University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2002 *Paper* 122

Construction of Counterfactuals and the G-computation Formula

Zhuo Yu* Mark J. van der Laan†

*Department of Statistics, University of California, Berkeley, zyu@stat.berkeley.edu
†Division of Biostatistics, School of Public Health, University of California, Berkeley

Construction of Counterfactuals and the G-computation Formula

Zhuo Yu and Mark J. van der Laan

Abstract

Robins' causal inference theory assumes existence of treatment specific counterfactual variables so that the observed data augmented by the counterfactual data will satisfy a consistency and a randomization assumption. Gill and Robins [2001] show that the consistency and randomization assumptions do not add any restrictions to the observed data distribution. In particular, they provide a construction of counterfactuals as a function of the observed data distribution. In this paper we provide a construction of counterfactuals as a function of the observed data itself. Our construction provides a new statistical tool for estimation of counterfactual distributions. Robins [1987b] shows that the counterfactual distribution can be identified from the observed data distribution by a G-computation formula under an additional identifiability assumption. He proves this for discrete variables. Gill and Robins [2001] prove the G-computation formula for continuous variable under some additional conditions and modifications of the consistency and the randomization assumptions. We prove that if treatment is discrete, then Robins' G-computation formula holds under the original consistency, randomization assumptions and a generalized version of identifiability assumption.

1 Introduction

In a series of papers, Robins [1986, 1987, 1989, 1997] develops a systematic approach for causal inference in complex longitudinal studies. His approach depends on introducing counterfactual variables which link the variables observed in the real world to variables expressing what would happen should the subject have received treatment a. The keys to link the observed variables and counterfactual variables are the so called consistency, randomization and identifiability assumptions under which the counterfactual distributions can be recovered from observed data distribution if all variables are discrete (Robins [1987]).

Suppose a subject will visit a clinic at K fixed time points. At visit $k = 1, \ldots, K$, medical tests are done yielding some data L_k (when the doctor assigns a treatment A_k, this could be the quantity of a certain drug). The data $L_1, A_1 \ldots, L_{k-1}, A_{k-1}$ from earlier visits is available. Of interest is some response Y, to be thought of as representing the state of the subject after the complete treatment regime. Thus in time sequence the complete history of the subject results in the alternating sequence of covariates (or responses) and treatments

$$L_1, A_1, \ldots, L_K, A_K, Y \equiv L_{K+1}.$$

We assume without mention from now on that all the random variables in this paper are multivariate real valued and are defined on a given common probability space $(\Omega, \mathcal{F}, P)$. We write $\bar{L}_k = (L_1, \ldots, L_k)$, $\bar{A}_k = (A_1, \ldots, A_k)$. We abbreviate $\bar{L}_{K+1}$ and $\bar{A}_K$ to $\bar{L}$ and $\bar{A}$. A treatment regime or plan, denoted g, is a rule which specifies treatment at each time point, given the data available at that moment. In other words, it is a collection $(g_k)_{k=1}^K$ of functions g_k, the k'th defined on sequences of the first k covariate values, where $a_k = g_k(\bar{l}_k)$ is the treatment to be administered at the k'th visit given covariate values $\bar{l}_k = (l_1, \ldots, l_k)$ up till then. Define $\bar{g}_k(\bar{l}_k) = (g_1(l_1), g_2(l_1, l_2), \ldots, g_k(l_1, \ldots, l_k))$. We abbreviate $\bar{g}_K$ to $\bar{g}$. Let $\mathcal{G}$ denote the set of all treatment plans. A fundamental assumption for Robins causal inference methodology is the existence of treatment specific random variables $(Y^{\bar{g}}; g \in \mathcal{G})$ such that the following assumptions hold.

A1 (Consistency) There exists a subset $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ such that $Y(\omega) = Y^{\bar{g}}(\omega)$ on $\{\omega \in \Omega_0 : \bar{A}(\omega) = \bar{g}_K(\bar{L}_K(\omega))\}$,

A2 (Randomization) $A_k \perp (Y^{\bar{g}}; \bar{g} \in \mathcal{G}) | (\bar{A}_{k-1}, \bar{L}_k)$. That is, A_k is independent of $(Y^{\bar{g}}; \bar{g} \in \mathcal{G})$ given the observed data history $(\bar{A}_{k-1}, \bar{L}_k)$ before A_k.

and that the the following identifiability assumption holds:

A3 (Identifiability) For any $\bar{l}_k$ with $P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{l}_{k-1}), \bar{L}_k = \bar{l}_k) > 0$, it follows that $P(\bar{A}_k = \bar{g}_k(\bar{l}_k), \bar{L}_k = \bar{l}_k) > 0$.

