribution at $a_0 = g_0(\ell_0)$, $f(a_1|\ell_1, a_0, \ell_0)$ is replaced by a degenerate distribution at $a_1 = g_1(\ell_0, \ell_1)$, and, in general, $f\left(a_k \mid \overline{\ell}_k, \overline{a}_{k-1}\right)$ is replaced by a degenerate distribution at $a_k = g_k\left(\overline{\ell}_k\right)$.

In the following, let $g\left(\overline{\ell}_k\right) \equiv \left(g_0\left(\overline{\ell}_0\right), \ldots, g_k\left(\overline{\ell}_k\right)\right)$ and $g_k\left(\overline{\ell}_k\right)$ denote realizations of $\overline{A}_k$ and A_k respectively. Then the marginal density $f_g(y)$ of Y under the distribution $F_g(\cdot)$ is

$$f_g(y) = \int f_g(y, \overline{\ell}_K) d\mu(\overline{\ell}_K) \equiv \int \cdots \int f(y \mid \overline{\ell}_K, g(\overline{\ell}_K)) \prod_{j=0}^{K} f\left(\ell_j \mid \overline{\ell}_{j-1}, g(\overline{\ell}_{j-1})\right) d\mu(\ell_j) \ .$$

Similarly, the marginal distribution function of Y under $F_g(\cdot)$ is

$$F_g(y) = \int \cdots \int pr\left[Y < y \mid \overline{\ell}_K, g\left(\overline{\ell}_K\right)\right] \prod_{j=0}^{K} f(\ell_j \mid \overline{\ell}_{j-1}, g(\overline{\ell}_{j-1})) d\mu(\ell_j) \ . \qquad (3.2)$$

Robins (1986) referred to (3.2) as the G-computation algorithm formula or functional for the effect of regime g on outcome Y. Throughout we assume that for each $g \in \mathcal{G}$, $f\left(\overline{\ell}_k, g\left(\overline{\ell}_k\right)\right) \neq 0$ implies $f\left(a \mid \overline{\ell}_k, g\left(\overline{\ell}_k\right)\right) \neq 0$ for all a in the support of A_k so that the RHS of (3.2) is well-defined. For A_k continuous, this positivity assumption needs to be slightly modified to properly account for measure theoretic difficulties (Gill and Robins, 1997), due to the existence of different versions of conditional distributions. Gill and Robins (1997) show that such difficulties do not arise if we assume that $f\left(\ell_k \mid \overline{\ell}_{k-1}, \overline{a}_{k-1}\right)$ is (weakly) continuous in all arguments that represent realizations of random variables with continuous distributions, which then suffices to make (3.2) a unique well-defined function of the joint distribution of the observable data V. Robins (1986) and Pearl and Robins (1995) give sufficient conditions under which (3.2) has a causal interpretation as the distribution of Y that would be observed if all subjects were treated with (i.e., forced to follow) plan g. A sufficient condition, exemplified by DAG 1a, is the following.

Assumption of g-identifiability: Any hidden variable U that is an ancestor of A_k on the causally sufficient graph generating the data is, for each k, d-separated from A_k conditional on the past $\left(\overline{L}_k, \overline{A}_{k-1}\right)$.

G-identifiability will hold in any sequential randomized trial and is assumed to hold throughout the remainder of the paper.

Informally, G-identifiability will be true if all determinants of the outcome Y that are used by patients and physicians to determine the dosage of treatment at each time k are recorded in $\left(\overline{L}_k, \overline{A}_{k-1}\right)$. It is a primary goal of epidemiologists conducting observational studies to collect data on a sufficient number of covariates in $\overline{L}_k$ to insure that G-identifiability will hold, at least approximately. However, the assumption of G-identifiability cannot be subjected to an empirical test. G-identifiability is equivalent to the assumption of sequential randomization or, equivalently, of no unmeasured confounders as used in the counter-factual approach to causal inference (Robins, 1997). It should be noted that the approach to causal inference used in this paper based on causally sufficient DAGs with hidden variables is mathematically equivalent to the approach based on counterfactuals (Robins, 1997; Pearl, 1995).

We emphasize that from this point on the paper is solely concerned with estimation of and tests concerning the functionals $F_g(y)$. As $F_g(y)$ is a function of the joint distribution of the observed data V, a reader skeptical about causal language and inference is free to regard the remainder of this paper as simply an exposition as to how one might model and estimate the functionals $F_g(y)$, putting aside why and when one might be interested in these particular functionals.

4. The Direct Effect "g"-null Hypothesis

In settings where the treatments $\overline{A} \equiv \overline{A}_K = (A_0, \ldots, A_K)$ represent a single type of treatment given at different times, an important first question is whether the "g"-null hypothesis of no effect of treatment on Y is true, i.e., whether

$$F_{g_1}(y) = F_{g_2}(y) \text{ for all } y \text{ and all } g_1, g_2 \in \mathcal{G}. \tag{4.1}$$

Eq. (4.1) is a hypothesis about the distribution of the observable V whether or not the g-identifiability assumption holds. When g-identifiability holds, (4.1) implies that the distribution of Y will be the same under any choice of regime g, and thus it does not matter whether the treatments A_k are given or withheld at each occasion k. When g-identifiability holds, we also refer to (4.1) as the g-null hypothesis, the removal of " " from around g indicating that (4.1) now has a causal interpretation.

In this paper we will not be interested in the "g"-null hypothesis (4.1). Rather, we suppose that *(i)* treatment $A_m = (A_{Pm}, A_{Zm})$ at time t_m is comprised of two different treatments A_{Pm} and A_{Zm}, (e.g., prophylaxis therapy and AZT treatment) and *(ii)* the treatment A_{Zm} is known to affect the outcome Y. The hypothetical trial in Sec. 1.1 is the special case in which $K = 1$, $L_0 \equiv 0$, $L_1 \equiv L$, $A_0 = (A_{P0}, A_{Z0})$ with $A_{Z0} \equiv 0$ for all subjects and $A_1 = (A_{P1}, A_{Z1})$ with $A_{P1} \equiv 0$ for all subjects since no subject received AZT therapy at t_0 or prophylaxis therapy at t_1. In such settings, an important first question is whether there is a direct effect of the treatment $\overline{A}_{PK} \equiv (A_{P0}, \ldots, A_{PK})$ on the outcome Y when the treatment $\overline{A}_{ZK}$ is set to any history $\overline{a}_{ZK}$.

Notational convention: We introduce the convention that the subscripts (k, ℓ, j) generally subscript the times at which treatments are given and the lack of a subscript indicates time t_K. Furthermore, the subscript i will be reserved to index study subjects $i = 1, \ldots, n$. Thus, for example, $\overline{A}_{Zi} \equiv \overline{A}_{ZKi}$ is the AZT history through t_K of subject i. Finally, as above, we will often suppress the subscript i and write $\overline{A}_{Zi}$ as $\overline{A}_Z$. Also we introduce the notational convention that for any random variable Z_k, both Z_{-1} and $\overline{Z}_{-1}$ are identically zero.

To formalize our no-direct-effect null hypothesis, let $g_P \equiv (g_{P0}, \ldots, g_{PK})$ be a collection of functions where $g_{Pm} : \overline{\mathcal{L}}_m \to \mathcal{A}_{Pm}$. Then, for history $\overline{a}_Z \in \overline{\mathcal{A}}_Z$, let $g = (g_P, \overline{a}_Z)$ be the treatment regime or plan given by $g_m\left(\overline{\ell}_m\right) = \left\{g_{Pm}\left(\overline{\ell}_m\right), a_{Zm}\right\}$. Let $\mathcal{G}_P$ be the set of all g_p. Then $F_{(g_P, \overline{a}_Z)}(y)$ is the distribution of Y that, given g-identifiability, would be observed if $\overline{A}_Z$

was set to $\overline{a}_Z$ and the treatments $\overline{A}_P$ were assigned, possibly dynamically, according to the plan g_P. If g_P is the non-dynamic regime $\overline{a}_P$, we write $F_{\overline{a}_P,\overline{a}_Z}(y)$.

Definition: The direct effect "g"-null hypothesis of no direct effect of $\overline{A}_P$ controlling for $\overline{A}_Z$ is the hypothesis

$$F_{g_{P1},\overline{a}_Z}(y) = F_{g_{P2},\overline{a}_Z}(y) \tag{4.2}$$

for all $\overline{a}_Z$ and $g_{P1}, g_{P2} \in \mathcal{G}_P$.

Remark: Given the g-identifiability assumption of Sec. 3, the hypothesis that, on the causally sufficient graph underlying our DAG G, all directed paths from any variable A_{Pm} to Y include some A_{Zk} implies the direct effect "g"-null hypothesis (4.2). Under a faithfulness assumption, the converse is also true.

An alternative characterization of the direct effect "g"-null hypothesis in terms of the conditional distributions $F_g\left(y \mid \overline{\ell}_k\right)$ for non-dynamic regimes g is provided in the following lemma whose proof is left to the reader.

Lemma 4.1: The direct effect "g"-null hypothesis (4.2) is true if and only if

$$F_{\overline{a}_P,\overline{a}_Z}\left(y \mid \overline{\ell}_k\right) \text{ does not depend on } \underline{a}_{Pk} \equiv \left(a_{Pk}, \ldots, a_{PK}\right) \tag{4.3}$$

for all $\overline{a}_P, \overline{a}_Z, y$ and $\overline{\ell}_k$.

Note $F_g\left(y \mid \overline{\ell}_k\right)$ is given by (3.2) except with the product taken from $k+1$ to K rather than from 0 to K. If we apply this Lemma to the simple example in Section 1, we recover (1.2). That is, the direct effect "g"-null hypothesis for the observed data $V = (A_0, A_1, L, Y)$ of Sec. 1 is precisely (1.2).

4.1. Failure of the usual parameterization for testing the direct effect "g"-null hypothesis

In Section 1, we saw that is was difficult to test the direct effect "g"-null hypothesis (4.3) using the usual parameterization of a DAG. These problems are exacerbated in the general case. Indeed, there are several difficulties. First, even if the densities appearing in the G-computation formula (3.2) were known for each $g \in \mathcal{G}$, since $F_g(y)$ is a high-dimensional integral, in general, it cannot be analytically evaluated for any g and thus, must be evaluated by a monte Carlo integral approximation — the monte Carlo G-computation algorithm (Robins, 1987, 1989). Second, even if $F_g(y)$ could be well-approximated for each regime g, the cardinality of the set $\mathcal{G}$ is enormous [growing at faster than an exponential rate in K (Robins, 1989)]. Thus it would be computationally infeasible to evaluate $F_g(y)$ for all g necessary to determine whether the direct effect "g"-null hypothesis held.

However, as we saw in Sec. 1, the most fundamental difficulty with the usual parameterization of a DAG in terms of the densities $f(v_j \mid pa_j)$ is that if we use standard parametric models for $f(v_j \mid pa_j)$, *(i)* there is no parameter, say ψ, which takes the value zero if and only if the direct effect "g"-null hypothesis is true, and *(ii)* the direct effect "g"-null hypothesis, even when true, may, with probability approaching 1, be rejected in large samples.

5. Direct Effect G-null Tests

As in the special case discussed in Sec. 1, an appropriate approach to testing the direct effect "g"-null hypothesis is based on the following theorem which is a corollary of Theorems 10.1 and 10.2 below. However, since Theorem 5.1 is perhaps the key result of this paper, we provide an independent proof in Sec. 5.2 below. This proof contains many of the important ideas of the paper. The proof is similar to the alternative proof of Corollary 2.1 above, except it avoids appealing to an underlying causally sufficient graph generating the data.

For any variable $\overline{X} = \overline{X}_K$, let $\underline{X}_m \equiv (X_m, \ldots, X_K)$ and let $W_m = \prod_{k=m}^{K} f\left(A_{Zk} \mid \overline{A}_{k-1}, \overline{L}_k\right)$. By convention $W_{K+1} \equiv 1$.

Theorem 5.1: Direct effect ''g''-null theorem: The direct effect "g"-null hypothesis (4.3) is true if and only if for each $m, m = 0, \ldots, K$, and each function $T_m = t_m\left(Y, \underline{A}_{Z(m+1)}\right)$,

$$E\left[T_m / W_{m+1} \mid \overline{A}_m, \overline{L}_m\right] \text{ does not depend on } A_{Pm} \text{ w.p.1} \tag{5.1}$$

whenever the expectation is finite.

We can use (5.1) to construct tests of the direct effect "g"-null hypothesis. We first provide, in Theorem 5.2 below, a quite general approach to testing (5.1). We then suggest a practical approach that allows the analyst to test (5.1) using easily available off-the-shelf software.

General Approach to Testing: First consider a sequential randomized trial where $f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m\right)$ and thus $f\left[A_{Pm} \mid A_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right], f\left[A_{Zm} \mid \overline{A}_{m-1}, \overline{L}_m\right]$ and W_{m+1} are known. Let $T^*_m\left(A_{Pm}, Y\right) \equiv t^*_m\left(A_{Pm}, Y, \underline{A}_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right)$ be a function chosen by the data analyst. Now define $U_m = \left\{T^*_m\left(A_{Pm}, Y\right) - \int T^*_m\left(a_{Pm}, Y\right) dF\left[a_{Pm} \mid A_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right]\right\} / W_{m+1}$ and let $U_\bullet = \sum_{m=0}^{K} U_m$. Then, by the direct effect "g"-null theorem, we obtain the following result.

Theorem 5.2: If the direct effect "g"-null hypothesis (4.3) is true and $var\left(U_\bullet\right) < \infty$, then $E\left(U_\bullet\right) = 0$ and $\chi \equiv \sum_i U_{\bullet i} / \left[\sum_i U^2_{\bullet i}\right]^{\frac{1}{2}}$ converges to a $N(0,1)$ random variable.

It follows from Theorem (5.2) that the so-called g-test that rejects when $\mid \chi \mid > 1.96$ is an asymptotically .05 level test of (4.3). The power of the test depends on the choice of the function t^*_m, as discussed later. The test given in Sec. 2 is the special case of the test χ with $K = 1$, $A_{Z1} \equiv A_1$, $A_{P0} \equiv A_0$, $T^*_0\left(A_{P0}, Y\right) \equiv t\left(A_1, Y\right) A_0$ and $T^*_1\left(A_{P1}, Y\right) \equiv 0$.

In an observational study, we would *(i)* specify a parametric model $f\left(A_k \mid \overline{A}_{k-1}, \overline{L}_k; \alpha\right)$ for the now unknown densities $f\left[A_k \mid \overline{A}_{k-1}, \overline{L}_k\right]$, then *(ii)* replace the unknown densities in the definition of W_{m+1} and U_m by their maximum likelihood estimates obtained by on maximizing the likelihood $\mathcal{L}(\alpha) = \prod_{i=1}^{n} \prod_{m=0}^{K} f\left(A_{ki} \mid \overline{A}_{(k-1)i}, \overline{L}_{ki}; \alpha\right)$ over α, and *(iii)* redefine $\widehat{\chi} = \sum_i U_{\bullet i}\left(\widehat{\alpha}\right) / \widehat{\nu}$, where $U_{\bullet i} \equiv U_{\bullet i}\left(\alpha_0\right)$ and the variance estimator $\widehat{\nu}$ appropriately adjusts for estimation of α. Specifically, $\widehat{\nu} = \sum_i \widehat{D}_i(0) \widehat{D}^\prime_i(0)$ where $\widehat{D}_i(0)$ is defined following Eq. (8.1a) below.

5.1. Practical approach to testing

We describe a practical approach to testing (5.1) that only requires access to easily available off-the-shelf software. Further, the approach has the advantage that we only need to correctly specify a model for the conditional mean of A_{Pm} given $A_{Zm}, \overline{A}_{m-1}, \overline{L}_{m-1}$ rather than a model for the entire conditional distribution. We describe our approach in several steps.

Step 1: Specify a parametric model $f\left(A_{Zk} \mid \overline{A}_{k-1}, \overline{L}_k; \alpha^{(1)}\right)$ and calculate the MLE $\widehat{\alpha}^{(1)}$ that maximizes $\prod\limits_i \prod\limits_{k=0}^{K} f\left(A_{Zki} \mid \overline{A}_{(k-1)\,i}, \overline{L}_{ki}; \alpha^{(1)}\right)$ and let $W_{mi}\left(\widehat{\alpha}^{(1)}\right)$ denote W_{mi} evaluated under the density indexed by $\widehat{\alpha}^{(1)}$.

Step 2: For $m = 0, \ldots, K$, specify a model for the conditional mean of A_{Pm} depending on a parameter vector $\alpha^{(0)}$

$$E\left[A_{Pm} \mid A_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right] = d\left(\alpha^{(0)\prime} \mathcal{Q}_m\right), \tag{5.2}$$

where $\mathcal{Q}_m$ is a known vector function of $A_{Zm}, \overline{A}_{m-1}, \overline{L}_m$ and $d(\bullet)$ is a known link function. For example, if A_{Pm} is dichotomous, we might choose $d(x) = \{1 + \exp(-x)\}^{-1}$. If A_{Pm} is continuous, we might choose $d(x) = x$.

Step 3: Compute an α-level test of the hypothesis that $\theta = 0$ in the extended model that adds the term $\theta \mathcal{Q}_m^* = \theta q_m^*\left(Y, A_{Zm}, \overline{A}_{m-1}, \overline{L}_m, \underline{A}_{Z(m+1)}\right) / W_{m+1}\left(\widehat{\alpha}^{(1)}\right)$ to the $\alpha^{(0)\prime}\mathcal{Q}_m$ in (5.2), where *(i)* $q_m^*(\bullet)$ is a known function chosen by the data analyst, *(ii)* in testing $\theta = 0$, we treat the $\mathcal{Q}_m^*$ as "fixed covariates" and *(iii)* we use off-the-shelf generalized estimating equation (GEE) software for "clustered" data available in S+ or SAS that regards the $A_{P0}, \ldots, A_{PK}$ on each subject as correlated (i.e., clustered). Often the software will ask the user to specify a so-called working covariance matrix; one can specify the independence covariance matrix.

It can be shown that, if our model for $f\left(A_{Zk} \mid \overline{A}_{k-1}, \overline{L}_{k-1}\right)$ and Eq. (5.2) are correctly specified, then under direct-effect null hypothesis (5.1), in large samples, the rejection rate of the α-level test in Step 3 will be less than or equal to α. That is, the test is "conservative." The reason that the test may reject, even in large samples, at a rate less than its nominal α-level is that the variance computed by the off-the-shelf software programs does not adjust for the effect of estimating $\alpha^{(1)}$ (although it does correctly adjust for the effect of estimating $\alpha^{(0)}$).

We refer to all the above tests as direct effect g-tests. A direct effect g-test is a semiparametric test since it only requires we specify a model for $f\left(A_m \mid \overline{L}_m, \overline{A}_{m-1}\right)$ rather than for the entire joint distribution of the observed data $V = \left(Y, \overline{L}, \overline{A}\right)$. In an observational study, the g-test is only guaranteed to reject at a rate no greater than its nominal level if the model for $A_m \mid \overline{A}_{m-1}, \overline{L}_m$ is correct. Again however, in contrast to the disturbing results of Lemma 1.1, any model for $f\left(A_m \mid \overline{L}_m, \overline{A}_{m-1}\right)$ will be compatible with the direct effect g-null hypothesis (4.3).

5.2. Proof of Theorem 5.1

We first state and prove a lemma. First, some definitions. Let DAG G^* be the manipulated subgraph of DAG G in which all arrows into each A_{Zk} have been removed except those from

$A_{Z0}, \ldots, A_{Z(k-1)}$ and equip G^* with the density

$$f^*(v) = t_2(\overline{a}_Z)\, f\left(y \mid \overline{a}_K, \overline{\ell}_K\right) \prod_{m=0}^{K} f\left(a_{Pm} \mid a_{Zm}, \overline{a}_{m-1}, \overline{\ell}_{m-1}\right) f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$$

where $t_2(\overline{a}_Z)$ is the new marginal density for the "composite" parentless variable $\overline{A}_Z = (A_{Z0}, \ldots, A_{ZK})$. The remaining factors are as in the density $f(v)$ of the true DAG G. Asterisks will refer to the DAG G^* and its associated density $f^*(v)$. We now outline the proof.