Under these three assumptions Robins [1987] proves that if $\bar{L}$ and $\bar{A}$ are discrete, the distribution of $Y^{\bar{g}}$ is given by the G-computation formula:

$$P(Y^{\bar{g}} \in \cdot) = \sum_{l_1} \dots \sum_{l_K} P(Y \in \cdot | \bar{A}_K = \bar{g}_K(\bar{l}_K), \bar{L}_K = \bar{l}_K) \times \prod_{k=1}^{K} P(L_k = l_k | \bar{A}_{k-1} = \bar{g}_{k-1}(\bar{l}_{k-1}), \bar{L}_{k-1} = \bar{l}_{k-1}) \tag{1}$$

In this paper, as in Gill and Robins [2001], we are concerned with answering the following two basic questions regarding Robins' theory. Firstly, given observed variables, whether there exist underlying counterfactual variables $(Y^{\bar{g}}; \bar{g} \in \mathcal{G})$ which satisfy the assumptions A1 and A2. More precisely, given observed random variables $(Y, \bar{A}_K, \bar{L}_K)$ defined on a given probability space, can one construct counterfactual variables $(Y^{\bar{g}}; \bar{g} \in \mathcal{G})$ (possibly after augmenting the probability space with some independent uniform variables in order to have some extra source of randomness) on the same space satisfying these assumptions. According to Gill and Robins [2001], "if the answer is no, adopting his approach means making restrictive implicit assumptions-not very desirable. If however the answer is yes, his approach is neutral. One can freely use it in modelling and estimation, exploring the consequences (for the unobserved variables) of the model". Secondly, whether the counterfactual distribution of $Y^{\bar{g}}$ can be identified by the observed data distribution under the consistency, randomization and identifiability assumptions.

For the first question, given a distribution of observed variables, Gill and Robins [2001] construct counterfactual and factual variables which satisfy these assumptions with the factual variables having the given observed data distribution. Our construction is directly based on observed variables rather than their distribution. In other words, our construction directly maps observed variables and their distribution into counterfactual variables. The estimated counterfactuals can be used to estimate the counterfactual distributions, and, and in particular, parameters of the counterfactual distribution such as a mean or median.

For the second question, we show that if the treatment value is discrete, the counterfactual distribution can be recovered from the observed data distribution by the G-computation formula (1) with the conditional probabilities replaced by conditional distributions under A1, A2 and A3*, here A3* is a generalized version of A3. We shall note that in most of the applications the treatment value is discrete, so that it is of interest to study the correctness of the G-computation formula in this case. Gill and Robins [2001]'s proof of the G-computation formula depends on additional continuity assumptions regarding the joint law of the counterfactual variables and the factual variables. See Assumptions **C** and **Cg** in Gill and Robins [2001]. They also modify the original consistency and randomization assumptions in terms of conditional distributions. See Assumptions **A1*** and **A2*** in Gill and Robins [2001]. We shall point out that Gill and Robins

[2001] deal with a more general situation where the treatment value could also be continuous. Our result also shows that the G-computation formula does not depend on how one chooses the conditional distributions which are not unique given the observed data distribution. If the treatment value is continuous, we show in the discussion section with an example that A1, A2 and A3* are not enough to identify the counterfactual distribution. This was already shown by Gill and Robins [2001] with another type of example. In this case, one needs more restrictions on the counterfactual variables. However, the continuous case is not within the scope of the current paper.

Organization of the paper. Section 2 provides our construction of counterfactuals satisfying A1 and A2. Under the additional assumption that the treatment value is discrete, we show in section 3 that the counterfactual distribution can be computed by the G-computation formula with conditional probabilities in (1) replaced by conditional distributions. In section 4, we provide statistical methods based on our construction of counterfactuals. Section 5 discusses some future directions.

2 Construction of counterfactuals

In this section, we state our construction of counterfactuals and provide the proof. Given observed variables defined on a given probability space, we can construct counterfactual variables defined on the same space which satisfy the consistency and randomization assumptions A1 and A2. This teaches us that the consistency and randomization assumptions are "free" assumptions in the sense that they do not add hidden restrictions on the observed data distribution.

Section 2.1 states the main theorem. In section 2.2 we provide some preliminaries on conditioning. 2.3 establishes some lemmas which we use to prove the main theorem. The proof of the theorem is given in section 2.4.

2.1 The main Theorem

Theorem 2.1. *(Construction of counterfactuals) Let* $O \equiv (\bar{A}_K, \bar{L}_K, Y \equiv L_{K+1})$ *be a random variable defined on a given probability space* $(\Omega, \mathcal{F}, P)$*. Let* $\mathcal{G}$ *be the set of all treatment plans. Suppose* $L_k = (L_{k,1}, \ldots, L_{k,p_k}) \in \mathbf{R}^{p_k}$*. Let* $\Delta_k \equiv (\Delta_{k,1}, \ldots, \Delta_{k,p_k})$*,* $k = 1, \ldots, K+1$*, where* $\Delta_{k,j}$ *are all uniformly independently distributed on* $(0,1]$ *and independent of* O*. Let* $\bar{L}_{k,j} \equiv (\bar{L}_{k-1}, L_{k,1}, \ldots, L_{k,j})$*, and* $\bar{L}_{k,0} \equiv \bar{L}_{k-1}$*. Similarly we define* $\bar{\Delta}_{k,j}$*. Let* $Q_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(dl_{k,j}; \bar{a}_{k-1}, \bar{l}_{k,j-1})$ *be a regular conditional distribution of* $L_{k,j}$ *given* $(\bar{A}_{k-1}, \bar{L}_{k,j-1})$*. Let* $F^{-1}(\cdot) \equiv inf_y\{F(y) \geq \cdot\}$ *for a univariate distribution function* F*. Set* $L_1^{\bar{g}} \equiv L_1$*. For each*

k and j, define $L^{\bar g}_{k,j}$ recursively by: for $k = 1, \ldots, K+1$, for $j = 1, \ldots, p_k$

$$L^{\bar g}_{k,j} \equiv Q^{-1}_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}\left(Q^{\Delta_{k,j}}_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}(L_{k,j}; \bar A_{k-1}, \bar L_{k,j-1}); \bar g_{k-1}(\bar L^{\bar g}_{k-1}), \bar L^{\bar g}_{k,j-1}\right),$$

where $\bar L^{\bar g}_{k,0} \equiv \bar L^{\bar g}_{k-1}$, $L^{\bar g}_{k} \equiv (L^{\bar g}_{k,1}, \ldots, L^{\bar g}_{k,p_k})$ and

$$\begin{aligned} Q^{\Delta_{k,j}}_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}(L_{k,j}; \bar A_{k-1}, \bar L_{k,j-1}) &\equiv \Delta_{k,j} Q_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}(L_{k,j}; \bar A_{k-1}, \bar L_{k,j-1}) \\ &\quad + (1-\Delta_{k,j}) Q_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}(L_{k,j}-; \bar A_{k-1}, \bar L_{k,j-1}). \end{aligned}$$

Then $(Y^{\bar g} \equiv L^{\bar g}_{K+1}; \bar g \in \mathcal{G})$ satisfies assumptions A1 and A2.

Remark: Note that the constructions depend on $\Delta_{k,j}$ if $L_{k,j}$ is discrete. So the *counterfactuals* satisfying A1 and A2 are not unique. However, we show in section 3 that if the treatment is discrete, then the counterfactual *distribution* is unique under the additional assumption A3*.

We next explain algorithmically how to construct the counterfactuals given the observed variables and part of the observed data likelihood:

$$\prod_{k=1}^{K+1} \prod_{j=1}^{p_k} Q_{L_{k,j}|\bar A_{k-1},\bar L_{k,j-1}}(L_{k,j}; \bar A_{k-1}, \bar L_{k,j-1}). \tag{2}$$

To simplify the explanation, we assume $p_l = 1$. We use lower case letters to denote the actual observed variables. At the beginning, generate $K+1$ independent $Unif(0,1]$ random variables $(\delta_1, \ldots, \delta_{K+1})$ and set the counterfactual variable of l_1, $l_1^{\bar g}$, to be l_1 itself. In order to generate the counterfactual variable of l_2, $l_2^{\bar g}$, first calculate the conditional distribution of L_2 given $(A_1 = a_1, L_1 = l_1)$ at l_2 and its left limit. Then take the weighted average of these two values with weights equal to δ_2 and $1-\delta_2$. Denote the result with u_1. Thereby, apply the inverse of the conditional distribution of L_2 given $(A_1 = g_1(l_1^{\bar g}), L_1 = l_1^{\bar g})$ to u_1. Set the result to be $l_2^{\bar g}$. Note that this procedure corresponds with applying a quantile-quantile function $X_2 = F_2^{-1}F_1^{\Delta}(X_1)$ as described in Lemma 2.4, where $X_1 = l_2, F_1 = Q_{L_2|A_1,L_1}(\cdot; a_1, l_1), F_2 = Q_{L_2|A_1,L_1}(\cdot; g_1(l_1^{\bar g}), l_1^{\bar g})$. Repeat this procedure till we obtain $y^{\bar g} \equiv l^{\bar g}_{K+1}$.

2.2 Preliminaries

Conditional distributions. Let Y, X denote two random variables. There exists a regular conditional distribution $Q_{Y|X}(dy; x)$ satisfying

(a) $B \to Q_{Y|X}(B; x)$ is a probability measure for any fixed x.

(b) $x \to Q_{Y|X}(B; x)$ is a measurable function for any Borel set B, and

$$E(h(X,Y)|X) = \int_y h(X,y) Q_{Y|X}(dy;X) \tag{3}$$

for any bounded measurable function $h(x,y)$. We note that Q is not unique. As a convention, we denote $Q_{Y|X}((-\infty, y] : x)$ with $Q_{Y|X}(y;x)$. It follows that $(y,x) \to Q_{Y|X}(y;x)$ is a measurable function as well.