In the proof of the preliminary lemma, we show that the direct-effect null hypothesis (4.3) is identical under G^* and G. However, when, as in G^*, $\overline{A}_Z$ is exogenous (parentless), testing for no direct effect of A_P is equivalent for testing for no over-all effect of A_P. From my previous work (Robins, 1986), it is known how to test for no over-all effect. Finally, we show that any test under G^* implies that a corresponding test under G is obtained by reweighting the test statistic by the inverse probability of treatment with $\overline{A}_Z$, i.e., by the W_m's.

Lemma 5.1: The direct-effect "g"-null hypothesis (4.2) holds (under $f(v)$) if and only if

$$A_{Pm} \coprod^* \left(Y, \underline{A}_{Z(m+1)}\right) \mid \overline{L}_m, A_{Zm}, \overline{A}_{m-1} \tag{5.3}$$

where the $\coprod^*$ means independence under the law $f^*(v)$.

Proof: Since $F_{(g_P, \overline{a}_Z)}(y)$ does not depend on the densities $f\left(a_{Zm} \mid \overline{a}_{m-1}, \overline{\ell}_{m-1}\right)$ modified in $f^*(v)$, we have

$$F^*_{(g_P, \overline{a}_Z)}(y) = F_{(g_P, \overline{a}_Z)}(y) \ . \tag{5.4}$$

Thus, (4.2) also represents the direct-effect "g"-null hypothesis under DAG G^*. However, because the composite variable $\overline{A}_Z$ is exogenous on DAG G^*, the distribution, under $f^*(v)$, of Y when $\overline{A}_P$ follows plan g_P and $\overline{A}_Z$ is set to plan $\overline{a}_Z$ equals the conditional distribution, given $\overline{A}_Z = \overline{a}_Z$, of Y when $\overline{A}_P$ is set to g_P. That is, by direct calculation,

$$f^*_{(g_P, \overline{a}_Z)}(y) = f^*_{g_P}\left(y \mid \overline{A}_Z = \overline{a}_Z\right) \tag{5.5}$$

where, by definition, $f^*_{g_P}(v)$ is the density obtained by replacing in $f^*(v)$ the density $f\left(a_{Pm} \mid a_{Zm}, \overline{a}_{m-1}, \overline{\ell}_m\right)$ by a degenerate distribution at $a_{Pm} = g_{Pm}\left(\overline{\ell}_m\right)$ for $m = 0, \ldots, K$. Note, and this is why we introduced DAG G^*, (5.5) is false for DAG G. It follows from (5.5) that we can rewrite the direct-effect null hypothesis (4.2) as

$$F^*_{g_{P1}}\left(y \mid \overline{A}_Z = \overline{a}_Z\right) = F^*_{g_{P2}}\left(y \mid \overline{A}_Z = \overline{a}_Z\right) \ \textit{for all } g_{P1}, g_{P2} \in \mathcal{G}_P \textit{ and all } \overline{a}_Z \tag{5.6}$$

However, it is easy to show (Robins, 1987) that (5.6) is equal to the "g"-null hypothesis of no over-all effect of $\overline{A}_P$ on Y.

$$F^*_{g_{P1}}(y) = F^*_{g_{P2}}(y) \ \textit{for all } g_{P1}, g_{P2} \in \mathcal{G}_P \ . \tag{5.7}$$

The g-null theorem of Robins (1986) states that (5.7) is equivalent to

$$A_{Pm} \coprod^* Y \mid \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z \ . \tag{5.8}$$

However, by d-separation on DAG G^*, we know

$$\underline{A}_{Z(m+1)} \coprod^* \left(\overline{L}_m, \overline{A}_{Pm}\right) \mid \overline{A}_{Zm} \ .$$

Hence, on DAG G^*, (5.8) is true if and only if (5.3) is true, proving the lemma.

Proof of Theorem 5.1: It is well-known that (5.3) is true if and only if for all functions $h_m\left(Y, \underline{A}_{Z(m+1)}\right)$

$$E^*\left[h_m\left(Y, \underline{A}_{Z(m+1)}\right) \mid \overline{A}_m = \overline{a}_m, \overline{L}_m = \overline{\ell}_m\right] \text{ does not depend on } a_{Pm} \text{ w.p.1} \tag{5.9}$$

whenever the expectation is finite. Writing (5.9) out explicitly, we have

$$\iiint E\left[h_m\left(Y, \underline{A}_{Z(m+1)}\right) \mid \overline{a}_K, \overline{\ell}_K\right] \tag{5.10}$$
$$\left\{\prod_{k=m+1}^{K} dF\left[\ell_k \mid \overline{\ell}_{k-1}, \overline{a}_{P(k-1)}\right] dF\left[a_{Pk} \mid a_{Zk}, \overline{a}_{P(k-1)}, \overline{\ell}_k\right] t_2\left(a_{Zk} \mid \overline{a}_{Z(k-1)}\right) d\mu\left(a_{Zk}\right)\right\} \ .$$

Multiplying the integral in (5.10) by $1 = \prod_{k=m+1}^{K} f\left(a_{Zk} \mid \overline{\ell}_k, \overline{a}_{k-1}\right) / \prod_{k=m+1}^{K} f\left(a_{Zk} \mid \overline{\ell}_k, \overline{a}_{k-1}\right)$, we obtain

$$\iiint E\left[h_m\left(Y, \underline{A}_{Z(m+1)}\right) \prod_{k=m+1}^{K} t_2\left(a_{Zk} \mid \overline{a}_{Z(k-1)}\right) / \left\{\prod_{k=m+1}^{K} f\left(a_{Zk} \mid \overline{\ell}_k, \overline{a}_{k-1}\right)\right\} \mid \overline{a}_K, \overline{\ell}_K\right]$$
$$dF\left(\overline{\ell}_K, \overline{a}_K \mid \overline{a}_m, \overline{\ell}_m\right)$$

which equals

$$E\left[T_m / W_{m+1} \mid \overline{A}_m = \overline{a}_m, \overline{L}_m = \overline{\ell}_m\right] \text{ with } T_m \equiv h_m\left(Y, \underline{A}_{Z(m+1)}\right) \prod_{k=m+1}^{K} t_2\left(A_{Zk} \mid \overline{A}_{Z(k-1)}\right) \ .$$

Since h_m was arbitrary, T_m is arbitrary as well. This proves the theorem.

6. Direct Effect "g"-null mean hypothesis

Now we would like to estimate various "direct effect" contrasts of interest, such as the difference in the distribution functions $F_{(g_{P1}, \overline{a}_Z)}(y) - F_{(g_{P2}, \overline{a}_Z)}(y)$ for regimes g_{P1}, g_{P2}, and $\overline{a}_Z$ of interest. Often, rather than focusing on the contrasts between entire distribution functions, one may be most concerned with its estimating the effect of treatment with $\overline{A}_{PZ}$ on the *mean* of Y when $\overline{A}_Z$ is set to some value $\overline{a}_Z$ of interest. That is, we are interested in the contrast

$$E_{(g_{P1}, \overline{a}_Z)}(Y) - E_{(g_{P2}, \overline{a}_Z)}(Y) \tag{6.1}$$

where $E_g(Y) \equiv \int y dF_g(y)$ and in the following direct effect "g"-null mean hypothesis.

Definition: The direct effect "g"-null mean hypothesis of no direct effect of $\overline{A}_P$ controlling for $\overline{A}_Z$ is the hypothesis

$$E_{(g_{P1},\overline{a}_Z)}(Y) = E_{(g_{P2},\overline{a}_Z)}(Y) \tag{6.2}$$

for all $\overline{a}_Z, g_{P1}, g_{P2}$.

Note that the direct effect "g"-null hypothesis (4.2) implies the direct-effect "g"-null mean hypothesis (6.2) but the converse is false since treatment with $\overline{A}_P$ could affect the distribution of Y without affecting its mean. Similar to Lemma 4.1, we have:

Lemma: Eq. (6.2) is true if and only if

$$E_{\overline{a}}\left(Y \mid \overline{\ell}_m\right) \textit{ does not depend on } \underline{a}_{Pm} \tag{6.3}$$

for all $\overline{a} \equiv \left(\overline{a}_P, \overline{a}_Z\right), \overline{\ell}_m$.

A further equivalent characterization is given in the following.

Theorem 6.1: Direct Effect "g"-Null Mean Theorem: Eq. (6.2) is true if and only if for each m and each function $T_m = Y t_m\left(\underline{A}_{Z(m+1)}\right)$ linear in Y, $E\left[T_m/W_{m+1} \mid \overline{A}_m, \overline{L}_m\right]$ does not depend on A_{Pm}, whenever the expectation exists.

Proof Sketch: The proof is analogous to that given in Sec. 5.2, except that the "g"-null mean hypothesis

$$E^*_{g_{P1}}(Y) = E^*_{g_{P2}}(Y) \tag{6.4}$$

of G^* is equivalent to the hypothesis $E^*\left[Y \mid \overline{A}_m, \overline{L}_m, \overline{A}_Z\right]$ does not depend on A_{Pm} for each m (Robins, 1994, 1997) which replaces (5.8) in the proof. Theorem 6.1 is also a direct corollary of Theorems 7.1 and 8.1 below.

It follows that the asymptotic α-level test that rejects whenever the direct-effect g-test statistic χ of Theorem 5.2 exceeds 1.96 in absolute value is an asymptotic α-level test of the direct effect g-null mean hypothesis (6.2) provided $T^*_m(A_{Pm}, Y) \equiv T^*_{m1}(A_{Pm})Y + T^*_{m2}(A_{Pm})$ used in the definition of χ is linear in Y. Similarly, the off-the-shelf test of Sec. 5.1 can be used, provided we choose q^*_m linear in Y.

We shall now develop a unified approach to testing the direct effect g-null mean hypothesis (6.2) and estimating the contrast (6.1) based on semiparametric g-estimation of direct effect structural nested mean models (SNMMs). We will then turn our attention to development of a unified approach to testing the direct-effect "g"-null hypothesis (4.3) and estimating distributional contrasts based on g-estimation of direct effect structural nested distribution models (SNDMs).

We consider direct effect structural nested mean models (SNMMs) in addition to direct effect structural nested distribution models (SNDMs), because *(i)* the former are conceptually much easier to understand than the latter, *(ii)* SNDMs are defined only for continuous Y while SNMMs allow Y to be continuous or discrete, *(iii)* the direct effect "g"-null mean hypothesis (6.2) may be of greater subject matter interest than the direct effect "g"-null hypothesis (4.2) and *(iv)* mean contrasts are much easier to compute than distributional contrasts..

7. A New Characterization of the Direct-effect "g"-Null Mean Hypothesis

The first step in defining direct-effect SNMMs is yet another new characterization of the direct effect "g"-null mean hypothesis (6.2). Given any treatment history $\overline{a} = (\overline{a}_P, \overline{a}_Z)$ in obvious notation, adopt the convention that the treatment history that *(i)* agrees with $\overline{a}_Z$ through t_K for $\overline{A}_Z$ and *(ii)* agrees with $\overline{a}_P$ through t_m and is zero thereafter for $\overline{A}_P$ will be denoted $(\overline{a}_{Pm}, \overline{a}_Z)$ or $\left(\{\overline{a}_{Pm}, 0\}\right), \overline{a}_Z)$. Denote $E_g\left(Y \mid \overline{\ell}_m\right)$ by $b\left(\overline{\ell}_m, g\right)$ and $E_g\left(Y\right)$ *by* $b\left(g\right)$. Write $b\left(\overline{\ell}_m, g = (\overline{a})\right)$ as $b\left(\overline{\ell}_m, \overline{a}\right) = b\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ when $\overline{a} = (\overline{a}_{Pm}, \overline{a}_Z)$. Then define the "blip function"

$$\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) \equiv b\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) - b\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \ .$$

Given the g-identifiability assumption, $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ is the direct effect on the mean of Y of one final blip of A_{Pm} treatment of magnitude a_{Pm} at time t_m among subjects with history $\left(\overline{a}_{(m-1)}, \overline{\ell}_m\right)$ when treatment $\overline{A}_Z$ is set to $\overline{a}_Z$. In particular, note the "blip" function satisfies

$$\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = 0 \ if \ a_{Pm} = 0 \ . \tag{7.1}$$

Our interest in this function is based on the following theorem.

Theorem 7.1: $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = 0$ for all $y, m, \overline{a}_{Pm}, \overline{a}_Z$ if and only if the direct effect "g"-null mean hypothesis (6.2) holds.

The theorem is a corollary of Theorem (8.6) below. It also follows by noting that the direct-effect "g"-null mean hypothesis (6.2) holds for DAG G if and only if the over-all "g"-null mean hypothesis (6.4) holds for DAG G^*. This later hypothesis holds if and only if $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) \equiv 0$ by results in Robins (1994, 1997).

8. Direct Effect SNMMs

In view of Theorem 7.1, our approach will be to construct a parametric model for $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ depending on a parameter ψ such that $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) \equiv 0$ if and only if the true value ψ_0 of the parameter is zero.

Definition: The distribution F of the observables V follows a direct effect pseudo-structural nested mean model for the effect of $\overline{A}_P$ controlling for $\overline{A}_Z$ if $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = \gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi_0\right)$ where $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ is a known function depending on a finite dimensional parameter ψ and $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right) = 0$ if $a_{Pm} = 0$ or $\psi = 0$, so $\psi_0 = 0$ represents the direct effect "g"-null hypothesis (6.2). As just one example, we might consider the model $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right) = \psi_1 a_{Pm} + \psi_2 a_{Pm} a_{P(m-1)} + \psi_3 a_{Pm} \ell^*_{m-1} + \psi_4 a_{Pm} a_{ZK}$ where ℓ^*_m is a given univariate function of $\overline{\ell}_m$. When g-identifiability holds, we also refer to $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ as a direct-effect SNMM, the removal of "pseudo" reflecting the fact that $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ now has a causal interpretation.

We now consider testing and estimation of ψ_0 in the semiparametric model *(a)* characterized by the restrictions that the law of V follows the direct effect pseudo-SNMM $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$, and as in a sequential randomized trial, $f\left(a_m \mid \overline{a}_{m-1}, \overline{\ell}_m\right)$ is known. This model is referred to as semiparametric since it parameterizes some but not all of the joint distribution of V.

Our fundamental tool is the following theorem. For any function $\gamma^*\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ satisfying $\gamma^*\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = 0$ if $a_{Pm} = 0$, define $H\left(\gamma^*\right) \equiv Y - \sum_{m=0}^{K} \gamma^*\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$. The following theorem gives a useful characterization of the true blip function $\gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$.

Theorem 8.1: $\gamma^*\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) = \gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$ w. p. 1 if and only if for each $m = 0, \ldots, K$ and any function $t_m(\bullet)$

$$E\left[t_m\left(\underline{A}_{Z(m+1)}\right) H\left(\gamma^*\right) / W_{m+1} \mid \overline{A}_m, \overline{L}_m\right] \textit{ does not depend on } A_{Pm} \textit{ w.p. } 1.$$

Proof: A direct proof is given in Appendix 1. Here we sketch an alternative proof. On DAG G^*, we know by results of Robins (1994, 1997) on overall effects that $\gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$ is uniquely characterized by $E\left[H(\gamma) \mid \overline{A}_{Pm}, \overline{L}_m, \overline{A}_Z\right]$ does not depend on A_{Pm}. We use this to replace Eq. (5.8) and proceed analogously to the proof of Theorem 5.1.

Remark: Note that Theorem 6.1 is a direct corollary of Theorems 7.1 and 8.1. We can use Theorem 8.1 to construct semiparametric direct-effect g-tests and g-estimates for ψ_0. In a parallel to Sec. 5, we first provide in Theorem 8.2 below a quite general approach to testing and estimation.. We then suggest a practical approach that allows the analyst to use easily available off-the-shelf software.

8.1. General approach to testing and estimation

To describe the asymptotic properties of the estimators introduced in this Section, it will be useful to define an asymptotically linear estimator. An estimator $\widehat{\psi}$ of ψ_0 is asymptotically linear with influence function B if

$$n^{\frac{1}{2}}\left(\widehat{\psi} - \psi_0\right) = n^{-\frac{1}{2}} \sum_{i=1}^{n} B_i + o_p(1),$$

where $E(B) = 0$, $E(B'B) < \infty$, and $o_p(1)$ represents a random variable converging to zero in probability. Thus, an asymptotically linear estimator is asymptotically equivalent to the sum of independent and identically distributed random variables B_i. It follows, therefore, that if $\widehat{\psi}$ is asymptotically linear, then, by the central limit theorem and Slutzky's theorem, $n^{\frac{1}{2}}\left(\widehat{\psi} - \psi_0\right)$ is asymptotically normal with mean zero and variance $E[BB']$. Nearly all commonly encountered estimators are asymptotically linear. For example, the maximum likelihood estimator and most Bayes estimators of the parameter vector indexing a parametric model will be asymptotically linear with influence function equal to the inverse of the expected information matrix multiplied by the score vector (i.e., the derivative of the log likelihood contribution of a single subject with respect to the parameter). Two asymptotically linear estimators, $\widehat{\psi}^{(1)}$ and $\widehat{\psi}^{(2}$ with the same influence function B are asymptotically equivalent in the sense that $n^{\frac{1}{2}}\left(\widehat{\psi}^{(1)} - \widehat{\psi}^{(2)}\right)$ goes to zero in probability.

It will also be useful to define a regular estimator: a regular estimator is one whose convergence to its limiting distribution is locally uniform (Bickel et al., 1993). Regularity is a technical condition that prohibits super-efficient estimators. Thus, when we say the maximum likelihood estimator or a Bayes estimator is efficient, we mean it is efficient within the class of regular estimators. Thus, an estimator being regular, asymptotically linear (RAL) is a highly desirable property. With this background, we are ready to describe our general approach to testing and estimation.

Let $T_m^*(A_{Pm}, y) = T_{m1}^*(A_{Pm})\, y + T_{m2}^*(A_{Pm})$ where, for $j = 1, 2$, $T_{mj}^*(A_{Pm}) = t_{mj}^*\left(A_{Pm}, \underline{A}_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right)$. Write $H(\psi) = Y - \sum_{m=0}^{K} \gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z, \psi\right)$ and let $U_m(\psi, t^*) = W_{m+1}^{-1}\left\{T_m^*(A_{Pm}, H(\psi)) - \int T_m^*\{a_{Pm}, H(\psi)\}\, dF\left[a_{Pm} \mid A_{Zm}, \overline{A}_{m-1}, \overline{L}_m\right]\right\}$ and $U_\bullet(\psi) \equiv U_\bullet(\psi, t^*) \equiv \sum_{m=0}^{K} U_m(\psi, t^*)$, where t^* is the collection of functions $\left\{t_{mj}^*(\bullet); m = 0, \ldots, K, j = 0, 1\right\}$ chosen by the investigator. Also given functions $r_m\left(A_{Zm}, \overline{A}_{m-1}, \overline{L}_{m-1}\right)$ chosen by the investigator, let $R_\bullet = \sum_{m=0}^{K} R_m$, where $R_m = r_m\left(A_{Zm}, \overline{A}_{m-1}, \overline{L}_{m-1}\right) - \int r_m\left(a_{Zm}, \overline{A}_{m-1}, \overline{L}_{m-1}\right) dF\left(a_{Zm} \mid \overline{A}_{m-1}, \overline{L}_{m-1}\right)$ and define $U_\bullet^*(\psi) \equiv U_\bullet^*(\psi, t^*, r) = U_\bullet(\psi, t^*) - R_\bullet$. In the following theorem, the functions t_{mj}^* and r_m are vector valued of the dimension of ψ.

Theorem 8.2: If $var\{U_\bullet^*(\psi_0, t^*, r)\}$ is finite, then *(i)* $E[U_\bullet^*(\psi_0)] = 0$, and if ψ is one-dimensional, $\chi(\psi_0) \equiv \sum_i U_{\bullet i}^*(\psi_0) / \left[\sum_i U_{\bullet i}^*(\psi_0)^2\right]^{\frac{1}{2}}$ converges to a $N(0,1)$ random variable. Further, under standard regularity conditions, with probability approaching 1, the unique solution $\widehat{\psi} \equiv \widehat{\psi}(t^*, r)$ to $0 = \sum_i U_{\bullet i}^*(\psi, t^*, r) = 0$ is a regular asymptotically linear (RAL) estimator of ψ_0 with influence function $-\{E[\partial U_\bullet^*(\psi_0, t^*, r)/\partial\psi]\}^{-1} U_\bullet^*(\psi_0, t^*, r)$.

Remark: The test statistic χ discussed in the paragraph following Theorem 6.1 is a special case of $\chi(\psi_0)$ with $\psi_0 = 0$ and $R_\bullet \equiv 0$.

Elsewhere we prove there further exists t_{eff}^* and r_{eff} such that $\widehat{\psi}\left(t_{eff}^*, r_{eff}\right)$ is the most efficient possible estimator of ψ_0 under the restrictions imposed by our semiparametric model *(a)*. That is, the asymptotic variance of $\widehat{\psi}\left(t_{eff}^*, r_{eff}\right)$ attains the semiparametric variance bound for the model. In particular, this implies that when ψ_0 is 1-dimensional, the direct-effect g-test based on $\chi \equiv \chi(0)$ of the null hypothesis $\psi_0 = 0$ that uses t_{eff}^* and r_{eff} is locally most powerful among all regular asymptotic α-level tests (Robins and Rotnitzky, 1992) of the direct effect "g"-null mean hypothesis (6.3) under the sole restriction that, as in a sequential randomized trial, the densities $f\left(a_m \mid \overline{a}_{m-1}, \overline{\ell}_{m-1}\right)$ are known. We introduced the additional term $R_\bullet$ in $U_\bullet^*(\psi)$ to be able to characterize the most efficient semiparametric procedure; the choice $R_\bullet \equiv 0$ of Secs. 5 and 6 is inefficient.

In observational studies, we can replace the unknown density $f\left(A_m \mid \overline{L}_m, \overline{A}_{m-1}\right)$ used in $U_m(\psi)$ and W_{m+1} by maximum likelihood estimates under a parametric model $f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}; \alpha\right]$. Let $U_\bullet^*(\psi, \alpha) = U_\bullet^*(\psi, t^*, r, \alpha)$ and $\widehat{\psi}(\alpha) \equiv \widehat{\psi}(t^*, r, \alpha)$ be $U_\bullet^*(\psi, t^*, r)$ and $\widehat{\psi}(t^*, r)$ with $f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}\right]$ replaced by $f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}; \alpha\right]$. Then if the parametric model $f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}; \alpha\right]$ is correctly specified, the estimator $\widehat{\psi}(t^*, r, \widehat{\alpha})$ with $\widehat{\alpha}$ the maximum likelihood estimator of α will be RAL with influence function $D = D(\psi_0) \equiv \Gamma^{-1}(\psi_0)\, D^*(\psi_0)$ where $\Gamma(\psi_0) \equiv -E[\partial U_\bullet^*(\psi_0, t^*, r)/\partial\psi]$ and $D^*(\psi_0) \equiv U^*(\psi_0, t^*, r) - E[U^*(\psi_0, t^*, r)\, S_\alpha'][E(S_\alpha S_\alpha')]^{-1} S_\alpha$, $S_\alpha \equiv S_\alpha(\alpha_0)$ and $S_\alpha(\alpha)$ is the score for

α (i.e., the derivative of the log-likelihood w.r.t α for a single subject). Furthermore, the asymptotic variance of $\widehat{\psi}(t^*, r, \widehat{\alpha})$ will be less than or equal to that of $\widehat{\psi}(t^*, r)$. In particular, $\widehat{\psi}\left(t^*_{eff}, r_{eff}\right)$ and $\widehat{\psi}\left(t^*_{eff}, r_{eff}, \widehat{\alpha}\right)$ will have the same efficient variance. A consistent estimator of the asymptotic variance of $n^{\frac{1}{2}}\left\{\widehat{\psi}(t, r, \widehat{\alpha}) - \psi_0\right\}$ is

$$n^{-1}\sum_i \widehat{D}_i\left(\widehat{\psi}\right)\widehat{D}_i\left(\widehat{\psi}\right)' \tag{8.1a}$$

where $\widehat{D}(\psi) = \widehat{\Gamma}^{-1}\left(\widehat{\psi}\right)\widehat{D}^*(\psi)$ and $\widehat{\Gamma}(\psi)$ and $\widehat{D}^*(\psi)$ are the estimators of $\Gamma(\psi)$ and $D^*(\psi)$ obtained by replacing α by $\widehat{\alpha}$ and expectations by sample averages over the n study subjects. Similarly, with ψ 1-dimensional, if $\psi = \psi_0$

$$\widehat{\chi}(\psi) \equiv \sum_i U^*_{\bullet i}(\psi, \hat{\alpha}) / \left\{\sum_i \widehat{D}^{*2}_i(\psi)\right\}^{\frac{1}{2}} \tag{8.1b}$$

converges to a $N(0,1)$ random variable.

A practical approach to testing and estimation: A conservative α-level test of the hypothesis $\psi = \psi_0$ can be obtained using off-the-shelf software by following the practical testing algorithm of Sec. 5.1, except in Step 3 we replace $\mathcal{Q}^*_m$ by $\mathcal{Q}^*_m(\psi) =$
$q^*_m\left[H(\psi), A_{Zm}, \overline{A}_{m-1}, \overline{L}_m, \underline{A}_{Z(m+1)}\right]$ with q^*_m linear in its first argument and treat $\mathcal{Q}^*_m(\psi)$ as a fixed covariate in the testing procedure. A RAL estimator $\widetilde{\psi}$ of ψ_0 is then obtained as the value of ψ that makes this test statistic precisely zero. A conservative 95% confidence interval (i.e., an interval that is guaranteed in large samples to cover ψ_0 at least 95% of the time) can be obtained as the set of ψ for which the conservative .05-level test of $\psi = \psi_0$ of Step 3 fails to reject.

We refer to all the above estimators as direct-effect g-estimators. It will also be pedagogically useful to reconsider the simple toy example of Sec. 2 with $K = 1$, $A_{P0} \equiv A_0$, $A_{Z\,0} \equiv 0$, $A_{P1} \equiv 0$, $A_{Z\,1} \equiv A_1$, $L_0 \equiv 0$, $L_1 \equiv L$, $W_1 = f(A_1 \mid L, A_0)$, and $H(\psi) = Y - \psi A_0$. Then a test of the hypothesis $\psi = \psi_0$ is obtained by using standard software to test whether the variable $\mathcal{W}(\psi) \equiv H(\psi)\, t_2(A_1)/W_1$ is correlated with A_0 where $t_2(A_1)$ is chosen by the data analyst. Furthermore, an RAL estimator $\widehat{\psi}$ of ψ_0 is obtained by estimating the linear regression model $Y_i = \beta + \psi A_{0\,i} + \varepsilon_i$ by weighted least squares with subject i receiving the weight $t_2(A_{1\,i})/W_{1\,i}$. Note that no tests and estimators quite this simple are available in the more complex settings discussed above.

It can be shown that consideration of the abstract estimating functions $U^*_\bullet(\psi)$ is completely general in the sense that any other estimator $\widetilde{\psi}$ of ψ_0, such as the off-the-shelf g-estimator, is asymptotically equivalent to an estimator $\widehat{\psi}(t^*, r)$ for some choice of the functions t^* and r. That is, $\widetilde{\psi}$ and $\widehat{\psi}(t^*, r)$ have the same influence function and thus $n^{\frac{1}{2}}\left\{\widetilde{\psi} - \widehat{\psi}(t^*, r)\right\}$ converges in probability to zero.

Estimation of contrasts: When $\psi_0 \neq 0$, knowledge of ψ_0 alone will not allow one to calculate the mean contrasts $b(g_{P1}, \overline{a}_Z) - b(g_{P2}, \overline{a}_Z)$ with the following exception. Suppose that we were interested in contrasts between the non-dynamic regimes $g_{P1} \equiv \overline{a}_P^{(1)}$ and $g_{P2} \equiv \overline{a}_P^{(2)}$ and

$$\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) \equiv \gamma\left(\overline{a}_{Pm}, \overline{a}_Z\right) \tag{8.2}$$

does not depend on $\overline{\ell}_m$ for each m. Define $b\left(\overline{a}_Z\right) \equiv b\left(0, \overline{a}_Z\right)$ to be $b\left(g\right)$ for $g = \left(\overline{a}_P, \overline{a}_Z\right)$ with $\overline{a}_P$ identically zero. Note $b\left(\overline{a}_Z\right) - b\left(\overline{a}_Z \equiv 0\right)$ is, under g-identifiability, the effect of treatment with $\overline{A}_Z = \overline{a}_Z$ on the mean of Y when treatment with $\overline{A}_P$ is withheld. It follows from Theorem 8.3 below that $b\left(\overline{a}_P, \overline{a}_Z\right) - b\left(\overline{a}_Z\right)$ is $\sum_{m=0}^{K} \gamma\left(\overline{a}_{Pm}, \overline{a}_Z\right)$ which is only a function of the parameter ψ_0 of a correctly specified direct effect SNMM $\gamma\left(\overline{a}_{Pm}, \overline{a}_Z, \psi\right)$.

Suppose, however, we wish to estimate the ratio $b\left(\overline{a}_P^{(1)}, \overline{a}_Z\right) / b\left(\overline{a}_P^{(2)}, \overline{a}_Z\right)$ rather than the difference. The following theorem, which is a corollary to Theorem 8.6 below, indicates that in order to do so, we also need an estimate of $b\left(\overline{a}_Z\right)$.

Theorem 8.3: If (8.2) is true, $b\left(\overline{a}_P, \overline{a}_Z\right) = \sum_{m=0}^{K} \gamma\left(\overline{a}_{Pm}, \overline{a}_Z\right) + b\left(\overline{a}_Z\right)$.

With this motivation, our next goal will be to estimate $b\left(\overline{a}_Z\right)$. To do so, we modify our previous semiparametric model *(a)* to the more restrictive semiparametric model *(b)* which contains the additional assumption that $b\left(\overline{a}_Z\right) = b\left(\overline{a}_Z; \theta_0\right)$ where $b\left(\overline{a}_Z; \theta\right)$ is a known function of a finite dimensional parameter θ. The model $b\left(\overline{a}_Z; \theta\right)$ is a marginal structural model for the effect of $\overline{A}_Z$ on the mean of Y when $\overline{A}_P$ is set to zero (Robins, 1998, 1999). The key tool in constructing an estimator for θ in this model is the following characterization of $b\left(\overline{a}_Z\right)$.

Theorem 8.4: Let $\sigma\left(\gamma, b^*\right) = H\left(\gamma\right) - b^*\left(\overline{A}_Z\right)$. Then $b^*\left(\overline{A}_Z\right) = b\left(\overline{A}_Z\right)$ w. p. 1 if and only if, for all functions $t\left(\overline{a}_Z\right)$, $E\left[\sigma\left(\gamma, b^*\right) t\left(\overline{A}_Z\right) / W\right] = 0$, whenever the expectation is finite and $W \equiv W_0$.

Proof: A direct proof is given in Appendix 1. Alternately by Robins (1994, 1997), $E^*\left[H\left(\gamma\right) \mid \overline{A}_Z\right] = b\left(\overline{A}_Z\right)$ on G^*, and thus we can use the proof methods of Sec. 5.2.

Theorem 8.4 suggests the following estimation procedure for θ_0. Define $\sigma\left(\psi, \theta\right) = H\left(\psi\right) - b\left(\overline{A}_Z; \theta\right)$. Let $S\left(\theta, \psi\right) \equiv S\left(\theta, \psi, t\right) \equiv \sigma\left(\psi, \theta\right) t\left(\overline{A}_Z\right) / W$ where $t\left(\overline{A}_Z\right)$ is now a vector-valued function of the dimension of θ chosen by the data analyst. Let $\widehat{\theta} \equiv \widehat{\theta}\left(\widehat{\psi}\right) \equiv \widehat{\theta}\left(\widehat{\psi}, t\right)$ solve $0 = \sum_i S_i\left(\theta, \widehat{\psi}, t\right)$ where $\widehat{\psi} = \widehat{\psi}\left(t^*, r\right)$ is the estimator of ψ_0 defined previously. Then since Theorem 8.4 implies $E\left[S\left(\theta_0, \psi_0, t\right)\right] = 0$ under semiparametric model *(b)*, we obtain the following theorem.

Theorem 8.5: Given semiparametric model *(b)*, under standard regularity conditions, $\widehat{\rho}' = \left(\widehat{\psi}', \widehat{\theta}'\right)$ is a RAL estimator of $\rho_0' = \left(\psi_0', \theta_0'\right)$ with influence function $-\left\{E\left[\partial S^*\left(\rho_0\right) / \partial \rho\right]\right\}^{-1} S^*\left(\rho_0\right)$ where $S^*\left(\rho\right) = \left(U\left(\psi\right)', S\left(\theta, \psi\right)'\right)'$.

Remark: Note $\widehat{\psi} \equiv \widehat{\psi}\left(t^*_{eff}, r_{eff}\right)$, although efficient semiparametric model *(a)*, will in general not be an efficient estimator of ψ_0 in semiparametric model *(b)*. To see why, consider the extreme case where θ_0 was known *a priori*. Then since, for any function $t\left(\overline{A}_Z\right)$ of the dimension of ψ, $E\left[S\left(\theta_0, \psi_0, t\right)\right] = 0$ under semiparametric model *(b)*, we can use this fact to obtain an estimator, say $\widetilde{\psi}\left(t^*, r, t\right)$, that is more efficient than $\widehat{\psi}\left(t^*, r\right)$. Unfortunately, the estimator $\widetilde{\psi}\left(t^*, r, t\right)$, in contrast to $\widehat{\psi}\left(t^*, r\right)$, will be inconsistent for ψ_0 if the semiparametric model *(b)* is false due to the model $b(\overline{a}_Z;\theta)$ being misspecified. It follows that due to this possible misspecification, using $\widetilde{\psi}(t^*,r,t)$ can lead us to falsely conclude that there exists a direct effect of $\overline{A}_P$ on the mean of Y (i.e., $\psi_0 \neq 0$) even in a sequential randomized experiment with known randomization probabilities. This error is avoided by basing inference on the less efficient estimator $\widehat{\psi}(t^*,r)$.

Further contrasts: Suppose our goal is to estimate the contrast $b(g_{P1},\overline{a}_Z) - b(g_{P2},\overline{a}_Z)$ and either (8.2) is false or g_{P1} and/or g_{P2} is a dynamic regime. According to the following theorem, proved in Sec. 9., it is then unnecessary to estimate $b(\overline{a}_Z)$, but we must now estimate the densities $f\left(\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right)$. Furthermore, if we wish to estimate the ratio of the means rather than the difference, we must also estimate $b(\overline{a}_Z)$.

Theorem 8.6: $b(g_P,\overline{a}_Z) \equiv$

$$b(\overline{a}_Z) + \iint \sum_{m=0}^{K} \gamma\left[\overline{\ell}_m, g_P\left(\overline{\ell}_m\right),\overline{a}_Z\right] \prod_{m=0}^{K} dF\left[\ell_m \mid \overline{\ell}_{m-1}, g_P\left(\overline{\ell}_{m-1}\right),\overline{a}_{Z(m-1)}\right] \ . \tag{8.3}$$

Proof: Robins (1994, 1997) shows $E^*_{g_P}\left[Y \mid \overline{A}_Z = \overline{a}_Z\right]$ is given by (8.3) under G^*. However, by $\overline{A}_Z$ exogenous on G^*, $E^*_{g_P}\left[Y \mid \overline{A}_Z = \overline{a}_Z\right] = E^*_{(g_P,\overline{a}_Z)}(Y) \equiv b^*(g_P,\overline{a}_Z)$. However, all functionals in (8.3) are common to $f(v)$ and $f^*(v)$, so (8.3) must also equal $b(g_P,\overline{a}_Z)$. A quite different alternative proof is provided in Section 9.

To estimate $b(g_P,\overline{a}_Z)$, we use our previous estimates $\left(\widehat{\psi},\widehat{\theta}\right)$. Estimates of $f\left(\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right)$ that converge at rate $n^{\frac{1}{2}}$ can be obtained by specifying a parametric model $f\left(\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1};\eta\right)$ and then estimating η by $\widehat{\eta}$ that maximizes $\prod_{i=1}^{n}\prod_{m=0}^{K} f\left(L_{mi} \mid \overline{L}_{(m-1),i},\overline{A}_{(m-1),i};\eta\right)$. The integral in (8.3) can be easily evaluated by Monte Carlo integration.

9. A Non-standard Parameterization and Parametric Likelihood-based Inference

In Sections 7 and 8, we discussed inference for the direct effect of $\overline{A}_P$ on Y based on the semiparametric estimation of SNMMs. In contrast to our approach, in the graphical modelling literature, the usual approach to estimating a functional $q(F)$ [such as $b(g_P,\overline{a}_Z)$] of the joint distribution $F(v)$ of the observed data $V = \left(Y,\overline{L},\overline{A}\right)$ is to specify a fully parametric model for $F(v)$ depending on a finite dimensional parameter ρ. Then one estimates ρ and the functional $q(\rho)$ by maximum likelihood. Alternatively, one can give ρ a prior distribution and estimate the functional $q(\rho)$ by its posterior mean or median. Keeping with this spirit, we describe a reparameterization of the joint distribution of V in terms of the functions $\gamma\left(\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ and $b(\overline{a}_Z)$ that will allow a fully parametric likelihood or Bayesian approach to testing the direct effect "g"-null mean hypothesis (6.3) and estimating the functionals $b(g_P,\overline{a}_Z)$. Such an approach is an alternative to the semiparametric methods described previously. To describe this approach, we shall need to define $\nu\left(\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) \equiv b\left(\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) - b\left(\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right)$. Since, by definition, $b\left(\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right) = \int b\left(\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) dF\left[\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right]$, we have

$$\int \nu\left(\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) dF\left[\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right] = 0 \ . \tag{9.1}$$

Write

$$\varepsilon = Y - E\left[Y \mid \overline{L}_K, \overline{A}_K\right] \equiv Y - b\left(\overline{L}_K, \overline{A}_K\right) \tag{9.2}$$

so

$$E\left[\varepsilon \mid \overline{L}_K, \overline{A}_K\right] = 0 \ . \tag{9.3}$$

Having defined ε, we next note that we have, from their definitions, that

$$b\left(\overline{\ell}_K, \overline{a}_K\right) = \sum_{m=0}^{K} \gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) + \nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) + b\left(\overline{a}_Z\right) \ . \tag{9.4}$$

Finally, we shall need the fact that (9.1) implies there exists a unique function $\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ satisfying the standardization condition

$$\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = 0 \ if \ \ell_m = 0 \tag{9.5}$$

such that

$$\begin{gathered} \nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = \\ \nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) - \int \nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) dF\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right) \ . \end{gathered} \tag{9.6}$$

Specifically, (9.5) and (9.6) imply

$$\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = \nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) - \nu\left(\left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \ . \tag{9.7}$$

Combining (9.2), (9.4), and (9.6), we obtain

$$\begin{gathered} Y = \varepsilon + \sum_{m=0}^{K} \gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) + b\left(\overline{A}_Z\right) + \\ \sum_{m=0}^{K} \left\{\nu^*\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) - \int \nu^*\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) dF\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}\right)\right\} \ . \end{gathered} \tag{9.8}$$

Thus, the density of $f(v)$ factors as follows.

$$f(V) \equiv f\left(Y, \overline{L}_K, \overline{A}_K\right) = f\left(\varepsilon \mid \overline{L}_K, \overline{A}_K\right) \prod_{m=0}^{K} f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}\right] f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}\right] \tag{9.9}$$

where *(i)* ε is defined in terms of $\left(Y, \overline{L}_K, \overline{A}_K\right)$ by (9.8). Thus we have reparameterized the density $f(V)$ in terms of *(i)* the function $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ satisfying (7.1), *(ii)* the functions $\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ satisfying (9.5), *(iii)* the functions $b\left(\overline{a}_Z\right)$, *(iv)* the density of $\varepsilon \mid \overline{\ell}_K, \overline{a}_K$ subject to (9.3), *(v)* the densities $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$, and *(vi)* the densities $f\left(am \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$. We are now in a position to provide an alternative proof of Theorem 8.6 that demonstrates the usefulness of the decomposition (9.8).