Support of a distribution. A support point of the law of a random variable X is a point x such that $P(X \in B(x,\delta)) > 0$ for all $\delta > 0$, where $B(x,\delta)$ is the open ball around x with radius δ. We define the support of X, denoted $Supp(X)$ or $Supp(F_X)$, to be the set of all support points.

Conditional independence. Let X, Y, Z be random variables. We have that $X \perp Y | Z$ if for any bounded continuous functions h_1 and h_2,

$$E(h_1(X)h_2(Y)|Z) = E(h_1(X)|Z)E(h_2(Y)|Z).$$

Another way to verify the conditional independence $X \perp Y|Z$ is to show that for any bounded measurable function $h(X)$, $E(h(X)|Y,Z)$ is only a function of Z. In order to prove that $X \perp (Y_t, t \in T)|Z$ for an arbitrary index set T, we only need to prove that for any finite subset $T_0 \subset T$, $X \perp (Y_t, t \in T_0)|Z$.

2.3 Lemmas

In this section we establish some lemmas needed for the proof of Theorem 2.1. In the following we use capital letters to denote random variables and small letters to denote realizations of the random variables.

We start with introducing $Supp'$ of a univariate distribution function F. We say $x \in Supp'(F)$, if x satisfies that for any $\delta > 0$, $F((x-\delta, x]) > 0$. Let $D(F)$ be all the continuity points of F satisfying the following conditions: (1) for any $x' > x$, $F(x') > F(x)$ and (2) there exists $x' < x$, such that $F(x') = F(x)$. We have that $D(F)$ is countable and thus has zero mass under $F(dy)$. It is not hard to show $Supp'(F) = Supp(F) \backslash D(F)$. Therefore,

$$1 = F\left(Supp(F)\right) = F\left(Supp'(F)\right) = \int_x I(x \in Supp'(F))F(dx) \tag{4}$$

Lemma 2.1. *Let F be a univariate distribution function. For any $\delta \in (0,1]$, we have*

$$F^{-1}\left(\delta F(y) + (1-\delta)F(y-)\right) = y, \quad \text{for all } y \in Supp'(F). \tag{5}$$

Proof: If y is a discontinuity point of $F(\cdot)$, then certainly $y \in Supp'(F)$ and (5) obviously holds. If $y \in Supp'(F)$ is a continuity point of $F(\cdot)$, then we have for any $y' < y$ that $F(y') < F(y)$. It is now easy to see that (5) holds in this case. □

Lemma 2.2. *Let Y and X be random variables and Y is univariate. Let $Q_{Y|X}(dy;x)$ be a regular conditional distribution of Y given X. Then*

$$(x, y) \to I\left(y \in Supp'\left(Q_{Y|X}(\cdot; x)\right)\right)$$

is a measurable function and

$$P\left(Y \in Supp'\left(Q_{Y|X}(\cdot; X)\right)\right) = 1$$

Proof: Firstly, we will show that $(x, y) \to I\left(y \in Supp'\left(Q_{Y|X}(\cdot; x)\right)\right)$ is a measurable function. We have that, $y \in Supp'\left(Q_{Y|X}(\cdot; x)\right)$ if and only if

$$Q_{Y|X}\left((y - \frac{1}{n}, y]; x\right) > 0 \text{ for any n.}$$

Thus

$$I\left(y \in Supp'\left(Q_{Y|X}(\cdot; x)\right)\right) = \lim_{n\to\infty} I\left(Q_{Y|X}\left((y - \frac{1}{n}, y]; x\right) > 0\right).$$

Since this defines $I\left(y \in Supp'\left(Q_{Y|X}(\cdot; x)\right)\right)$ as a pointwise limit of measurable functions, it is a measurable function itself. In addition, we have

$$\begin{aligned} P\left(Y \in Supp'\left(Q_{Y|X}(\cdot; X)\right)\right) &= EP\left(Y \in Supp'\left(Q_{Y|X}(\cdot; X)\right) | X\right) \\ &= E\int_y I\left(y \in Supp'\left(Q_{Y|X}(\cdot; X)\right)\right) Q_{Y|X}(dy; X) \\ &= 1. \end{aligned}$$

The second equality is due to (3) and the last equality is due to (4). □

Lemma 2.3. *Let Y be a univariate random variable with distribution F. Let Δ be uniformly distributed on $(0, 1]$ and independent of Y. Then*

$$F^{\Delta}(Y) \equiv \Delta F(Y) + (1 - \Delta)F(Y-)$$

is uniformly distributed on $(0, 1]$.