Alternative Proof of Theorem 8.6: Note from its definition in Sec. 3, $f_g(v)$ is given by (9.8) and (9.9) except with the law of A_m given $\overline{L}_m, \overline{A}_{m-1}$ putting all its mass on $A_m = g_m(\overline{L}_m)$. Since $b(g_P, \overline{a}_Z) \equiv \int y dF_g(y)$ for $g = (g_P, \overline{a}_Z)$, Theorem 8.6 follows from Eq. (9.8)-(9.9) and the fact that, by (9.3) and (9.1), the terms in (9.8) in ε and ν^* have mean zero both under $f(v)$ and $f_g(v)$.

If the support of Y is the whole real line, the reparameterization *(i)-(vi)* is unrestricted in the sense that given any functions and densities satisfying *(i)-(vi)*, we can use the densities *(iv)-(vi)* to generate random variables $(\varepsilon, \overline{L}_K, \overline{A}_K)$ and then use the function *(i)-(iii)* to generate Y via (9.8). The resulting $V = (Y, \overline{L}, \overline{A})$ has a density $f(v)$ satisfying (9.9) with the functions *(i)-(iii)* the appropriate functionals of $f(v)$.

Remark A: Note, to obtain an unrestricted parameterization, it is essential to replace $\nu(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$ by $\nu^*(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$ since arbitrary functions $\nu(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$ and densities $f(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1})$ will fail to satisfy (9.1). To be more precise, each possible density $f(v)$ of the data is generated by one and only one collection of functions *(i)-(vi)*. However, if, in generating data, we modify *(i)-(vi)* by replacing $\nu^*(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$ by an unrestricted function $\nu(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$ and then replace the terms in set braces in (9.8) by this $\nu(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)$, then each density $f(v)$ is generated by many different collections of the modified functions and densities *(i)-(vi)*. Now consider a particular density $\mathring{f}(v)$ and its implied functions $\mathring{\gamma}$, $\mathring{b}$, $\mathring{\nu}$, and densities $\mathring{f}(\varepsilon \mid \overline{\ell}_K, \overline{a}_K)$, $\mathring{f}(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1})$, $\mathring{f}(a_m \mid \overline{\ell}_m, \overline{a}_{m-1})$. In particular, the function $\mathring{\nu}$ satisfies (9.1) under $\mathring{f}(v)$. Then $\mathring{f}(v)$ is the image of many different modified collections *(i)-(vi)*, precisely one of which has the functions γ, b, ν and densities *(iv)-(vi)* equal to those implied by $\mathring{f}(v)$ - the collection for which ν and density $f(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1})$ satisfy (9.1). We now consider conceptually quite distinct further restrictions on *(i)-(vi)* induced by Y having a restricted sample space.

If, as discussed further below, the support of Y is restricted (e.g., Y is discrete with bounded support), then the representation of $F(v)$ in terms of the densities and (unmodified) functions *(i)-(vi)* still holds except that the reparameterization is no longer unrestricted. These densities and functions must satisfy additional side constraints. For example, if Y is a non-negative random variable, a direct effect SNMM fails to automatically impose the constraint that $E_g(Y) \equiv b(g)$ is positive. If the mean Y can take any non-negative value (e.g., Y is a Poisson or over-dispersed Poisson random variable), we can automatically impose this restriction by modelling $\ell n\left\{b(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z) / b(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)\right\}$ by $\gamma(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi)$ which we refer to as a direct-effect multiplicative SNMM model. Although we do not investigate this possibility further, the relationship between direct-effect SNMMs and direct-effect multiplicative SNMMs is similar to that between standard SNMMs and standard multiplicative SNMMs discussed in Robins (1994, 1997) and Robins et. al. (1999). Neither direct-effect SNMMs nor direct-effect multiplicative SNMMs automatically impose the true restriction $0 \leq E_g(Y) \leq 1$ when Y is a Bernoulli random variable. Direct effect logistic SNMMs that model $\ell ogit\left\{b(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z)\right\} - \ell ogit\{b(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z)\}$ by $\gamma(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi)$ naturally impose this restriction but do not admit simple semiparametric estimators of ψ even when $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$ is known. This is because all influence functions for ψ depend on a high dimensional smooth, i.e., a conditional expectation which is left unrestricted by the model.

9.1. A fully parametric model

It follows from (9.9) that if we specify a direct effect SNMM $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ and parametric models $b\left(\overline{a}_Z, \theta\right)$, $f\left(\varepsilon \mid \overline{\ell}_K, \overline{a}_K; \eta_1\right)$, $f\left(\ell_m \mid \overline{\ell}_{m-1}\overline{a}_{m-1}; \eta_3\right)$, $\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z; \eta_2\right)$, and $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}; \alpha\right)$ subject to the restrictions

$$\int \varepsilon dF\left[\varepsilon \mid \overline{\ell}_K, \overline{a}_K; \eta_1\right] = 0 \tag{9.10}$$

and

$$\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z; \eta_2\right) = 0 \text{ } if \text{ } \ell_m = 0 \tag{9.11}$$

the contribution to the likelihood for a single subject can be written

$$f\left(V;\rho\right) \equiv f\left[Y \mid \overline{L}_K, \overline{A}_K; \psi, \theta, \eta\right] \prod_{m=0}^{K} f\left(A_m \mid \overline{L}_m, \overline{A}_{m-1}; \alpha\right) f\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}; \eta_3\right) \tag{9.12}$$

with $\eta = \left(\eta_1', \eta_2', \eta_3'\right)'$, $\rho \equiv \left(\psi', \theta', \eta', \alpha'\right)'$,

$$f\left[Y \mid \overline{L}_K, \overline{A}_K; \psi, \theta, \eta\right] \equiv f\left[\varepsilon\left(\psi, \theta, \eta_2, \eta_3\right) \mid \overline{L}_K, \overline{A}_K; \eta_1\right], \tag{9.13}$$

$$\begin{aligned}\varepsilon\left(\psi, \theta, \eta_2, \eta_3\right) = \sigma\left(\psi, \theta\right) - \{ &\sum_{m=0}^{K} \nu^*\left(\overline{L}_m, \overline{A}_{\rho(m-1)}, \overline{A}_Z; \eta_2\right) \\ - &\int \nu^*\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z; \eta_2\right) dF\left(L_m \mid \overline{A}_{m-1}, \overline{L}_{m-1}; \eta_3\right) \}\end{aligned} \tag{9.14}$$

where, again,

$$\sigma\left(\psi, \theta\right) \equiv H\left(\psi\right) - b\left(\overline{A}_Z; \theta\right), H\left(\psi\right) \equiv Y - \sum_{m=0}^{K} \gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z; \psi\right) \; .$$

Note that $f\left(V;\rho\right)$ is a non-standard parameterization of the DAG G since, for example, the parameter η_3 occurs both in $f\left(Y \mid \overline{L}_K, \overline{A}_K; \psi, \theta, \eta\right)$ and in $f\left[L_m \mid \overline{A}_{m-1}, \overline{L}_{m-1}; \eta_3\right]$. If $E\left(Y \mid \overline{L}_K, \overline{A}_K\right)$ can take any value in $(-\infty, \infty)$ the above parametrization, although non-standard, is unrestricted and so can be chosen variation independent in the sense that the parameter space for $\rho' \equiv \left(\psi', \theta', \eta', \alpha'\right)$ is the product of the parameter spaces for $\psi', \theta', \eta_1', \eta_2', \eta_3'$, and α'. If $E\left(Y \mid \overline{L}_K, \overline{A}_K\right)$ can only take any value in $(0, \infty)$, it is necessary to use multiplicative SNMMs to obtain an unrestricted parametrization that can be chosen variation independent. If $E\left(Y \mid \overline{L}_K, \overline{A}_K\right)$ can only take any value in $(0, 1)$, it is necessary to use logistic SNMMs to obtain an unrestricted parametrization that can be chosen variation independent.

9.2. Estimation

We assume we have an unrestricted variation-independent parametrization. Let $\widehat{\rho}_{MLE}$ maximize the likelihood $\prod_{i=1}^{n} f(V_i;\rho)$. Due to the fact that α only occurs in the terms $f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m; \alpha\right)$, the maximum likelihood estimators $\left(\widehat{\psi}_{MLE}, \widehat{\theta}_{MLE}, \widehat{\eta}_{MLE}\right)$ of (ψ,θ,η) are the same whether $f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m\right)$ is known (as in a sequential randomized trial), follows a parametric model depending on α, or is completely unknown. On the other hand, our semiparametric g-estimator $\widehat{\psi}$ of ψ_0 required that we either know (as in a sequential randomized trial) or model $f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m\right)$, but allowed us to leave *(a)* $f\left[\varepsilon \mid \overline{L}_K, \overline{A}_K\right]$, *(b)* $f\left(\overline{\ell}_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$, *(c)* $\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$, and *(d)* $b\left(\overline{a}_Z\right)$ completely unrestricted. In contrast, to compute $\widehat{\psi}_{MLE}$, we need to model the densities and functions *(a)-(d)*, but the densities $f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m\right)$ can be left completely unrestricted.

If all models are correctly specified, $\widehat{\psi}_{MLE}$ will be more efficient than even our most efficient g-estimator $\widehat{\psi}\left(t^*_{eff}, r_{eff}\right)$. Unfortunately, $\widehat{\psi}_{MLE}$ will be inconsistent for ψ_0 if any of the models for the densities and functions *(a)-(d)* are misspecified. [However, in contrast with a standard parameterization, there will always exist a joint distribution for V compatible with $f(V;\rho)$ for which the direct effect "g"-null mean hypothesis is true. Indeed, it will be true for any distribution $f(V;\rho)$ with $\psi = 0$.] As discussed previously, the g-estimator $\widehat{\psi}$ will be consistent if, as in a randomized trial, $f\left(a_m \mid \overline{a}_{m-1}, \overline{\ell}_m\right)$ is known. In an observational study, it will be inconsistent if the model for these densities is misspecified. However, since *(i)* we believe it is much more feasible to specify a realistic model for $f\left(a_m \mid \overline{a}_{m-1}, \overline{\ell}_m\right)$ than to specify parametric models for the densities and functions *(a)-(d)* above, and *(ii)* we do not wish to conclude that $\overline{A}_P$ has a direct effect when in truth it does not (i.e., $\psi_0 = 0$), we prefer, in the interest of robustness, a g-estimator $\widehat{\psi}$ to either the MLE $\widehat{\psi}_{MLE}$ or a Bayes estimator $\widehat{\psi}_B$ (which is asymptotically equivalent to $\widehat{\psi}_{MLE}$). We also prefer the g-estimate $\widehat{\psi}$ to $\widehat{\psi}_{MLE}$ because of computational convenience, since the likelihood $\prod_{i=1}^{n} f(V_i;\rho)$ can be quite difficult to maximize.

Remark B: The greatest computational difficulty in maximizing $\prod_{i=1}^{n} f(V_i;\rho)$ will be due to the term in set braces in (9.14). Therefore, we might consider modifying the likelihood by replacing the terms in set braces in (9.14) by a model $\nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{A}_Z; \eta_2\right)$ for $\nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{A}_Z\right)$ and maximize the modified likelihood subject to the equality constraint

$$0 = \int \nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z; \eta_2\right) dF\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}; \eta_3\right) \tag{9.15}$$

as required by (9.1). Unfortunately, this may create difficulties similar to that found in Lemma 1.1 and Theorem 1.2. Often the only parameter values (η_2, η_3) for which (9.15) holds will be (η_2^*, η_3) such that $\nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{A}_Z; \eta_2^*\right) \equiv 0$, which is a strong restriction we would not wish to impose (since it implies that L_m is not a predictor of Y). Suppose therefore we choose to maximize the modified likelihood without imposing the constraint (9.15).

This implies that we are using the modified parameterization *(i)-(vi)* described in a Remark A of Sec. 9. Hence the parameter ρ is no longer guaranteed to be identified, since many modified collections of functions *(i)-(vi)* imply the same distribution $f(v)$. However, if the parameter ρ is of small or moderate dimension, ρ will usually be identified. That is, the statement $f(V;\rho_1) = f(V;\rho_2)$ w.p.1 will be false for all ρ_1, ρ_2 in the parameter space. Assuming identification, let ρ^* be the probability limit of $\widehat{\rho}$. Suppose, as will surely be the case, (9.15) fails under ρ^*. Then even if *(i)* the model is correctly specified so that $f(V;\rho^*)$ actually generated the data and *(ii)* $\widehat{\psi}_{MLE}$ converges to $\psi^* = 0$, we cannot conclude that the direct-effect "g"-null mean hypothesis (6.2) is true. To see why, adopt the notation of Remark A of Sec. 9 and write $\overset{\circ}{f}(v) \equiv f(v;\rho^*)$ to represent the distribution generating the data with associated blip function $\overset{\circ}{\gamma}\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$. As discussed in Remark A, *(a)* the function $\overset{\circ}{\gamma}\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ will differ from the function $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z; \psi^*\right) = 0$ when (9.15) fails at ρ^* and *(b)* the direct-effect "g"-null hypothesis is the hypothesis that $\overset{\circ}{\gamma}\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = 0$. Thus, unconstrained maximization of the modified likelihood is to be avoided.

9.3. Contrasts revisited

As noted above, to estimate $b\left(g_P^{(1)}, \overline{a}_Z\right) - b\left(g_P^{(2)}, \overline{a}_Z\right)$ we must specify and estimate models for $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ and $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$. To estimate ratio contrasts between these means, we must estimate the functional $b\left(\overline{a}_Z\right)$ as well. Finally, suppose we wish to estimate the contrasts between distribution functions corresponding to the regimes $\left(g_P^{(1)}, \overline{a}_Z\right)$ and $\left(g_P^{(2)}, \overline{a}_Z\right)$. To do so we must specify and estimate a parametric model for $\nu^*\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ in which case we will have estimated enough of the joint distribution $F(v)$ of the observables that we can compute an estimate of the distribution function $F_g(y)$ for any regime g.

One relatively robust approach would be to calculate a g-estimate $\widehat{\psi}$ for the parameter ψ_0 of a SNMM and then maximize the likelihood $\prod_i f(V_i;\rho)$ of a fully parametric model over (θ, η) with ψ held fixed at $\widehat{\psi}$. Then, our estimation of the distribution function $F_g(y)$ will remain consistent with our earlier semiparametric g-test. In particular, in a sequential randomized trial with $f\left(a_m \mid \overline{a}_{m-1}, \overline{\ell}_{m-1}\right)$ known, the actual rejection rate of the direct effect "g"-null mean hypothesis will equal the nominal level. Indeed, because robustness in this sense is insured, one might even be willing, as an approximation, to maximize the modified likelihood of Remark B of Sec. 9.2 without imposing the constraint (9.15).

10. Direct-effect Structural Nested Distribution Models

In this Section, we describe the class of direct effect structural nested distribution models (SNDMs) for a continuous outcome Y. We purposely re-use (through redefinition) much of the notation introduced in the section on SNMMs in order that the connection between SNDMs and SNMMs is as clear as possible.

10.1. A New Characterization of the Direct-effect "g"-Null Hypothesis

The first step in constructing a direct-effect SNDM is a new characterization of the direct-effect "g"-null hypothesis (4.3). We assume the conditional distribution of Y given$\left(\overline{\ell}_m, \overline{a}_m\right)$ has a continuous positive density with respect to Lebesgue measure. The quantile-quantile function $\gamma\left(y\right) = F_1^{-1}\left\{F_2\left(y\right)\right\}$ mapping quantiles of $F_2\left(y\right)$ into quantiles of $F_1\left(y\right)$ is the unique function such that if Y_1 and Y_2 are distributed $F_1\left(y\right)$ and $F_2\left(y\right)$, then $\gamma\left(Y_2\right)$ is distributed $F_1\left(y\right)$.

Let $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ be the quantile-quantile function mapping quantiles of $F_{g=\left(\overline{a}_{Pm}, \overline{a}_Z\right)}\left(y \mid \overline{\ell}_m\right)$ into quantiles of $F_{g=\left(\overline{a}_{P(m-1)}, \overline{a}_Z\right)}\left(y \mid \overline{\ell}_m\right)$.

It follows from its definition as a quantile-quantile function that: *(a)* $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = y$ if $a_{Pm} = 0$; *(b)* $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ is increasing in y; and *(c)* the derivative of $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ w.r.t. y is continuous. Examples of such functions are

$$\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = y + 2a_{Pm} + 3a_{Pm}a_{P(m-1)} + 4a_{Pm}\ell_m^* + 5a_{Pm}a_{ZK} \tag{10.1a}$$

$$\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = y \exp\left\{2a_{Pm} + 3a_{Pm}a_{P(m-1)} + 4a_{Pm}\ell_m^* + 5a_{Pm}a_{ZK}\right\} \tag{10.1b}$$

where ℓ_m^* is a given univariate function of $\overline{\ell}_m$. Given the g-identifiability assumption, $\gamma\left(\overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ is the direct effect on the quantiles of the distribution of Y of one final blip of A_{Pm} treatment of magnitude a_{Pm} at time t_m among subjects with history $\left(\overline{a}_{P(m-1)}, \overline{\ell}_m\right)$ when treatment $\overline{A}_Z$ is set to $\overline{a}_Z$. Our interest in $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ is based on the following theorem.

Theorem 10.1: $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = y$ for all $y, m, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z$ if and only if the direct-effect "g"-null hypothesis (4.3) holds.

Proof: By Robins (1989, 1997) and Robins and Wasserman (1997), $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = y$ if and only if the (over-all) "g"-null hypothesis (5.7) holds on DAG G^* which (by the arguments in the proof of Lemma 5.1) is true if and only if (4.3) holds on DAG G.

10.2. Direct-effect Structural Nested Distribution Models

In view of Theorem 10.1, our approach will be to construct a parametric model for $\gamma(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z)$ depending on a parameter ψ such that $\gamma(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z) = y$ if and only if the true value ψ_0 of the parameter is 0.

Definition: The distribution F of V follows a direct-effect pseudo-structural nested distribution model $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ if $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = \gamma(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi_0)$ where (1) $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ is a known function; (2) ψ_0 is a finite vector of unknown parameters to be estimated; (3) for each value of ψ, $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ satisfies the conditions *(a)*, *(b)*, and *(c)* that were satisfied by $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$; (4) $\partial\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)/\partial\psi'$ and $\partial^2\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)/\partial\psi'\partial y$ are continuous for all ψ; and (5) $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right) = y$ if and only if $\psi = 0$ so that $\psi_0 = 0$ represents the direct-effect "g-" null hypothesis.

An example of an appropriate function $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ can be obtained from Eq. (10.1) by replacing the quantities 2, 3, 4 and 5 by the components of $\psi' = (\psi_1, \psi_2, \psi_3, \psi_4)$. We call models for $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ pseudo-structural because they are models for the distribution F of the observables V regardless of whether this distribution has a structural (i.e. causal) interpretation (as it would in a sequential randomized trial or, more generally, whenever the assumption of g-identifiability holds). When $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ does have a causal interpretation as well, we refer to our models as direct-effect SNDMs.

10.3. Semiparametric Estimation

We now consider testing and estimation of ψ_0 in the (redefined) semiparametric model *(a)* characterized by *(i)* the direct-effect SNDM $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z,\psi\right)$ and *(ii)* the densities $f\left(a_m\mid\overline{a}_{m-1},\overline{\ell}_m\right)$ are known. Our fundamental tool is the following theorem. For any function $\gamma^*\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ satisfying conditions *(a)-(c)*, satisfied by $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$, we recursively redefine the following random variables: $H_K\left(\gamma^*\right)=\gamma^*\left(Y,\overline{L}_K,\overline{A}_{Pk},\overline{A}_Z\right)$, $H_m\left(\gamma^*\right)=\gamma^*\left(H_{m+1}\left(\gamma^*\right),\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right)$, and set $H\left(\gamma^*\right)\equiv H_0\left(\gamma^*\right)$. The following theorem, proved in the Appendix, characterizes the true quantile-quantile function $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$.

Theorem 10.2: $\gamma^*\left(Y,\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right)=\gamma\left(Y,\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right)$ w.p.1 if and only if for $m=0,\dots K$ and any functions $t_m\left(\cdot,\cdot\right)$,

$$E\left[t_m\left(\underline{A}_{Z(m+1)},H\left(\gamma^*\right)\right)/W_{m+1}\mid\overline{A}_m,\overline{L}_m\right]\ \textit{does not depend on } A_{Pm}\ \textit{w.p.}1.$$

Proof: A direct proof is given in Appendix 1. Alternatively, by Robins (1989, 1997) and Robins and Wasserman (1997), the function γ is uniquely characterized on DAG G^* by $H\left(\gamma\right)\coprod^* A_{Pm}\mid\overline{A}_{P(m-1)},\overline{L}_m,\overline{A}_Z$ which can be used in place of Eq. (5.8) and the proof strategy of Sec. 5.2 adopted.

Given a SNDM, define $H\left(\psi\right)$ to be $H\left(\gamma^*\right)$ with γ^* the function $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z,\psi\right)$. We can then construct direct effect g-tests and g-estimates for ψ_0 analogous to those in Sec. 8 since Theorem 8.2 remains true with the functions $T_m^*\left(A_{Pm},y\right)$ now arbitrary functions of Y rather than only linear functions. Again there exists a t_{eff}^* and r_{eff} such that the asymptotic variance of $\widehat{\psi}\left(t_{eff}^*,r_{eff}\right)$ attains the semiparametric efficiency bound for ψ_0 in semiparametric model *(a)*. Generalizations to observational studies with unknown $f\left(A_m\mid\overline{L}_m,\overline{A}_{m-1}\right)$ is as described following Theorem 8.2 above. Further, the practical tests and estimators of Sec. 8 are available without requiring the functions q_m^* to be linear in their first argument.

10.4. Estimation of Some Contrasts

In this section, it will be convenient to adopt the following notation. Write $F_{(g_P,\overline{a}_Z)}\left(y\mid\overline{\ell}_m\right)$ as $b\left(y,\overline{\ell}_m,g_P,\overline{a}_Z\right)$, $F_{g_P,\overline{a}_Z}\left(y\right)$ as $b\left(y,g_P,\overline{a}_Z\right)$. When g_P is the $\overline{A}_P$-regime $\left(\overline{a}_{Pm},0\right)$, we write $b\left(y,\overline{\ell}_m,g_P,\overline{a}_Z\right)$ as $b\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ and $b\left(y,g_P,\overline{a}_Z\right)$ as $b\left(y,\overline{a}_{Pm},\overline{a}_Z\right)$. Finally for the $\overline{A}_P$-regime that is identically zero, we write $b\left(y,g_P,\overline{a}_Z\right)$ as $b\left(y,0,\overline{a}_Z\right)\equiv b\left(y,\overline{a}_Z\right)$. In this Section, we will propose methods to estimate some contrasts $b\left(y,g_P^{(1)},\overline{a}_Z\right)-b\left(y,g_P^{(2)},\overline{a}_Z\right)$. To do so, we shall need to estimate the quantile-quantile function $\tau\left(y,\overline{a}_Z\right)$ that maps quantiles of $b\left(y,\overline{a}_Z\right)\equiv b\left(y,0,\overline{a}_Z\right)$ into those of $b\left(y,0\right)\equiv b\left(y,0,0\right)$, i.e., $b\left(y,0,\overline{a}_Z\right)\equiv b\left\{\tau\left(y,\overline{a}_Z\right),0,0\right\}$. Note that

$\tau(y,\overline{a}_Z)$ is increasing in y and $\tau(y,\overline{a}_Z) = y$ if $\overline{a}_Z \equiv 0$. To estimate $\tau(y,\overline{a}_Z)$ we replace the semiparametric model *(a)* with the semiparametric model *(b)* which imposes the additional restriction that $\tau(y,\overline{a}_Z) = \tau(y,\overline{a}_Z;\theta_0)$ where $\tau(y,\overline{a}_Z;\theta)$ is a known increasing function of y satisfying $\tau(y,\overline{a}_Z;\theta) = y$ if $\overline{a}_Z \equiv 0$ or $\theta = 0$. Thus $\theta_0 = 0$ reflects the hypothesis $b(y,0,\overline{a}_Z) = b(y,0,0)$ of no effect of treatment $\overline{A}_Z$ on the marginal distribution of Y when treatment with $\overline{A}_P$ is withheld. The model $\tau(y,\overline{a}_Z;\theta)$ is a marginal structural transformation model in the sense of Robins (1998, 1999) for the effect of $\overline{A}_Z$ on Y when treatment with $\overline{A}_P$ is withheld. We could have used any other marginal structural model for $b(y,\overline{a}_Z)$. We have chosen the transformation model because of its relationship to the ordinary structural nested models discussed in Sec. 12. The fundamental tool in constructing an estimate for θ is the following characterization of $\tau(y,\overline{a}_Z)$. Let $\tau^*(y,\overline{a}_Z)$ be any increasing function of y satisfying $r^*(y,\overline{a}_Z) = 0$ if $\overline{a}_Z = 0$.

Define $\sigma(\tau^*) \equiv \sigma(\gamma,\tau^*) \equiv \tau^*\left\{H(\gamma),\overline{A}_Z\right\}$. Given any function $t(\bullet,\bullet)$ and density $t_2(\overline{a}_Z)$ for $\overline{A}_Z$, i.e.,

$$\int t_2(\overline{a}_Z)\, d\mu(\overline{a}_Z) = 1, \tag{10.2}$$

let $c_1(\overline{a}_Z) = E\left[t\left\{\sigma(\tau^*),\overline{a}_Z\right\} t_2\left(\overline{A}_Z\right)/W\right]$ and $c_2(y) = E\left[t\left\{y,\overline{A}_Z\right\}/W\right]$ with $W = W_0$.

Theorem 10.3: The following are equivalent. *(i)* $\tau^*\left(y,\overline{A}_Z\right) = \tau\left(y,\overline{A}_Z\right)$ w.p.1., *(ii)* for any function $t(\bullet,\bullet)$ and density $t_2(\overline{a}_Z)$

$$E\left\{W^{-1}\left[t\left\{\sigma(\tau^*),\overline{A}_Z\right\} - c_1\left(\overline{A}_Z\right)\right]\right\} = 0 \tag{10.3}$$

whenever the expectation is finite.

(iii) for any $t(\bullet,\bullet)$ and density $t_2(\overline{a}_Z)$

$$E\left\{W^{-1}\left[t\left\{\sigma(\tau^*),\overline{A}_Z\right\} - t_2\left(\overline{A}_Z\right) c_2\left\{\sigma(\tau^*)\right\}\right]\right\} = 0 \tag{10.4}$$

whenever the expectation is finite.

Proofs of Theorems 10.3 - 10.5 are given at the end of this section.

Remark: The theorem is false if $t_2(\overline{a}_Z)$ is not a density.

Given a g-estimate $\widehat{\psi}$ of ψ_0, Eq. (10.3) suggests the following estimator $\widehat{\theta}^{(1)}$ of θ_0. Define $\sigma(\psi,\theta) = \tau\left\{H(\psi),\overline{A}_Z;\theta\right\}$ and let

$$\widehat{S}_1(\theta,\psi) \equiv \widehat{S}_1(\theta,\psi,t,t_2) =$$
$$W^{-1}\left[t\left\{\sigma(\psi,\theta),\overline{A}_Z\right\} - n^{-1}\sum_{i=1}^{n} t_2\left(\overline{A}_{Z\,i}\right) t\left\{\sigma_i(\psi,\theta),\overline{A}_Z\right\}/W_i\right],$$

where $t_2(\overline{a}_Z)$ satisfies (10.2). Let $\widehat{\theta}^{(1)} = \widehat{\theta}^{(1)}\left(\widehat{\psi}\right) \equiv \widehat{\theta}^{(1)}\left(\widehat{\psi},t,t_2\right)$ solve $0 = \sum\limits_{i=1}^{n} \widehat{S}_{1\,i}\left(\theta,\widehat{\psi},t,t_2\right)$.

Similarly, Eq. (10.4) suggests the following alternative estimator $\widehat{\theta}^{(2)}$. Let

$$\widehat{S}_2(\theta,\psi) = \widehat{S}_2(\theta,\psi,t,t_2) =$$
$$W^{-1}\left[t\left\{\sigma(\psi,\theta),\overline{A}_Z\right\} - t_2\left(\overline{A}_Z\right)\left[n^{-1}\sum_{i=1}^{n} t\left\{\sigma(\psi,\theta),\overline{A}_{Zi}\right\}/W_i\right]\right]$$

and let $\widehat{\theta}^{(2)} = \widehat{\theta}^{(2)}\left(\widehat{\psi}\right) \equiv \widehat{\theta}^{(2)}\left(\widehat{\psi}, t, t_2\right)$ solve $0 = \sum_{i=1}^{n} \widehat{S}_{2i}\left(\theta, \widehat{\psi}, t, t_2\right)$. Then, under regularity conditions, by Theorem 10.4, $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$ will be RAL estimators of θ_0.

Given RAL estimators $\widehat{\theta}$ and $\widehat{\psi}$ of θ_0 and ψ_0, we can use the following theorem to estimate the distribution $b(y,0) \equiv b(y,0,0)$ of Y had all subjects remained untreated with either treatment.

Theorem 10.4: For any function $t(\overline{a}_Z)$, not necessarily a density,

$$E\left[W^{-1} t\left(\overline{A}_Z\right)\{I[\sigma(\tau) \leq y] - b(y,0,0)\}\right] = 0 \tag{10.5}$$

whenever the expectation is finite where, for any event Z, $I(Z) = 1$ if Z is true and 0 otherwise.

Theorem 10.4 implies that, having chosen $t(\overline{a}_Z)$ to insure integrability, if $\overline{A}_Z$ is continuous, the estimator $\widehat{b}(y,0,0) = b\left(y,0,0;\widehat{\psi},\widehat{\theta}\right) \equiv \sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right) I\left[\sigma_i\left(\widehat{\psi},\widehat{\theta}\right) \leq y\right] / \sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right)$ solving $\sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right)\left\{I\left[\sigma_i\left(\widehat{\psi},\widehat{\theta}\right) \leq y\right] - b(y,0,0)\right\} = 0$ will be a RAL estimator of $b(y,0,0)$. Thus $\widehat{b}(y,0,0)$ is discrete with support at the $\sigma_i\left(\widehat{\psi},\widehat{\theta}\right)$ with density

$$\widehat{f}_{0,0}(y) = \sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right) I\left\{\sigma_i\left(\widehat{\psi},\widehat{\theta}\right) = y\right\} / \sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right) \ . \tag{10.6}$$

Furthermore, we can use the following theorem to estimate the non-dynamic contrast $b\left(y,\overline{a}_P^{(1)},\overline{a}_Z\right) - b\left(y,\overline{a}_P^{(2)},\overline{a}_Z\right)$ whenever, for $m = 0,\ldots,K$,

$$\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right) = \gamma\left(y,\overline{a}_{Pm},\overline{a}_Z\right) \text{ does not depend on } \overline{\ell}_m \ . \tag{10.7}$$

First we need to develop some additional notation. Let $h\left(y,\overline{\ell},\overline{a}\right)$ be the function such that

$$H \equiv H(\gamma) = h\left(Y,\overline{L},\overline{A}\right) \ . \tag{10.8}$$

This function is increasing in y and satisfies $h\left(y,\overline{\ell},\overline{a}\right) = h(y,\overline{a})$ if (10.7) is true. For any function $q(y,\bullet,\bullet)$ increasing in y, we define $q^{-1}(y,\bullet,\bullet)$ to be the function satisfying $q^{-1}(u,\bullet,\bullet) = y$ if $q(y,\bullet,\bullet) = u$.

Theorem 10.5: If (10.7) holds, then for any regime $\overline{a} = (\overline{a}_P,\overline{a}_Z)$

$$b(y,\overline{a}_P,\overline{a}_Z) = pr\left[h^{-1}\left\{\tau^{-1}(X,\overline{a}_Z),\overline{a}\right\} < y\right]$$

where X is a random variable with distribution $b(y,0,0)$.

Thus to estimate $b(y,\overline{a}_P,\overline{a}_Z)$ under (10.7), we compute $pr\left[h^{-1}\left\{\tau^{-1}\left(X,\overline{a}_Z;\widehat{\theta}\right),\overline{a},\widehat{\psi}\right\} < y\right]$ with X drawn from the density $\widehat{f}_{0,0}(y)$ of (10.6), where $h^{-1}(\bullet,\bullet,\psi)$ is the function $h^{-1}(\bullet,\bullet)$ based on the blip function $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z,\psi\right) = \gamma\left(y,\overline{a}_{Pm},\overline{a}_Z,\psi\right)$. We obtain $b\left(y,\overline{a}_P,\overline{a}_Z;\widehat{\psi},\widehat{\theta}\right) = \sum_i I\left[h^{-1}\left\{\tau^{-1}\left(\sigma_i\left(\widehat{\psi},\widehat{\theta}\right),\overline{a}_Z;\widehat{\theta}\right),\overline{a},\widehat{\psi}\right\} < y\right] W_i^{-1} t\left(\overline{A}_{Zi}\right) / \sum_i W_i^{-1} t\left(\overline{A}_{Zi}\right)$.

When either (10.7) is false or g_p is a dynamic regime, it is much more difficult to obtain an estimator of $b(y,g_P,\overline{a}_Z)$ that is consistent with our direct-effect SNDM. We return to this issue in Sec. 11.

Proof of Theorem 10.3: A direct proof is given in Appendix 1. Here is an alternative proof. Under DAG G^*, $pr^*\left[H(\gamma) > y \mid \overline{A}_Z\right] = b\left(y,0,\overline{A}_Z\right)$ by results in Robins (1989, 1997) on estimation of the overall effect of $\overline{A}_P$ on G^*. Furthermore, these results imply that $\tau(\bullet,\bullet)$ is the unique function such that

$$\sigma \equiv \tau\left\{H(\gamma), \overline{A}_Z\right\} \coprod^* \overline{A}_Z \ . \tag{10.9}$$

However, (10.9) is equivalent to

$$E^*\left[q(\sigma, \overline{a}_Z) - c_2(\sigma)\right] = 0 \tag{10.10}$$

for all $q(\bullet,\bullet)$ where $\sigma \equiv \sigma(\tau) \equiv \sigma(\gamma,\tau)$ and $c_2(y) \equiv E^*\left\{q\left(y, \overline{A}_Z\right)\right\}$. However, for any random variable N,

$$E^*(N) = E\left[Nt_2\left(\overline{A}_Z\right)/W\right], \tag{10.11}$$

where $t_2(\overline{a}_Z)$ is the density of $\overline{A}_Z$ on G^*. Invoking (10.11) and defining $t(\sigma,\overline{a}_Z) \equiv q(\sigma,\overline{a}_Z)\, t_2(\overline{a}_Z)$, we find that (10.10) is equivalent to (10.4) with $\sigma(\tau^*) \equiv \sigma$.

On the other hand, (10.9) is also equivalent to

$$E^*\left[q\left(\sigma, \overline{A}_Z\right) - c_1^*\left(\overline{A}_Z\right)\right] = 0 \tag{10.12}$$

for all $q(\bullet,\bullet)$ where $c_1^*(\overline{a}_Z) = E^*\left[q(\sigma,\overline{a}_Z)\right]$. Using (10.11) and putting $t(\sigma,\overline{a}_Z) = t_2(\overline{a}_Z)\, q(\sigma,\overline{a}_Z)$ and noting $t_2(\overline{a}_Z)\, c_1^*(\overline{a}_Z) = E^*\left[t_2(\overline{a}_Z)\, q(\sigma,\overline{a}_Z)\right] = c_1(\overline{a}_Z)$, we conclude (10.12) is equivalent to (10.3) with $\sigma(\tau^*) = \sigma$.

Proof of Theorem 10.4: Under DAG G^*, it follows from Robins (1989, 1997) and (10.9) that $\sigma = \sigma(\gamma,\tau)$ has distribution $b(y,0,0)$. Hence, $E^*\left[I\{\sigma < y\} - b(y,0,0)\right] = 0$. Invoking Eq. (10.11) completes the proof. An alternative direct proof is given in the Appendix.

Proof of Theorem 10.5: This is an immediate consequence of the definitions of $\tau(y,\overline{a}_Z)$, $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$, and $h\left(y,\overline{\ell}_K,\overline{a}_K\right)$.

11. Parametric Likelihood-based Inference for Direct Effect SNDMs

We now describe a reparameterization of the distribution of V in terms of the functions $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ and $\tau(y,\overline{a}_Z)$ that will allow a fully parametric likelihood or Bayesian approach to testing the direct effect "g"-null hypothesis and estimating the functionals $b(y,g_P,\overline{a}_Z)$. To describe this approach, we first define $\nu\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$ by

$$b\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) = b\left\{\nu\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right),\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right\} \ . \tag{11.1}$$

That is, $\nu\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$ is the unique function mapping quantiles of $F_{\overline{a}}\left(y \mid \overline{\ell}_m\right)$ into quantiles of $F_{\overline{a}}\left(y \mid \overline{\ell}_{m-1}\right)$ with $\overline{a} = \left\{\left(\overline{a}_{P(m-1)},0\right),\overline{a}_Z\right\}$.

Thus, since by definition, $\int b\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) dF\left(\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right) = b\left(y,\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right)$, it follows that

$$\int b\left[\nu\left\{y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right\},\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right] dF\left[\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_{Z(m-1)}\right] = b\left(y,\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z\right) \tag{11.2}$$

Now we recursively define the random variables

$$\mathcal{H}_K, \mathcal{M}_K, \mathcal{H}_{K-1}, \mathcal{M}_{K-1}, \ldots, \mathcal{H}_0, \mathcal{M}_0$$

by $\mathcal{H}_K = \gamma\left(Y, \overline{L}_K, \overline{A}_{PK}, \overline{A}_Z\right)$, $\mathcal{M}_k = \nu\left(\mathcal{H}_k, \overline{L}_k, \overline{A}_{P(k-1)}, \overline{A}_Z\right)$, $\mathcal{H}_k = \gamma\left(\mathcal{M}_{k+1}, \overline{L}_k, \overline{A}_{Pk}, \overline{A}_Z\right)$. Finally, define $\varepsilon \equiv \tau\left(\mathcal{M}_0, \overline{A}_Z\right)$. We also write $\varepsilon = d\left[Y, \overline{L}_K, \overline{A}_K\right]$ to emphasize that ε is a function $d(\cdot,\cdot,\cdot)$ of the data $\left(Y, \overline{L}_K, \overline{A}_K\right)$ that is increasing in Y. Thus we can write $Y = d^{-1}\left(\varepsilon, \overline{L}_K, \overline{A}_K\right)$. We prove the following in Appendix 2.

Theorem 11.1: *(i)*

$$pr\left[\mathcal{H}_m > y \mid \overline{L}, \overline{A}\right] = b\left(y, \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) \tag{11.3}$$

so that

$$\mathcal{H}_m \coprod \left(\underline{A}_{Pm}, \underline{L}_{m+1}\right) \mid \overline{L}_m, \overline{A}_Z, \overline{A}_{P(m-1)} \ ; \tag{11.4}$$

(ii)

$$pr\left[\mathcal{M}_m > y \mid \overline{L}, \overline{A}\right] = b\left(y, \overline{L}_{m-1}, \overline{A}_{P(m-1)}, \overline{A}_Z\right) \tag{11.5}$$

so

$$\mathcal{M}_m \coprod \left(\underline{L}_m, \underline{A}_{Pm}\right) \mid \overline{L}_{m-1}, \overline{A}_{P(m-1)}, \overline{A}_Z \ ; \tag{11.6}$$

(iii)

$$pr\left[\varepsilon > y \mid \overline{L}, \overline{A}\right] = b\left(y, 0, 0\right) \tag{11.7}$$

so

$$\varepsilon \coprod \left(\overline{L}, \overline{A}\right) \ . \tag{11.8}$$

Remark: $H_m \equiv H_m(\gamma)$ and $\sigma \equiv \sigma(\gamma, \tau)$, in contrast to $\mathcal{H}_m$ and ε, are not conditionally independent of any components of $\left(\overline{L}, \overline{A}\right)$ given any other components. In particular

$$H_m \coprod A_{Pm} \mid \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z \tag{11.9}$$

is false (although true on DAG G^*) for, were it true, the inverse weight W_{m+1} would not be necessary in Eq. (5.1) or in the expectation in Theorem 10.2. Similarly, if σ were independent of $\overline{A}_Z$, the inverse weight W would not be necessary in Eqs. (10.3) and (10.4) of Theorem 10.3.