Proof of Lemma 2.3: Let F_n be the convolution of F and $Unif(0, \frac{1}{n}]$, that is,

$$F_n(\cdot) \equiv n \int_{(0,\frac{1}{n}]} F(\cdot - u)du.$$

Note that F_n is a continuous distribution function. Since F is right continuous and has left limit, it is easy to verify that for any $\delta \in (0, 1]$ and $y \in \mathbf{R}$,

$$F_n(y + \frac{\delta}{n}) \longrightarrow \delta F(y) + (1 - \delta)F(y-), \text{ for } n \to \infty.$$

Thus

$$F_n(Y + \frac{\Delta}{n}) \longrightarrow \Delta F(Y) + (1-\Delta)F(Y-), \text{ a.s. for } n \to \infty. \tag{6}$$

The desired result follows now from (6) and the fact that $F_n(Y+\Delta/n)$ is uniformly distributed on $(0,1]$. □

The following Lemma allows us to define a q-q function for discrete distribution functions.

Lemma 2.4. *Let X_1 and X_2 be univariate random variables with distribution functions F_1 and F_2. Let Δ be independent of (X_1, X_2). Then $F_2^{-1}\left(F_1^{\Delta}(X_1)\right)$ is distributed as F_2.*

Proof: The proof is straightforward by Lemma 2.3. □

Lemma 2.5. *Let X, Y and Z be random variables and Z is independent of (X, Y). Let $Q_{Y|X}(dy; x)$ be a regular conditional distribution of Y given X. Then*

(1) $Q_{Y,Z|X}(dy, dz; x) \equiv P(Z \in dz)Q_{Y|X}(dy; x)$ is a regular conditional distribution of (Y, Z) given X.

(2) $Q_{Y|X}(dy; x)$ is a regular conditional distribution of Y given (X, Z).

Proof: The proof is straightforward and omitted. □

Lemma 2.6. *Let Y be a univariate random variable, $X \in \mathbf{R}^d$ be a random variable and $Q_{Y|X}(dy; x)$ be a regular conditional distribution of Y given X. Let $\Delta \sim Unif(0,1]$ be independent of (X, Y). Let $Z \in \mathbf{R}^e$ be a random variable and independent of (Δ, X, Y). Let $h : \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}^e \to \mathbf{R}$ be a bounded measurable function. Define*

$$Q^{\Delta}_{Y|X}(Y; X) \equiv \Delta Q_{Y|X}(Y; X) + (1-\Delta)Q_{Y|X}(Y-; X).$$

We have

$$E\left(h(Q^{\Delta}_{Y|X}(Y; X), X, Z)|X, Z\right) = \int_{u\in(0,1]} h(u, X, Z)du, \quad a.s.$$

Lemma 2.6 teaches us that $Q^{\Delta}_{Y|X}(Y; X)$ is uniformly distributed on $(0,1]$ given (X, Z).

Proof: Lemma 2.6 is a straightforward result of Lemma 2.3 and Lemma 2.5. □

2.4 Proof of Theorem 2.1

Proof: *Consistency.* Let

$$\Omega_0 \equiv \bigcap_{k,j} \left\{ \omega : L_{k,j} \in Supp' \left(Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(\cdot; \bar{A}_{k-1}, \bar{L}_{k,j-1}) \right) \right\}.$$

By Lemma 2.2, $P(\Omega_0) = 1$. It follows from Lemma 2.1 immediately that

$$L_{2,1} = L^{\bar{g}}_{2,1} \equiv Q^{-1}_{L_{2,1}|A_1,L_1} \left(Q^{\Delta_{2,1}}_{L_{2,1}|A_1,L_1}(L_{2,1}; A_1, L_1); g_1(L_1), L_1 \right) \tag{7}$$

on $\{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(L_K)\}$. Now it can be verified by deduction that for $k = 1, \ldots, K+1$ and for $j = 1, \ldots, p_k$,

$$L_{k,j} = L^{\bar{g}}_{k,j}, \text{ on } \{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(L_K)\}.$$

Therefore $L_k = L^{\bar{g}}_k$ on $\{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(L_K)\}$ for $k = 2, \ldots, K+1$ and the consistency assumption A1 follows.

Randomization. Given a finite subset $\mathcal{G}_0 \subset \mathcal{G}$ and a bounded measurable function of $(Y^{\bar{g}}; \bar{g} \in \mathcal{G}_0)$, denoted with $h(Y^{\bar{g}}; \bar{g} \in \mathcal{G}_0)$. We will show that $E(h|\bar{A}_m, \bar{L}_m)$ is only a function of $(\bar{A}_{m-1}, \bar{L}_m)$. We calculate $E(h|\bar{A}_m, \bar{L}_m)$ by sequentially conditioning on $(\bar{A}_K, \bar{L}_{K+1,p_{K\pm1}-1}, \bar{\Delta}_{K+1,p_{K+1}-1})$, $(\bar{A}_K, \bar{L}_{K+1,p_{K+1}-2}, \bar{\Delta}_{K+1,p_{K+1}-2})$, $\ldots$, $(\bar{A}_m, \bar{L}_m, \bar{\Delta}_m)$ and $(\bar{A}_m, \bar{L}_m)$. h is actually only a function of

$$\left(Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j}; \bar{A}_{k-1}, \bar{L}_{k,j-1}); 1 \leq k \leq K+1, 1 \leq j \leq p_{K+1} \right).$$

Denote it with

$$h_{K+1,p_{K+1}} \left(Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j}; \bar{A}_{k-1}, \bar{L}_{k,j-1}); k \leq K+1, 1 \leq j \leq p_{K+1} \right).$$

Firstly, take the conditional expectation of $h_{K+1,p_{K+1}}$ given $(\bar{A}_K, \bar{L}_{K+1,p_{K+1}-1}, \bar{\Delta}_{K+1,p_{K+1}-1})$. We have that $Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}} \left(L_{k,j}; \bar{A}_{k-1}, \bar{L}_{k,j-1} \right)$ is uniformly distributed on $(0, 1]$ given $(\bar{A}_{k-1}, \bar{L}_{k,j-1}, \bar{\Delta}_{k,j-1})$ by Lemma 2.6. The result of the conditional expectation can be obtained as follows: first replace $Q^{\Delta_{K+1,p_{K+1}}}_{L_{K+1,p_{K+1}}|\bar{A}_K,\bar{L}_{K+1,p_{K+1}}}$ $(L_{K+1,p_{K+1}}; \bar{A}_K, \bar{L}_{K+1,p_{K+1}})$ by u and then integrate w.r.t u on $(0, 1]$. We obtain a function of

$$\begin{aligned}(Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}&(L_{k,j}; \bar{A}_{k-1}, \bar{L}_{k,j-1}); k \leq K+1,\\ &1 \leq j \leq p_k \text{ for } 1 \leq k \leq K \text{ and } 1 \leq j \leq p_{K+1} - 1 \text{ for } k = K+1).\end{aligned}$$

Denote this function with $h_{K+1,p_{K+1}-1}$. Iterating this process (i.e. take the conditional expectation of $h_{K+1,p_{K+1}-1}$ given $(\bar{A}_K, \bar{L}_{K+1,p_{K+1}-2}, \bar{\Delta}_{K+1,p_{K+1}-2})$ and so on) and ending with conditioning on $(\bar{A}_m, \bar{L}_m, \bar{\Delta}_m) = (\bar{A}_m, \bar{L}_{m,p_m}, \bar{\Delta}_{m,p_m})$, one ends up with h_{m,p_m}, which is only a function of

$$\left(Q^{\Delta_{k,j}}_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j}; \bar{A}_{k-1}, \bar{L}_{k,j-1}); 1 \leq k \leq m, 1 \leq j \leq p_k\right).$$

This is obviously a function of $(\bar{A}_{m-1}, \bar{L}_m, \bar{\Delta}_m)$. Denote this function with $f(\bar{A}_{m-1}, \bar{L}_m, \bar{\Delta}_m)$. Now we obtain

$$E(h|\bar{A}_m, \bar{L}_m, \bar{\Delta}_m) = f(\bar{A}_{m-1}, \bar{L}_m, \bar{\Delta}_m)$$

Since $\Delta_{k,j}$ are $I.I.D.\ Unif(0,1]$ and independent of $\bar{A}$ and $\bar{L}$, we have

$$\begin{aligned} E(h|\bar{A}_m, \bar{L}_m) &= E\left(E(h|\bar{A}_m, \bar{L}_m, \bar{\Delta}_m)|\bar{A}_m, \bar{L}_m\right) \\ &= E(f(\bar{A}_{m-1}, \bar{L}_m, \bar{\Delta}_m)|\bar{A}_m, \bar{L}_m) \\ &= \int_{(0,1]} \cdots \int_{(0,1]} f(\bar{A}_{m-1}, \bar{L}_m, \bar{u}_m) du_{1,1} du_{1,2} \ldots du_{m,p_m}. \end{aligned}$$

We conclude that $E(h|\bar{A}_m, \bar{L}_m)$ is only a function of $(\bar{A}_{m-1}, \bar{L}_m)$ which completes the proof. □

3 G-computation formula

In section 2, we showed that the counterfactuals $(Y^{\bar{g}} : \bar{g} \in \mathcal{G})$ satisfying A1 and A2 can be constructed given the observed variables. In this section we show that the counterfactual distribution can be identified by the observed data distribution under A1, A2 and A3* which is a generalized version of A3. assuming that treatment $\bar{A}$ is discrete valued. The counterfactual distribution is indeed given by the general form of the G-computation formula where we replace conditional probabilities in (1) by conditional distributions. The result of G-computation formula does not depend on how one chooses the conditional distributions.