Theorem 11.1 implies that the density $f(V)$ factors as follows.

$$f(V) \equiv f\left(Y, \overline{L}_K, \overline{A}_K\right) \equiv \left\{\partial\varepsilon/\partial Y\right\} f(\varepsilon) \prod_{m=0}^{K} f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}\right] f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1}\right] \tag{11.10}$$

where $\partial\varepsilon/\partial Y \equiv \partial d\left(Y, \overline{L}_K, \overline{A}_K\right)/\partial Y \equiv \left\{\partial d^{-1}\left(\varepsilon, \overline{L}_K, \overline{A}_K\right)/\partial\varepsilon\right\}^{-1}$. Thus we have reparameterized $f(V)$ in terms of the functions *(i)* $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$, *(ii)* $\tau\left(y, \overline{a}_Z\right)$, *(iii)* $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$

satisfying (11.2), *(iv)* the density $f(\varepsilon)$ of ε, *(v)* the densities $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$, and *(vi)* the densities $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$.

However, this reparameterization is not unrestricted since the constraint (11.2) will not hold for arbitrary densities $f\left(\ell_m \mid \overline{\ell}_{m-1}, a_{m-1}\right)$ and functions $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$. An unrestricted parameterization is necessary in order to allow unconstrained likelihood-based inferences, and, more importantly, to avoid difficulties analogous to those discussed in Remark B of Sec. 9.2. Therefore, we replace in our parameterization the functions $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ by the hazard ratio $\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$, where

$$\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = \lambda\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) / \lambda\left(y, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \quad (11.11)$$

so that

$$\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = 1 \textit{ if } \ell_m = 0 \textit{ and } \nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \textit{ is non-negative} . \quad (11.12)$$

Here $\lambda\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \equiv -\partial\left\{\ln\left[1 - b\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)\right]\right\} / \partial y$. It may appear from Eq. (11.11) that in addition to the function ν^* our parameterization also depends on the hazard function $\lambda\left(y, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$. However, $\lambda\left(y, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ is completely determined (and thus not part of the parameterization) by the law $F_\varepsilon(y)$, $\tau(y, \overline{a}_Z)$, $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$, $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)$ and $\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$.

Specifically, the $\lambda\left(y, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ are obtained recursively from these other functions and densities as follows. First, $b(y, 0, \overline{a}_Z) \equiv b\left(y, \overline{\ell}_{-1}, \overline{a}_{-1}, \overline{a}_Z\right)$ is given by $b\left(\tau^{-1}(y, \overline{a}_Z), 0, 0\right) \equiv F_\varepsilon\left(\tau^{-1}(y, \overline{a}_Z)\right)$, where we have used our convention $\overline{z}_{-1} \equiv 0$. Now for $m = 0, \ldots, K-1$, given $b\left(y, \overline{\ell}_{m-1}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$, it follows from Robins, Rotnitzky and Scharfstein (1999, Sec.8.7a) that $\lambda\left(y, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ is the unique solution $r\left(y, \overline{\ell}_{m-1}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ to the Volterra-like integral equation

$$r\left(y, \overline{\ell}_{m-1}, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = \frac{\partial\left\{b\left(y, \overline{\ell}_{m-1}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)\right\} / \partial y}{\int \exp\left[-\int_{-\infty}^{y} \nu^*\left(u, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) r\left(u, \overline{\ell}_{m-1}, \overline{a}_{P(m-1)}, \overline{a}_Z\right) du\right] \nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) dF\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}\right)}.$$

Thus $b\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = 1 - \exp\left[-\int_{-\infty}^{y} \nu^*\left(u, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) \lambda\left(u, \left\{\overline{\ell}_{m-1}, \ell_m = 0\right\}, \overline{a}_{P(m-1)}, \overline{a}_Z\right)\right] du$. Further $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ can now be obtained by (11.1). Finally, $b\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) = b\left\{\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right), \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right\}$. We thus are in a position now to repeat the recursion with m substituted for $m-1$.

It follows from the factorization (11.10) that if we specify a direct effect SNDM $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z, \psi\right)$ and parametric models $\tau(y, \overline{a}_Z; \theta)$, $f(\varepsilon; \eta_1)$, $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}, \eta_3\right)$, $\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z; \eta_2\right)$ and $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}; \alpha\right)$ subject to the restrictions $\tau(y, \overline{a}_Z; \theta) = y$ if $\overline{a}_Z \equiv 0$ and the nonnegative function $\nu^*\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z; \eta_2\right) = 1$ if $\ell_m = 0$, then the contribution to the likelihood for a single subject can be written

$$f(V;\rho) =$$

$$\{\partial\varepsilon(\psi,\theta,\eta_2,\eta_3)/\partial Y\} f[\varepsilon(\psi,\theta,\eta_2,\eta_3);\eta_1] \prod_{m=0}^{K} f\left[A_m \mid \overline{L}_m, \overline{A}_{m-1};\alpha\right] f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1};\eta_3\right] .$$

The parameterization is non-standard because the parameter η_3 occurs in two terms. Note that $\varepsilon(\psi,\theta,\eta_2,\eta_3)$ has no simple relationship to $\sigma(\psi,\theta)$ in contrast to what we found with SNMM models. The robustness properties of $\widehat{\psi}_{MLE}$ are analogous to those described for direct-effect SNMM models. The above parametrization is variation-independent in the sense that any value of any parameter can occur with any value of any other parameter. However, the parametrization (11.11) in terms of hazard ratios requires that the measure $b\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$ be absolutely continuous with respect to $b\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$ for all ℓ_m, which we henceforth assume to be the case.

Remark: We could have tried to avoid the assumption of absolute continuity mentioned above by redefining the function $\nu^*\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$ to be the unique function satisfying

$$\nu^*\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) = \nu\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) - \nu\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$$

and

$$\nu^*\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right) = 0 \ if \ \ell_m = 0 .$$

It may appear that, in addition to the function ν^*, this alternative parameterization also depends on the function $\nu\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$. However, it can be shown that $\nu\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$ is completely determined (and thus not part of the parameterization) by the law $F_\varepsilon(y)$, $\tau(y,\overline{a}_Z)$, $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$, $f\left(\ell_m \mid \overline{\ell}_{m-1},\overline{a}_{m-1}\right)$ and $\nu^*\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$. However, as discussed in Sec.(8.7a) of Robins, Rotnitzky and Scharfstein (1999), this alternative parametrization will not be unrestricted (i.e.,variation independent) if one allows $\nu^*\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_Z\right)$ to be non-monotone in y. Non-monotonicity is required if all possible laws of the observed data are to be represented by the parametrization.

The computational difficulty in computing $\widehat{\rho}_{MLE}$ is much greater for direct-effect SNDMs than for direct-effect SNMMs, since at each iteration $\lambda\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$ must be recursively computed by solving a series of Volterra-like integral equations,as described above. With an estimator of ρ in hand, we can estimate, through (11.10), the entire joint distribution $F(v)$ of the observables and thus $F_g(y)$ for any regime g. A robust and computationally less demanding approach to estimating ρ than maximum likelihood is to *(i)* calculate a g-estimate $\widehat{\psi}$ of ψ_0, *(ii)* compute $\widehat{\theta} = \widehat{\theta}\left(\widehat{\psi}\right)$ as described in Sec. 10, *(iii)* estimate $\hat{\eta}_3$ by maximizing $\prod_i \prod_{m=0}^{K} f\left(L_{mi} \mid \overline{L}_{(m-1)i},\overline{A}_{(m-1)i};\eta_3\right)$, and *(iv)* finally estimate η_1 and η_2 by maximizing $\prod_i f(V_i;\rho)$ over (η_1,η_2) with (ψ,θ,η_3) fixed at $\left(\widehat{\psi},\widehat{\theta},\widehat{\eta}_3\right)$. This guarantees the estimator of $F_g(y)$ is consistent with our semiparametric g-test of Sec. 3.

However, even this option may be computationally difficult because of the need to estimate the $\lambda\left(y,\left\{\overline{\ell}_{m-1},\ell_m=0\right\},\overline{a}_{P(m-1)},\overline{a}_Z\right)$ at each iteration. In the next Section, we discuss a different parameterization which will turn out to have difficulties of its own but also, under further assumptions, much to recommend it.

12. An Alternative Approach to Estimation of Contrasts

Consider again estimation of $F_{g_P,\overline{a}_Z}(y)\equiv b(y,g_P,\overline{a}_Z)$. It follows from Robins (1989, 1997) and Robins and Wasserman (1997) that, under the law $f^*(v)$ of DAG G^* of Sec. 5.1 with $\overline{A}_Z$ exogenous, $F^*_\sigma(y)=b(y,0,0)$, reflecting the fact that $b(y,0,0)\equiv F^*_{(g_P\equiv 0,\overline{a}_Z\equiv 0)}(y)=F^*_{g_P\equiv 0}\left(y\mid\overline{A}_Z\equiv 0\right)$. Further, we obtain independent realizations, say Y^*, from $F^*_{g_P}(y\mid\overline{a}_Z)$ by the following Monte Carlo algorithm:

Step 1: Draw σ from $b(y,0,0)$;

Step 2: Draw L_0 from $f^*(\ell_0\mid\overline{a}_Z,\sigma)$;

Step 3: Recursively for $m=1,\dots,K$, draw L_m from $f^*\left(\ell_m\mid\overline{L}_{m-1},g_P\left(\overline{L}_{m-1}\right),\overline{a}_Z,\sigma\right)$;

Step 4: Compute $Y^*=h^{-1}\left\{\tau^{-1}(\sigma,\overline{a}_Z),\overline{L}_K,g_P\left(\overline{L}_K\right),\overline{a}_Z\right\}$ where again h^{-1} is the inverse function to $H=h\left(Y,\overline{L}_K,\overline{A}_{PK},\overline{A}_{ZK}\right)$ and is a functional of $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$.

However, as shown in the proof of Lemma 5.1, $F^*_{g_P}(y\mid\overline{a}_Z)=F^*_{g_P,\overline{a}_Z}(y)=F_{g_P,\overline{a}_Z}(y)$, so we can in principle use the above algorithm to draw from our target distribution $F_{g_P,\overline{a}_Z}(y)$ as follows. From a g-estimator $\widehat{\psi}$, we obtain an estimate $h^{-1}\left(\cdot,\cdot,\cdot,\cdot,\widehat{\psi}\right)$ of $h^{-1}(\cdot,\cdot,\cdot,\cdot)$. From an estimator $\widehat{\theta}=\widehat{\theta}\left(\widehat{\psi}\right)$ of θ_0, we obtain an estimator $\tau^{-1}\left(\cdot,\cdot,\widehat{\theta}\right)$ of $\tau^{-1}(\cdot,\cdot)$. Further, from $\left(\widehat{\theta},\widehat{\psi}\right)$, we can draw from $b(y,0,0)$ using the estimated distribution $b\left(y,0,0;\widehat{\theta},\widehat{\psi}\right)$ with density $\widehat{f}_{0,0}(y)$ given in (10.6). Hence to implement an estimated version of the algorithm, it only remains necessary to estimate

$$f^*\left(\ell_m\mid\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z,\sigma\right)\ . \tag{12.1}$$

Suppose we could correctly specify a parametric model for (12.1) depending on a parameter η.

$$f\left(\ell_m\mid\overline{\ell}_{m-1},\overline{a}_{P(m-1)},\overline{a}_Z,\sigma;\eta\right)\ . \tag{12.2}$$

Then had the data been generated under G^*, the score $S(\eta)=$
$\partial\ell n\, f\left(L_m\mid\overline{L}_{m-1},\overline{A}_{P(m-1)},\overline{A}_Z,\sigma;\eta\right)/\partial\eta$ has mean zero under $f^*(v)$ at the true value of η. Eq. (10.11) then implies

$$E\left[t_2\left(\overline{A}_Z\right)S(\eta)/W\right]=0\ . \tag{12.3}$$

Write $S(\eta)=S(\eta,\sigma)$ to emphasize the dependence of the score on σ. Then (12.3) suggests estimating η by $\widehat{\eta}$ solving

$$0=\sum\nolimits_i t_2\left(\overline{A}_{Zi}\right)S_i\left(\eta,\sigma_i\left(\widehat{\psi},\widehat{\theta}\right)\right)/W_i\ . \tag{12.4}$$

for some $t_2(\overline{a}_Z)$. The solution $\widehat{\eta}$ to (12.4) will be a RAL estimator of η if the models for *(i)* $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$, *(ii)* $\tau(y,\overline{a}_Z)$, and *(iii)* the density (12.1) are correctly specified. However, except in the special case described later, it is unlikely *(i)*-*(iii)* will all be correct, since a constraint must be satisfied. Specifically, as noted previously, by using d-separation on DAG G^*, we have

$$\underline{A}_{Zk} \coprod^* L_k \mid \overline{L}_{k-1}, \overline{A}_{k-1} \tag{12.5}$$

and further, by (10.9), on G^*, $\sigma \coprod^* \overline{A}_Z$. Also $\sigma \coprod^* A_{Pm} \mid \underline{A}_{Zm}, \overline{A}_{m-1}, \overline{L}_m$, since, by results in Robins (1997) for over-all effects, $H \coprod^* A_{Pm} \mid \underline{A}_{Zm}, \overline{A}_{m-1}, \overline{L}_m$. These independencies imply that

$$\prod_{m=0}^{k} f^*\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}\right) = \int \prod_{m=0}^{k} f^*\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}, \underline{A}_{Zm}, \sigma\right) f^*(\sigma)\, d\sigma \tag{12.6}$$

for $k = 0, \ldots, K$.

If we specify a model (12.2) which imposes

$$\underline{A}_{Zm} \coprod^* L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}, \sigma \tag{12.7}$$

then (12.6) holds. However, (12.7) is a restriction on the allowable densities $f(v)$ for the data. If (12.7) is not true, it will be essentially impossible to specify a model (12.2) such that, for some value of η_2, *(i)* (12.7) is false and *(ii)* the constraint (12.6) holds. Hence when (12.7) is false, a parameterization in terms of *(i)* $F^*_\sigma(y) \equiv b(y,0,0)$, *(ii)* the densities (12.1), and the functions *(iii)* $\gamma\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_Z\right)$ and *(iv)* $\tau(y,\overline{a}_Z)$ which fails to impose (12.6) will suffer from the type of difficulties described in Remark A of Sec. 9. However, such a parameterization cannot be realized if (12.6) is imposed. Thus it becomes important to characterize (12.7) in some equivalent fashion with a clear causal interpretation, so that we can better judge when it is substantively reasonable to impose (12.7). We will show that (12.7) can be characterized as a particular restriction on the functional form of a standard SNDM as studied in Robins (1997) and Robins and Wasserman (1997).

To characterize this restriction, we recall the definition of a standard SNDM. Let $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ be the quantile-quantile function satisfying $b\left(y,\overline{\ell}_m,\overline{a}_m\right) = b\left[\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right),\overline{\ell}_m,\overline{a}_{m-1}\right]$ where $b\left(y,\overline{\ell}_m,\overline{a}_m\right) \equiv b\left(y,\overline{\ell}_m,\overline{a}_{Pm},\overline{a}_{Zm}\right)$ and $b\left(y,\overline{\ell}_m,\overline{a}_{Pk},\overline{a}_{Zm}\right) \equiv F_g\left(y \mid \overline{\ell}_m\right)$ for the regime g that has *(i)* A_p history $\overline{a}_{Pk}$ through t_k and zero thereafter and *(ii)* A_Z history $\overline{a}_{Zm}$ through t_m and zero thereafter. It follows that $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is *(a)* increasing and continuously differentiable in y and *(b)* $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right) = y$ if $a_m = 0$ (i.e., $a_{Pm} = a_{Zm} = 0$). Robins (1989, 1997) shows *(i)* $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right) \equiv y$ if and only if the "g"-null hypothesis (4.1) holds. Now define $\overset{\bullet}{H} \equiv \overset{\bullet}{H}\left(\overset{\bullet}{\gamma}\right) \equiv \overset{\bullet}{H}_0\left(\overset{\bullet}{\gamma}\right) \equiv \overset{\bullet}{h}\left(y,\overline{L}_K,\overline{A}_K\right)$ recursively by $\overset{\bullet}{H}_{K+1}\left(\overset{\bullet}{\gamma}\right) \equiv Y$ and $\overset{\bullet}{H}_k\left(\overset{\bullet}{\gamma}\right) \equiv \overset{\bullet}{h}_k\left(Y,\overline{L}_K,\overline{A}_K\right) \equiv \overset{\bullet}{\gamma}\left[\overset{\bullet}{H}_{k+1}\left(\overset{\bullet}{\gamma}\right),\overline{L}_k,\overline{A}_k\right]$. Robins (1989, 1997) shows that

$$\overset{\bullet}{H} \coprod A_m \mid \overline{L}_m, \overline{A}_{m-1}\ . \tag{12.8}$$

Definition: The distribution $F(v)$ follows a standard SNDM $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta\right)$ if $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right) = \overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta_0\right)$ and *(i)* $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta\right)$ satisfies *(a)-(b)* above and $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta\right) \equiv y$ if and only if $\delta = 0$, so $\delta_0 = 0$ represents the "g"-null hypothesis (4.1). An example would be

$$\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta\right) = y + \delta_1 a_{Zm} a_{Z(m-1)} + \delta_2 a_{Zm} \ell_m^* + \delta_3 a_{Zm} a_{Pm} \ell_m^* + \delta_4 a_{Zm} a_{P(m-1)} + \delta_5 a_{Pm} + \delta_6 a_{Pm} a_{Z(m-1)} + \delta_7 a_{Pm} \ell_m^* a_{Z(m-1)} \tag{12.9}$$

where $\delta = (\delta_1, \dots, \delta_7)'$ and ℓ_m^* is a known function of $\overline{\ell}_m$. We shall need the following.

Definition: We say that $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is A_P-direct-effect consistent if $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is not a function of $\overline{\ell}_m$ when $a_{Pm} = 0$.

Example: Putting $a_{Pm} = 0$ in (12.9), we are left with the non-zero terms

$$\delta_1 a_{Zm} a_{Z(m-1)} + \delta_2 a_{Zm} \ell_m^* + \delta_4 a_{Zm} a_{P(m-1)} \ .$$

We thus deduce that $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is A_P-direct-effect consistent if and only if the true value δ_{20} of δ_2 is zero. The importance of this definition is the following.

Theorem 12.1: If $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is A_P-direct-effect consistent, the direct-effect "g"-null hypothesis (4.3) holds $\Leftrightarrow \overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ does not depend on $\overline{a}_{Pm} \Leftrightarrow \overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right) = \overset{\bullet}{\gamma}\left(y,\overline{a}_{Zm}\right)$ does not depend on $\overline{\ell}_m$ or $\overline{a}_{Pm}$.

Example: Suppose $\delta_{20} = 0$ in model (12.9) so it is A_P-direct-effect consistent. Then the direct-effect "g"-null hypothesis holds $\Leftrightarrow \delta_{30} = \ldots = \delta_{70} = 0$. That is, $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m,\delta_0\right) = \delta_{10} a_{Zm} a_{Z(m-1)}$. Note if $\delta_{20} \neq 0$ in (12.9), the direct-effect "g"-null hypothesis can be false even if $\delta_{30} = \ldots = \delta_{70} = 0$ since, for example, A_{P0} might affect ℓ_m^* which then, in turn, affects Y via the interaction $a_{Zm}\ell_m^*$. Indeed, we developed direct-effect SNDMs model precisely because, in the presence of terms such as $\delta_2 a_{Zm} \ell_m^*$, we are unable to use standard SNDMs to test the direct-effect "g"-null hypothesis. See Robins and Wasserman (1997, Sec. 8.1) for further discussion. Before giving our main theorem, we give another condition equivalent to A_P-direct-effect consistency.