Section 3.1 gives the theorem. The proof of the G-computation formula is given in section 3.2

3.1 G-computation formula

Theorem 3.1. *(G-computation formula) Let $(\bar{A}_K, \bar{L}_K, Y \equiv L_{K+1})$ be a random variable defined on a given probability space $(\Omega, \mathcal{F}, P)$ and let A_k be discrete, $k = 1, \ldots, K$. Assume that the consistency and randomization assumptions A1 and A2 hold. Let $Q_{L_{k+1}|\bar{A}_k,\bar{L}_k}(dl_{k+1}; \bar{a}_k, \bar{l}_k)$ be a regular conditional distribution of L_{k+1} given $(\bar{A}_k, \bar{L}_k)$ for $k = 1, \ldots K$ (note that $L_{K+1} \equiv Y$). Let $Q_{L_{k+1}|\bar{A}_k,\bar{L}_k}(dl_{k+1}; \bar{a}_k, \bar{l}_k)$ denote $P(L_1 \in dl_1)$ when $k = 0$. Let $\bar{g}$ be a treatment plan. Assume that $\bar{g}_K$ satisfies the following identifiability assumption.*

A3* *For any Borel set C with $P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), \bar{L}_k \in C) > 0$, we have $P(\bar{A}_k = \bar{g}_k(\bar{L}_k), \bar{L}_k \in C) > 0$.*

We have (recall that $l_{K+1} \equiv y$)

$$P(Y^{\bar{g}} \in dy) = \int_{l_1} \cdots \int_{l_K} \prod_{k=0}^{K} Q_{L_{k+1}|\bar{A}_k, \bar{L}_k}(dl_{k+1}; g_k(\bar{l}_k), \bar{l}_k). \tag{8}$$

Remark We note that when L_k is discrete, A3* is equivalent to the discrete version of A3 given in Section 1. A sufficient condition for A3* to hold is

A3** For any $\bar{l}_k$ and $\bar{a}_k = \bar{g}_k(\bar{l}_k)$ with $(\bar{a}_{k-1}, \bar{l}_k) \in Supp(\bar{A}_{k-1}, \bar{L}_k)$, it follows that $Q_{A_k|\bar{A}_{k-1}, \bar{L}_k}(\{a_k\}; \bar{a}_{k-1}, \bar{l}_k) > 0$, where $Q_{A_k|\bar{A}_{k-1}, \bar{L}_k}(\cdot; \bar{a}_{k-1}, \bar{l}_k)$ is a regular conditional distribution of A_k given $(\bar{A}_{k-1}, \bar{L}_k)$.

A3** is easier to verify practically.

3.2 Proof of Theorem 3.1

We first provide a definition which defines a conditional expectation conditioning on an event and a sub σ-field. Then we establish a Lemma which we need in the proof of Theorem 3.1. We will first remind the definition of conditional expectation. Let $\mathcal{H} \subset \mathcal{F}$ be a sub σ-field of $\mathcal{F}$. $E(X|\mathcal{H})$ is defined as the unique $\mathcal{H}$-measurable random variable ξ which satisfies $E(XI_H) = E(\xi I_H)$ for any $H \in \mathcal{H}$. If $EY^2 < \infty$, then $\xi = E(X|\mathcal{H})$ is the unique (a.s. sense) random variable which minimizes $E(X - \xi)^2$ among $\mathcal{H}$-measurable random variable ξ.

Definition 3.1. *Let Y be a random variable defined on a given probability space $(\Omega, \mathcal{F}, P)$. Let $F \in \mathcal{F}$ and $\mathcal{H} \subset \mathcal{F}$ be a sub σ-field. We define $E(Y|F, \mathcal{H})$ as follows:*

$$E(Y|F, \mathcal{H}) \equiv \frac{E(YI_F|\mathcal{H})}{P(F|\mathcal{H})} I\left(P(F|\mathcal{H}) > 0\right),$$

where I_F denotes the indicator function.

By Definition 3.1, the following Lemma is straightforward.

Lemma 3.1. *$E(Y|F, \mathcal{H})$ as defined by Definition 3.1 satisfies the following properties*

(1) $E(aX + bY|F, \mathcal{H}) = aE(X|F, \mathcal{H}) + bE(X|F, \mathcal{H})$, *if X and Y are integrable.*

(2) $E(Y|F, \mathcal{H}) = E(YI_F|F, \mathcal{H})$.

(3) $E(Y|F, \mathcal{H}) = YI\left(P(F|\mathcal{H}) > 0\right)$ *if Y is $\mathcal{H}$-measurable.*

(4) If $F \in \sigma(\mathcal{H}_1, \mathcal{H}_2)$, where $\mathcal{H}_i, i = 1, 2$ are sub σ-fields, then

$$E(Y|F, \mathcal{H}_1) = E\left(E(Y|\mathcal{H}_1, \mathcal{H}_2)|F, \mathcal{H}_1\right).$$

(5) If Y is $\mathcal{H}$-measurable, then $E(XY|F, \mathcal{H}) = E(X|F, \mathcal{H})Y$.

(6) If $X \perp F|\mathcal{H}$, then $E(X|F, \mathcal{H}) = E(X|\mathcal{H})I\left(P(F|\mathcal{H}) > 0\right)$.