Definition: There is no $L - A_Z$ interaction if, for each m, the quantile-quantile function $\gamma_Z\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_{Zm}\right)$ satisfying $b\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_{Zm}\right) =$
$b\left\{\gamma_Z\left(y,\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_{Zm}\right),\overline{\ell}_m,\overline{a}_{P(m-1)},\overline{a}_{Z(m-1)}\right\}$ does not depend on $\overline{\ell}_m$.

We are now ready to state our main theorem of this section which gives alternative characterizations of (12.7).

Theorem 12.2: The following are equivalent:

(1) Eq. (12.7) holds.

(2) $\overset{\bullet}{\gamma}\left(y,\overline{\ell}_m,\overline{a}_m\right)$ is A_P-direct-effect consistent.

(3) There is no $L - A_Z$ interaction.

(4) $\sigma = \overset{\bullet}{H}$ w.p.1.

(5) $\sigma \coprod A_m \mid \overline{L}_m, \overline{A}_{m-1}$.

(6) $\sigma \coprod^* A_m \mid \overline{L}_m, \overline{A}_{m-1}$.

Corollary 12.1: Eq. (12.7) implies
$f^*\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}, \sigma\right) = f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}, \sigma\right)$ and $f(\sigma) = f^*(\sigma)$.

Theorem 12.2 indicates that weighting by W_m^{-1} is unnecessary when we impose (12.7). In particular, it follows from parts (4)-(5) of Theorem 12.1 that (ψ, θ) can be estimated jointly by standard g-estimation of a standard SNDM as in Robins (1992, 1997). Further, by part (4) it follows that, given estimators $\left(\widetilde{\psi}, \widetilde{\theta}\right)$, we can estimate the law $F_\sigma(y) = b(y, 0, 0)$ by $n^{-1} \sum_i I\left[\sigma_i\left(\widetilde{\psi}, \widetilde{\theta}\right) > y\right]$ as in Robins (1997). Finally, Corollary (12.1) indicates that, given (12.7), the model (12.2) can be fit by maximum likelihood without needing to reweight, since the model is also true for the density $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}, \sigma\right)$. Note, however, that if one wanted an estimate of ψ_0 that would be consistent even under misspecification of the model $\tau(y, \overline{a}_z; \theta)$, one would have to use the weighted g-estimator $\widehat{\psi}$ of Sec. 10, rather than the joint standard g-estimator of (ψ, θ) mentioned above.

The no $L - A_Z$ interaction characterization of (12.7) is probably the most easily interpreted characterization of (12.7) from a substantive point of view. In assessing the reasonableness of the no $L - A_Z$ interaction assumption, it is important to recognize that this absence of interaction is only on a particular scale, the quantile-quantile transformation scale as represented by the function $\gamma_Z\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right)$ in the definition of no $L - A_Z$ interaction. The no $L - A_Z$ interaction assumption can be empirically tested by specifying a standard SNDM and testing for no A_P-direct effect consistency. For example, given SNDM (12.9), we would test the hypothesis $\delta_{20} = 0$ using a standard g-test as in Robins (1992, 1997).

13. Discussion

Following Robins (1998, 1999), we could have used direct effect marginal structural models (MSMs) rather than direct effect SNMs. A direct effect marginal structural distribution model is a semiparametric model for $b(y, \overline{a}_P, \overline{a}_Z) = F_{\overline{a}_P, \overline{a}_Z}(y)$ depending on a finite dimensional parameter β_0 such that $\beta_0 = 0$ if and only if $F_{0, \overline{a}_Z}(y) = F_{\overline{a}_P, \overline{a}_Z}(y)$ for all $\overline{a}_P$. The direct effect g-null hypothesis (4.2) implies $\beta_0 = 0$ but the converse is false. An SNDM with no $L - A_Z$ interaction (*i.e.*, $\delta_{20} = 0$ apriori) is a direct effect marginal structural model, although the converse is false..

A direct effect marginal structural mean model is a semiparametric model for $b(\overline{a}_P, \overline{a}_Z) = E_{\overline{a}_P, \overline{a}_Z}(Y)$ depending on a finite dimensional parameter β_0 such that $\beta_0 = 0$ if and only if $b(\overline{a}_P, \overline{a}_Z) = b(0, \overline{a}_Z)$ for all $\overline{a}_P$. The direct effect g-null mean null hypothesis (6.2) implies $\beta_0 = 0$ but the converse is false. Robins (1998, 1999) further discusses the advantages and disadvantages of marginal structural versus structural nested models.

Appendix 1:

Proof of Theorems 8.1 and 8.4: We shall need the following lemma. Let ε be as defined in Eq. (9.2).

Lemma: For $j = -1, 0, \ldots, K$, and any function $t_j(\bullet)$

$$E\left[\varepsilon W_{j+1}^{-1} t_j\left(\underline{A}_{Z(j+1)}\right) \mid \overline{A}_j, \overline{L}_j\right] = 0 \tag{A.1}$$

$$E\left[b\left(\overline{A}_Z\right) W_{j+1}^{-1} t_j\left(\overline{A}_{Z(j+1)}\right) \mid \overline{A}_j, \overline{L}_j\right] = \iint b\left(\overline{A}_Z\right) t_j\left(\underline{A}_{Z(j+1)}\right) \prod_{k=j+1}^{K} d\mu\left(A_{Zk}\right) \tag{A.2}$$

$$E\left[\nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) W_{j+1}^{-1} t_j\left(\underline{A}_{Z(j+1)}\right) \mid \overline{A}_j, \overline{L}_j\right] = 0 \textit{ for } j < m \leq K \tag{A.3}$$

$$\begin{gathered} E\left[\nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) W_{j+1}^{-1} t_j\left(\underline{A}_{Z(j+1)}\right) \mid \overline{A}_j, \overline{L}_j\right] = \\ \iint \nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) t_j\left(\underline{A}_{Z(j+1)}\right) \prod_{k=j+1}^{K} d\mu\left(A_{Zk}\right), 0 \leq m < j \leq K \; . \end{gathered} \tag{A.4}$$

Proof of Lemma: Consider (A.3). By Fubini's theorem, the left hand side of (A.3) is

$$\begin{gathered} \int \prod_{k=j+1}^{K} d\mu\left(A_{Zk}\right) t_j\left(\underline{A}_{Z(j+1)}\right) \\ \left\{ \iint \prod_{k=j+1}^{m-1} dF\left(A_{Pk} \mid \overline{A}_{k-1}, A_{Zk}\right) dF\left(L_k \mid \overline{L}_{k-1}, \overline{A}_{k-1}\right) \right. \\ \left. \left[\int \nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) dF\left(L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}\right)\right] \right\} . \end{gathered}$$

The term in the square brackets is zero by (9.1). (A.1), (A.2), and (A.4) are established by similar calculations.

Proof of Theorem 8.1: $\Rightarrow$Since the right hand sides of (A.1)-(A.4) do not depend on A_j, the conclusion of the theorem follows from (9.2) and (9.4) when we set $m = j$.

$\Leftarrow$ Suppose $\gamma^*\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$ was a second function satisfying the premise of the theorem that differed from $\gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$ on a set of positive probability. Let m^* be the largest value of m such that $\Delta\left(L_m, \overline{A}_{Pm}, \overline{A}_Z\right) \equiv \gamma^*\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) - \gamma\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$ is a function of A_{Pm} with positive probability. By the assumption that $\gamma^*\left(\overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) = 0$ if $A_{Pm} = 0$, we are guaranteed that $m^* \geq 0$. It follows that the expectation in the theorem will be a function of A_{Pm^*} if and only if $\int d\mu\left(\underline{A}_{Z(m^*+1)}\right) t\left\{\underline{A}_{Z(m^*+1)}\right\} \Delta\left(\overline{L}_{m^*}, \overline{A}_{Pm^*}, \overline{A}_Z\right)$ does not depend on A_{Pm^*}. But this clearly is false for a suitable choice of the function $t(\bullet)$.

Proof of Theorem 8.4: $\Rightarrow$ We must show that

$$E\left[\sigma(\gamma, b)\, t\left(\overline{A}_Z\right)/W\right] \equiv E\left[\sigma(\gamma, b)\, t\left(\overline{A}_Z\right)/W_0 \mid \overline{L}_{-1}, \overline{A}_{-1}\right] = 0 \; . \tag{A.5}$$

But by (9.4) and (9.2),

$$\sigma(\gamma, b) \equiv H(\gamma) - b\left(\overline{A}_Z\right) = \varepsilon + \sum_{m=0}^{K} \nu\left(\overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) \; .$$

(A.5) now follows from (A.1) and (A.3).

$\Leftarrow$ This can be proved analogously to Theorem 8.1.

Proof of Theorem 10.2-10.4:

We shall require some preliminary results and definitions. First note for regime $g = \overline{a}$

$$f_{\overline{a}}\left(y \mid \overline{\ell}_m\right) = \int f_{\overline{a}}\left(y \mid \overline{\ell}_{m+1}\right) dF\left(\ell_{m+1} \mid \overline{L}_m = \overline{\ell}_m, \overline{A}_m = \overline{a}_m\right) \ . \tag{A.6}$$

Now let

$$f\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right) \equiv \partial b\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)/\partial y \tag{A.7}$$

denote $f_g\left(y \mid \overline{\ell}_m\right)$ for $g = \overline{a}, \overline{a} = \{(\overline{a}_{Pm}, 0), \overline{a}_Z\}$. Also note that, from (A.7) and the definition of $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$, we have

$$f\left(u, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right) = f\left[\gamma^{-1}\left(u, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right), \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right] \partial\gamma^{-1}\left(u, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)/\partial u. \tag{A.8}$$

Proof of Theorem 10.2: $\Rightarrow$. We will show by induction that, whenever the expectation is finite,

$$E\left[t\left(\overline{A}_Z, H_m, \overline{L}_m, \overline{A}_{P(m-1)}\right)/W_{m+1} \mid \overline{A}_m, \overline{L}_m\right] = \tag{A.9}$$
$$\int \left\{\int t\left(\overline{A}_Z, u, \overline{L}_m, \overline{A}_{P(m-1)}\right) f\left(u, \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) du\right\} d\mu\left(\underline{A}_{Z(m+1)}\right),$$

which is not a function of A_{Pm}. This proves the theorem, since $H \equiv H(\gamma)$ is a deterministic function of $H_m, \overline{A}_Z, \overline{L}_{m-1}, \overline{A}_{P(m-1)}$.

Case 1: $m = K$: (A.9) holds since *(i)* $H_K = \gamma\left(Y, \overline{\ell}_K, \overline{a}_{PK}, \overline{A}_Z\right)$ has density $f\left(u, \overline{L}_K, \overline{A}_{P(K-1)}, \overline{A}_Z\right)$ given $\left(\overline{A}_K, \overline{L}_K\right)$, and *(ii)* $W_{K+1} \equiv 1$ and $A_{Z(K+1)} \equiv 0$ w.p.1. by convention.

Case 2: Assume (A.9) is true with $m+1$ replacing m. We show it as true for m. Now the LHS of (A.9) is

$$E\left\{q\left(\overline{A}_{m+1}, \overline{L}_{m+1}\right)\left\{f\left(A_{Z(m+1)} \mid \overline{A}_m, \overline{L}_m\right)\right\}^{-1} \mid \overline{A}_m, \overline{L}_m\right\} \tag{A.10}$$

with

$$q\left(\overline{A}_{m+1}, \overline{L}_{m+1}\right) = E\left[W_{m+2}^{-1} t\left\{\overline{A}_Z, \gamma\left(H_{m+1}, \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right), \overline{L}_m, \overline{A}_{P(m-1)}\right\} \mid \overline{A}_{m+1}, \overline{L}_{m+1}\right] \tag{A.11}$$

where we have used the definition of H_m in terms of H_{m+1}. Now, by the induction hypothesis, $q\left(\overline{A}_{m+1}, \overline{L}_{m+1}\right) = \int d\mu\left(\underline{A}_{Z(m+2)}\right)\left\{\int t\left[\overline{A}_Z, u, \overline{L}_m, \overline{A}_{P(m-1)}\right] f\left(h, \overline{L}_{m+1}, \overline{A}_{Pm}, \overline{A}_Z\right) dh\right\}$ with $u \equiv \gamma\left(h, \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right)$, which is not a function of $A_{P(m+1)}$. Hence, (A.10) equals

$$\int d\mu\left(\underline{A}_{Z(m+1)}\right)\left\{t\left[\overline{A}_Z, u, \overline{L}_m, \overline{A}_{P(m-1)}\right]\left[\int f\left(h, \overline{L}_{m+1}, \overline{A}_{Pm}, \overline{A}_Z\right) dF\left(L_{m+1} \mid \overline{L}_m, \overline{A}_m\right)\right] dh\right\} =$$
$$\int d\mu\left(\underline{A}_{Z(m+1)}\right)\left\{\int t\left[\overline{A}_Z, u, \overline{L}_m, \overline{A}_{P(m-1)}\right] f\left(h, \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) dh\right\}, \text{ by (A.6).}$$

We now change the variable of integration from h to u in this last expression to obtain

$$\int d\mu\left(\underline{A}_{Z(m+1)}\right)\{\int t\left(\overline{A}_Z,u,\overline{L}_m,\overline{A}_{P(m-1)}\right)f\left[\gamma^{-1}\left(u,\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right),\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right]$$
$$\left[\partial\gamma^{-1}\left(u,\overline{L}_m,\overline{A}_{Pm},\overline{A}_Z\right)/\partial u\right]du\},$$

which is the RHS of (A.9) by (A.8).

$\Leftarrow$ The proof is analogous to that of Theorem 8.1.

Proof of 10.3: $\Rightarrow$: It is straightforward to show part *(i)* of the Theorem implies parts *(ii)* and *(iii)* if we can show

$$E\left[t\left(\overline{A}_Z,\sigma\right)/W\right]\equiv E\left[t\left\{\overline{A}_Z,\tau\left(H,\overline{A}_Z\right)\right\}/W\right]=\int d\mu\left(\overline{A}_Z\right)\left\{\int t\left(\overline{A}_Z,h\right)f\left(h,0,0\right)dh\right\} \tag{A.12}$$

with $f(h,0,0)=\partial b(h,0,0)/\partial h$. Now the LHS of (A.12) is

$$E\left[E\left\{t\left[\overline{A}_Z,\tau\left(H,\overline{A}_Z\right)\right]/W_1\mid A_0,L_0\right\}\left\{f\left[A_{Z0}\mid L_0\right]\right\}^{-1}\right]$$
$$=E\left[\left\{\int d\mu\left(\underline{A}_{Z1}\right)\int t\left[\overline{A}_Z,\tau\left(u,\overline{A}_Z\right)\right]f\left[u,L_0,0,\overline{A}_Z\right]du\right\}\left\{f\left[A_{Z0}\mid L_0\right]\right\}^{-1}\right]$$

by (A.9),

$$=\int d\mu\left(\overline{A}_Z\right)\left[\int t\left[\overline{A}_Z,\tau\left(u,\overline{A}_Z\right)\right]\left\{\int f\left(u,L_0,0,\overline{A}_Z\right)dF\left(L_0\right)\right\}du\right]$$
$$=\int d\mu\left(\overline{A}_Z\right)\left[\int t\left[\overline{A}_Z,\tau\left(u,\overline{A}_Z\right)\right]f\left(u,0,\overline{A}_Z\right)d\mu\right] \tag{A.13}$$

with $f\left(u,0,\overline{A}_Z\right)=\partial b\left(u,0,\overline{A}_Z\right)/\partial u$. But, by the change of variables $h=\tau\left(u,0,\overline{A}_Z\right)$, (A.13) equals

$$\int d\mu\left(\overline{A}_Z\right)\left[\int t\left(\overline{A}_Z,h\right)f\left\{\tau^{-1}\left(h,\overline{A}_Z\right),0,\overline{A}_Z\right\}\left\{\partial\tau^{-1}\left(h,\overline{A}_Z\right)/\partial h\right\}dh\right]=$$
$$\int d\mu\left(\overline{A}_Z\right)\left[\int t\left(\overline{A}_Z,h\right)f\left(h,0,0\right)dh\right]$$

by definition of $\tau\left(h,\overline{A}_Z\right)$. The other parts of the theorem we leave for the reader.

Proof of Theorem 10.4: Define $t\left(\overline{A}_Z,\sigma\right)\equiv t\left(\overline{A}_Z\right)\left[I\left(\sigma>y\right)-b\left(y,0,0\right)\right]$. Then, by (A.12), we have $E\left[W^{-1}t\left(\overline{A}_Z,\sigma\right)\right]=\int d\mu\left(\overline{A}_Z\right)t\left(\overline{A}_Z\right)\left[\int\left\{I\left[h>y\right]-b\left(y,0,0\right)\right\}f\left(h,0,0\right)dh\right]=0$, proving Theorem 10.4.

Appendix 2:

Proof of Theorem 11.1: First note $pr\left[\mathcal{H}_K>y\mid\overline{L},\overline{A}\right]\equiv$
$pr\left[Y>\gamma^{-1}\left(y,\overline{L}_K,\overline{A}_{PK},\overline{A}_Z\right)\mid\overline{L},\overline{A}\right]=b\left[\gamma^{-1}\left(y,\overline{L}_K,\overline{A}_{PK},\overline{A}_Z\right),\overline{L}_K,\overline{A}_{PK},\overline{A}_Z\right]=$
$b\left[y,\overline{L}_K,\overline{A}_{P(K-1)},\overline{A}_Z\right]$ by definition of the function γ. The proof is now completed by induction using the following two lemmas.

Lemma 1: If (11.3) is true, then (11.4) is true.

Proof: $pr\left[\mathcal{M}_m > y \mid \overline{L}, \overline{A}\right] \equiv pr\left[\mathcal{H}_m > \nu^{-1}\left(y, \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right) \mid \overline{L}, \overline{A}\right] =$ $b\left\{\nu^{-1}\left(y, \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right), \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right\} = b\left(y, \overline{L}_{m-1}, \overline{A}_{P(m-1)}, \overline{A}_Z\right)$ by definition of ν.

Lemma 2: If (11.4) is true with $m+1$ replacing m, then (11.3) is true.

Proof: $pr\left[\mathcal{H}_m > y \mid \overline{L}, \overline{A}\right] \equiv pr\left[\mathcal{M}_{m+1} > \gamma^{-1}\left(y, \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right) \mid \overline{L}, \overline{A}\right] =$ $b\left[\gamma^{-1}\left(y, \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right), \overline{L}_m, \overline{A}_{Pm}, \overline{A}_Z\right] = b\left(y, \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z\right)$ by definition of the function γ.

Appendix 3:

Throughout, we use the convention that for any function $q(\cdot)$ with argument $\overline{a}_Z = \overline{a}_{ZK}$, $q(\overline{a}_{Zm})$ is $q(\overline{a}_Z)$ with $\overline{a}_Z = (\overline{a}_{Zm}, 0)$, and $q(0)$ has $\overline{a}_Z \equiv 0$. Analogous remarks hold for functions of $\overline{a}_P$.

Proof of Theorem 12.1: We will show (1)⇒(6)⇒(5)⇒(4)⇒(2)⇔(3), (3)⇒(4), and (4)⇒(1).

(1)⇒(6): We have noted in Section 12 that, whether or not (12.7) is imposed, $\sigma \coprod^* \overline{A}_Z$ and $A_{Pm} \coprod^* \sigma \mid \overline{L}_m, \overline{A}_{P(m-1)}, \overline{A}_Z$. Combining these restrictions with (12.7), we obtain the conditional independencies in Part (6).

(6)⇒(5): It is sufficient to show that

$$E\left[q(\sigma) \mid \overline{A}_m, \overline{L}_m\right] = E^*\left[q(\sigma) \mid \overline{L}_m, \overline{A}_m\right] \tag{A3.1}$$

for any function $q(\bullet)$.

Proof by Induction: In the proof we use the identity

$$E^*\left[q(\sigma) \mid \overline{L}_m, \overline{A}_m\right] = E\left[q(\sigma)\, t_2\left(\underline{A}_{Z(m+1)} \mid \overline{A}_{Zm}\right) / W_{m+1} \mid \overline{L}_m, \overline{A}_m\right] \tag{A3.2}$$

where where $t_2\left(\underline{A}_{Z(m+1)} \mid \overline{A}_{Zm}\right)$ is the conditional density of $\underline{A}_{Z(m+1)}$ given $\overline{A}_{Zm}$ under $f^*(v)$.

Case 1: $m = K$: $E\left[q(\sigma) \mid \overline{A}_K, \overline{L}_K\right] = E^*\left[q(\sigma) \mid \overline{A}_K, \overline{L}_K\right]$ by (A3.2).

Case 2: We assume (A3.1) is true for $m+1$ and will prove it true for m.

$E\left[q(\sigma) \mid \overline{A}_m, \overline{L}_m\right] = E\left[E\left\{q(\sigma) \mid \overline{A}_{m+1}, \overline{L}_{m+1}\right\} \mid \overline{A}_m, \overline{L}_m\right] = E\left[E^*\left\{q(\sigma) \mid \overline{A}_m, \overline{L}_{m+1}\right\} \mid \overline{A}_m, \overline{L}_m\right] =$

$E^*\left[W_{m+1}\left\{t_2\left(\underline{A}_{Z(m+1)} \mid \overline{A}_{Zm}\right)\right\}^{-1} E^*\left\{q(\sigma) \mid \overline{A}_m, \overline{L}_{m+1}\right\} \mid \overline{A}_m, \overline{L}_m\right] =$ $E^*\left[E^*\left\{q(\sigma) \mid \overline{A}_m, \overline{L}_{m+1}\right\} \mid \overline{A}_m, \overline{L}_m\right] = E^*\left[q(\sigma) \mid \overline{A}_m, \overline{L}_m\right]$, where the second inequality is by the induction hypothesis and $\left(\sigma, \underline{A}_{Z(m+1)}\right) \coprod^* A_{Pm} \mid \overline{L}_m, \overline{A}_{m-1}, A_{Zm}$, the third by (A3.2), and the fourth by integration under $f^*(v)$.

(5)⇒(4): We know from Robins (1989, 1995, 1997) and Robins and Wasserman (1997) that, given $f(v)$ there are unique densities *(i)* $f_{\overset{\bullet}{H}}(h)$, *(ii)* $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}, h\right), f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$ and unique functions $\overset{\bullet}{\gamma}\left(y, \overline{\ell}_m, \overline{a}_m\right)$ which have a continuous positive derivative with respect to y and satisfy $\overset{\bullet}{\gamma}\left(y, \overline{\ell}_m, \overline{a}_m\right) = 0$ if $a_m = 0$ such that

$$f(V) = \left\{\partial \overset{\bullet}{H} / \partial Y\right\} f\left(\overset{\bullet}{H}\right) \prod_{m=0}^{K} f\left(L_m \mid \overline{L}_{m-1} \overline{A}_{m-1}, \overset{\bullet}{H}\right) f\left(A_m \mid \overline{A}_{m-1}, \overline{L}_m\right) \tag{A3.3}$$

where $\overset{\bullet}{H} = \overset{\bullet}{h}\left(Y, \overline{L}_K, \overline{A}_K\right)$ is defined in terms of the $\overset{\bullet}{\gamma}\left(y, \overline{\ell}_m, \overline{a}_m\right)$ as in Sec. 12. Hence, in view of (5), (4) follows if we can show that $\sigma \equiv \tau\left[h\left(Y, \overline{L}_K, \overline{A}_K\right), \overline{A}_Z\right]$ is equal to some function, say $h^*\left(Y, \overline{L}_K, \overline{A}_K\right)$, that is *(a)* recursively defined in terms of functions $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right)$ just as the function $\overset{\bullet}{h}$ was recursively defined in terms of the functions $\overset{\bullet}{\gamma}$ and *(b)* $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right)$ has a continuous positive derivative with respect to y and satisfies $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right) = 0$ if $a_m = 0$. To do so, let $h^{-1}\left(y, \overline{\ell}_K, \overline{a}_{Pm}, \overline{a}_{Zk}\right)$ be shorthand for $h^{-1}\left(y, \overline{\ell}_K, \left(\overline{a}_{Pm}, 0\right), \left(\overline{a}_{Zk}, 0\right)\right)$ and let $\tau^{-1}\left(y, \overline{a}_{Zm}\right)$ be shorthand for $\tau^{-1}\left(y, \left(\overline{a}_{Zm}, 0\right)\right)$. Note $h^{-1}\left(y, \overline{\ell}_K, \overline{a}_{Pm}, \overline{a}_{Zm}\right) \equiv h^{-1}\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_{Zm}\right)$ does not depend on $\underline{\ell}_{m+1}$, as is easily shown from its definition in terms of the $\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$. Now define $\gamma^{*-1}\left(y, \ell_0, a_0\right) = h^{-1}\left\{\tau^{-1}\left(y, a_{Z0}\right), \ell_0, a_{P0}, a_{Z0}\right\}$ and define $\gamma^{*-1}\left(y, \overline{\ell}_m, \overline{a}_m\right)$ recursively by $\gamma^{*-1}\left(y, \overline{\ell}_m, \overline{a}_m\right) = h^{-1}\left\{\tau^{-1}\left[\gamma^*\left(y, \overline{\ell}_{m-1}, \overline{a}_{m-1}\right), \overline{a}_{Zm}\right], \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_{Zm}\right\}$, where $\gamma^*\left(y, \bullet\right)$ is the inverse function to $\gamma^{*-1}\left(y, \bullet\right)$. It is then straightforward to check that $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right)$ has a continuous positive derivative with respect to y and satisfies $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right) = 0$ if $a_m = 0$. Furthermore, one can check that $h^*\left(y, \overline{\ell}_K, \overline{a}_K\right)$ can be recursively obtained from the $\gamma^*\left(y, \overline{\ell}_m, \overline{a}_m\right)$ in the appropriate manner. This completes the proof.

(4)⇒(2): It follows from the above proof that (5)⇒(4) that when $a_{Pm} = 0$, $\overset{\bullet}{\gamma}^{-1}\left(y, \ell_0, a_0\right) = \tau^{-1}\left(y, a_{Z0}\right)$ does not depend on ℓ_0, and that $\overset{\bullet}{\gamma}^{-1}\left(y, \overline{\ell}_m, \overline{a}_m\right) = \tau^{-1}\left[\overset{\bullet}{\gamma}\left(y, \overline{\ell}_{m-1}, \overline{a}_{m-1}\right), \overline{a}_{Zm}\right]$ does not depend on ℓ_m. Combining these results, we conclude that when $a_{Pm} = 0$, $\overset{\bullet}{\gamma}^{-1}\left(y, \overline{\ell}_m, \overline{a}_m\right)$ does not depend on $\overline{\ell}_m$, concluding the proof.

(2)⇔(3): From their definitions, $\overset{\bullet}{\gamma}\left(y, \overline{\ell}_m, \overline{a}_m\right) = \gamma_Z\left\{\gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_{Zm}\right), \ell_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right\}$ so, when $a_{Pm} = 0$, $\overset{\bullet}{\gamma}\left(y, \overline{\ell}_m, \overline{a}_m\right) = \gamma_Z\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right)$.

(3)⇒(4): We need to show that

$$\overset{\bullet}{h}\left(y, \overline{\ell}, \overline{a}\right) = \tau\left[h\left(y, \overline{\ell}, \overline{a}\right), \overline{a}_Z\right] \tag{A3.4}$$

under the no $L - A_Z$ interaction assumption

$$\gamma_Z\left(y, \overline{\ell}_k, \overline{a}_{P(k-1)}, \overline{a}_{Zk}\right) = \gamma_Z\left(y, \overline{a}_{P(k-1)}, \overline{a}_{Zk}\right) \ . \tag{A3.5}$$

We shall require two preliminary lemmas.

Lemma A3.1: (A3.5) implies, for $-1 \le k \le m$,

$$b\left(y, \overline{\ell}_k, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right) = b\left\{u, \overline{\ell}_k, \overline{a}_{P(m-1)}, \overline{a}_{Z(m-1)}\right\} \tag{A3.6}$$

where $u \equiv \gamma_Z\left(y, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right)$.

Proof of Lemma A3.1: By definition of γ_Z, (A3.6) holds for $k = m$. Hence it suffices to show that if (A3.6) is true for k, it is true for $k-1$. Now $b\left(y, \overline{\ell}_{k-1}, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right) \equiv \int b\left(y, \overline{\ell}_k, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right) dF\left(\ell_k \mid \overline{\ell}_{k-1}, \overline{a}_{k-1}\right) = \int b\left(u, \overline{\ell}_k, \overline{a}_{P(m-1)}, \overline{a}_{Z(m-1)}\right) dF\left(\ell_k \mid \overline{\ell}_{k-1}, \overline{a}_{k-1}\right) \equiv$

$b\left(u, \overline{\ell}_{k-1}, \overline{a}_{P(m-1)}, \overline{a}_{Z(m-1)}\right)$, where the second to last equality is by the induction hypothesis.

Lemma A3.2: Eq. (A3.5) implies

$$\gamma_Z\left(u, 0, \overline{a}_{Zm}\right) = \gamma\left(u^*, \ell_0, a_{P0}, \overline{a}_{Z(m-1)}\right) \tag{A3.7}$$

with $u \equiv \gamma\left(y, \ell_0, a_{P0}, \overline{a}_{Zm}\right)$ and $u^* \equiv \gamma_Z\left(y, a_{P0}, \overline{a}_{Zm}\right)$.

Proof: $b\left(y, \ell_0, a_{P0}, \overline{a}_{Zm}\right) = b\left[u^*, \ell_0, a_{P0}, \overline{a}_{Z(m-1)}\right] = b\left[\gamma\left(u^*, \ell_0, a_{P0}, \overline{a}_{Z(m-1)}\right), \ell_0, 0, \overline{a}_{Z(m-1)}\right]$ where the first equality is by Lemma A3.1 with $\overline{a}_{P(m-1)} = \left(a_{P0}, 0\right)$ and the last equality is by definition of the direct effect blip function γ. However, $b\left(y, \ell_0, a_{P0}, \overline{a}_{Zm}\right) = b\left[u, \ell_0, 0, \overline{a}_{Zm}\right] = b\left[\gamma_Z\left(u, 0, \overline{a}_{Zm}\right), \ell_0, 0, \overline{a}_{Z(m-1)}\right]$. Hence, since $b\left(y, \ell_0, 0, \overline{a}_{Z(m-1)}\right)$ is monotone increasing in y, the lemma is proved.

Proof that (A3.5)⇒(A3.4): We shall use induction on K.

Case 1: $K=1$: Note $\overset{\bullet}{h}\left(s, \overline{\ell}_1, \overline{a}_1\right) = \gamma_Z\left(x^*, \ell_0, 0, a_{Z0}\right)$ with $x^* = \gamma\left[u^*, \ell_0, a_{P0}, a_{Z0}\right]$ where $u^* = \gamma_Z\left(y, a_{P0}, a_{Z1}\right)$ and $y \equiv \gamma\left(s, \overline{\ell}_1, \overline{a}_{P1}, \overline{a}_{Z1}\right)$. Similarly, $\tau\left[h\left(s, \overline{\ell}_1, \overline{a}_1\right), \overline{a}_{Z1}\right] = \gamma_Z\left[x, \ell_0, 0, a_{Z0}\right]$ where $x = \gamma_Z\left[u, 0, \overline{a}_{Z1}\right]$, $u = \gamma\left(y, \ell_0, a_{P0}, \overline{a}_{Z1}\right)$, and y is as defined above. Hence we must show that $x = x^*$, which is just (A3.7) with $m=1$.

Case 2: Assume the theorem is true for $K = K^*$. We shall show it is true for $K = K^*+1$. Given $\left(\overline{\ell}, \overline{a}\right) \equiv \left(\overline{\ell}_{K^*+1}, \overline{a}_{K^*+1}\right)$, let $\gamma_m(y) \equiv \gamma\left(y, \overline{\ell}_m, \overline{a}_{Pm}, \overline{a}_Z\right)$ and $\gamma_{Zm}(y) \equiv \gamma_Z\left(y, 0, \overline{a}_{Zm}\right)$. Then

$$\tau\left[h\left(y, \overline{\ell}, \overline{a}\right), \overline{a}_Z\right] = \gamma_{Z0} \circ \ldots \circ \gamma_{ZK^*} \circ \gamma_{Z(K^*+1)} \circ \gamma_0 \circ \ldots \circ \gamma_{K^*} \circ \gamma_{K^*+1}(y) \ . \tag{A3.8}$$

We now apply (A3.7) successively for $m = K^*+1, K^*, \ldots, 1$ to obtain

$$\gamma_{Z0} \circ \ldots \circ \gamma_{Z(K^*+1)} \circ \gamma_0(y) = \gamma_{Z0} \circ \gamma_{P0} \circ \gamma_{Z01} \circ \gamma_{Z02} \circ \ldots \circ \gamma_{Z0(K^*+1)}(y) \tag{A3.9}$$

where $\gamma_{P0}(y) \equiv \gamma\left(y, \ell_0, a_{P0}, a_{Z0}\right)$ and $\gamma_{Z0m}(y) \equiv \gamma_Z\left(y, a_{P0}, \overline{a}_{Zm}\right)$. Now since by their definitions, $\gamma_{Z0} \circ \gamma_{P0}(y) = \overset{\bullet}{\gamma}\left(y, \ell_0, a_0\right)$ and $\overset{\bullet}{h}\left(y, \overline{\ell}, \overline{a}\right) = \overset{\bullet}{\gamma}\left[\overset{\bullet}{h}_1\left(y, \overline{\ell}, \overline{a}\right), \ell_0, a_0\right]$, the theorem is proved if we can show that

$$h_1\left(y, \overline{\ell}, \overline{a}\right) = \gamma_{Z01} \circ \ldots \circ \gamma_{Z0(K^*+1)} \circ \gamma_1 \circ \ldots \circ \gamma_{K^*+1}(y) \ . \tag{A3.10}$$

However, treating $\left(\ell_0, a_0\right)$ as fixed, we find that (A3.10) is true by the induction hypothesis since the RHS of (A3.10) is $\tau \circ h$ for a study with $K = \left(K^*+1\right) - 1 = K^*$.

(4)⇒(1): This implication follows from a probability calculation on G^* using Part (6).

Proof of Corollary 12.1: Parts (5) and (6) of Theorem 12.1 imply
$f^*(v) = \{\partial\sigma/\partial y\} f^*(\sigma) \prod_{m=0}^{K} f^*\left(\ell_m \mid \overline{L}_{m-1}, \overline{a}_{m-1}, \sigma\right) f^*\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$ and
$f(v) = \{\partial\sigma/\partial y\} f(\sigma) \prod_{m=0}^{K} f\left(\ell_m \mid \overline{L}_{m-1}, \overline{a}_{m-1}, \sigma\right) f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$. The corollary is now proved by noting that by the definition of $f^*(v)$, $f^*(v) / \prod_{m=0}^{K} f^*\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right) =$
$f(v) / \prod_{m=0}^{K} f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}\right)$.

Final Remark: It is interesting to note that in the context of a SNMM, the assumption of no $L - A_Z$ mean interaction, i.e., $\gamma_Z\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right) \equiv b\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right) -$
$b\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Z(m-1)}\right)$ does not depend on $\overline{\ell}_m$ is true if and only if $\nu\left(\overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ does not

depend on $\underline{a}_{Zm}$. However, it is not true, in the setting of a SNDM, that the assumption (A3.5) of no $L - A_Z$ interaction either implies or is implied by the assumption that $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ does not depend on $\underline{a}_{Zm}$ [unless the null hypothesis $\tau\left(y, \overline{a}_Z\right) \equiv y$ is also true]. This reflects the fact that the scale in which the $L - A_Z$ interaction is measured by the function $\nu\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_Z\right)$ differs from the scale on which it is measured by $\gamma_Z\left(y, \overline{\ell}_m, \overline{a}_{P(m-1)}, \overline{a}_{Zm}\right)$.

REFERENCES

Gill R and Robins JM. (1997). "From g-computation to G-computation." (unpublished manuscript).

Pearl, J. (1995), "Causal diagrams for empirical research." *Biometrika,* 82(4), 669-690.

Pearl, J. and Robins, J.M. (1995). "Probabilistic evaluation of sequential plans from causal models with hidden variables." From: **Uncertainty in Artificial Intelligence:** Proceedings of the Eleventh Conference on Artificial Intelligence, August 18-20, 1995, McGill University, Montreal, Quebec, Canada. San Francisco, CA: Morgan Kaufmann. pp. 444-453.

Pearl, J. and Verma, T. (1991). "A Theory of Inferred Causation." In: **Principles of Knowledge, Representation and Reasoning: Proceedings of the Second International Conference**. Eds. J.A. Allen, R. Fikes, and E. Sandewall. 441-452.

Robins, J.M. (1986), "A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect." *Mathematical Modelling*, 7, 1393-1512.

Robins, J.M. (1987), "Addendum to 'A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect'." *Computers and Mathematics with Applications*, 14(9-12), 923-945.

Robins, J.M. (1989), "The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies." In: **Health Service Research Methodology: A Focus on AIDS.** Eds. Sechrest, L., Freeman, H., Mulley, A., Washington, D.C.: U.S. Public Health Service, National Center for Health Services Research, pp. 113-159.

Robins, J.M. (1992), "Estimation of the time-dependent accelerated failure time model in the presence of confounding factors." *Biometrika*, 79(2), 321-334.

Robins, J.M., and Rotnitzky, A. (1992). "Recovery of information and adjustment for dependent censoring using surrogate markers." **AIDS Epidemiology - Methodological Issues**. Eds: Jewell N., Dietz K., Farewell V. Boston, MA: Birkhäuser. pp. 297-331.

Robins, J.M. (1993), "Analytic methods for estimating HIV-treatment and cofactor effects." In: **Methodological Issues in AIDS Mental Health Research**, eds. Ostrow, D.G., and Kessler, R.C., NY: Plenum Press, 213-290.

Robins, J.M. (1994), "Correcting for non-compliance in randomized trials using structural nested mean models." *Communications in Statistics,* 23(8), 2379-2412.

Robins, J.M. (1995). "Discussion of 'Causal Diagrams for empirical research' by J. Pearl." *Biometrika*, 82(4), 695-698.

Robins J.M. (1997). "Causal inference from complex longitudinal data." In: **Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)**, M. Berkane, Editor. NY: Springer Verlag, 69-117.

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.), San Francisco:Morgan Kaufmann, pp. 409-418.

Robins J.M. (1998). Marginal structural models. *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science,* pp. *1-10.*

Robins J.M. (1999). Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In: **Statistical Models in Epidemiology**: **The Environment and Clinical Trials.** Halloran, E. and Berry, D., eds. NY: Springer-Verlag (to appear).

Robins JM, Scharfstein D, Rotnitzky A. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: **Statistical Models in Epidemiology**: **The Environment and Clinical Trials.** Halloran, E. and Berry, D., eds. NY: Springer-Verlag (to appear).

Robins, J.M. (2000). "Estimating the Causal Effect of a Time-varying Treatment on Survival using Structural Nested Failure Time Models." *Statistica Neederlandica* (to appear).

Rosenbaum, P.R. (1984), "Conditional permutation tests and the propensity score in observational studies." *Journal of the American Statistical Association,* 79, 565-574.

Spirtes, P., Glymour, C., and Scheines, R. (1993). **Causation, Prediction, and Search**. New York: Springer Verlag.

Verma, T. and Pearl, J. (1991). "Equivalence and synthesis of causal models." Technical Report R-150, Cognitive Systems Laboratory, University of California, Los Angeles.