In the following, as a convention, when we condition on a random variable, we mean conditioning on the σ-field generated by the random variable. When we condition on an event and a random variable, we mean conditioning on the event and the σ-field generated by the random variable as defined in Definition 3.1. For example conditioning on $(A_k = g_k(\bar{L}_k), \bar{A}_{k-1}, \bar{L}_k)$ means conditioning on $(F \equiv \{A_k = g_k(\bar{L}_k)\}, \mathcal{H} \equiv \sigma(\bar{A}_{k-1}, \bar{L}_k))$.

Proof of Theorem 3.1: We begin with establishing the following Lemmas. We assume that all the conditions of Theorem 3.1 hold.

Lemma 3.2. $Y^{\bar{g}} \perp \{A_k = g_k(\bar{L}_k)\}|(\bar{A}_{k-1}, \bar{L}_k)$.

Proof: This is a straightforward consequence of the fact that

$$X \perp Y|Z \Longrightarrow X \perp h(Y, Z)|Z,$$

where X, Y and Z are random variables and h is a measurable function. □

Lemma 3.3. *Let $\mathcal{F}_k^{g_k} \equiv (A_k = g_k(\bar{L}_k), \bar{A}_{k-1}, \bar{L}_k)$. We have that for any bounded measurable function h*

$$E\left(h(A_k, \bar{A}_{k-1}, \bar{L}_k)|\mathcal{F}_k^{g_k}\right) = h(g_k(\bar{L}_k), \bar{A}_{k-1}, \bar{L}_k)I(P(A_k = g_k(\bar{L}_k)|\mathcal{F}_k) > 0).$$

Proof: The proof is straightforward by definition 3.1 and omitted. □

Lemma 3.4. *Let $\mathcal{F}_k = (\bar{A}_{k-1}, \bar{L}_k)$. We have $P(A_k = g_k(\bar{L}_k)|\mathcal{F}_k) > 0$ a.s. on $\{\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1})\}$.*

Proof: Let $F_k \equiv \{P(A_k = g_k(\bar{L}_k)|\mathcal{F}_k) = 0\} = \{(\bar{A}_{k-1}, \bar{L}_k) \in C\}$, where C is a Borel set. In order to prove the Lemma, we need to show that

$$P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), F_k) = 0. \qquad (9)$$

Suppose (9) is incorrect. That is,

$$\begin{aligned}
&P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), F_k) \\
&\qquad = P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), (\bar{A}_{k-1}, \bar{L}_k) \in C) \\
&\qquad = P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), (\bar{g}_{k-1}(\bar{L}_{k-1}), \bar{L}_k) \in C) \\
&\qquad = P(\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), \bar{L}_k \in D) > 0,
\end{aligned}$$

where D is a Borel set. By assumption A3*, the last formula implies

$$P\big(\bar{A}_k = \bar{g}_k(\bar{L}_k), \bar{L}_k \in D\big) = P\big(\bar{A}_k = \bar{g}_k(\bar{L}_k), F_k\big) > 0. \quad (10)$$

We note that $\{\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), F_k\}$ is an element of the σ-field generated by $\mathcal{F}_k$. By definition of conditional expectation, we have

$$EI_{A_k=g_k(\bar{L}_k)} I_{\bar{A}_{k-1}=\bar{g}_{k-1}(\bar{L}_{k-1}),F_k} = EP(A_k = g_k(\bar{L}_k)|\mathcal{F}_k) I_{\bar{A}_{k-1}=\bar{g}_{k-1}(\bar{L}_{k-1}),F_k}.$$

The right hand side is zero due to the definition of F_k. But the left hand side is positive by (10). This is a contradiction. □

We now continue to prove the theorem. We first show that for any bounded measurable function h,

$$\begin{aligned} E\left(h(Y^{\bar{g}})\right) &= EE\left(E\left(\ldots E\left(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\right)\ldots|\mathcal{F}_1^{g_1}\right)|L_1\right) && (11)\\ &= EE\left(E\left(\ldots E\left(E(h(Y)|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\right)\ldots|\mathcal{F}_1^{g_1}\right)|L_1\right), && (12)\end{aligned}$$

where $\mathcal{F}_k^+ \equiv (\bar{A}_k, \bar{L}_k)$. Recall that $\mathcal{F}_k \equiv (\bar{A}_{k-1}, \bar{L}_k)$ and $\mathcal{F}_k^{g_k} \equiv (\{A_k = g_k(\bar{L}_k)\}, \bar{A}_{k-1}, \bar{L}_k)$. Firstly, applying property (2) in Lemma 3.1 with $F = \{A_k = g_k(\bar{L}_k)\}$ shows that (11) equals

$$\begin{aligned} &EE\big(E\big(\ldots E\big(E\big(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\big)|\mathcal{F}_{K-1}^+\big)\\ &\qquad I\big(A_{K-1} = g_{K-1}(\bar{L}_{k-1})\big)\ldots I(A_1 = g_1(L_1))|\mathcal{F}_1^{a_1}\big)|L_1\big)\end{aligned}$$

Applying (5) of Lemma 3.1 allows us to move the indicators inside step by step, till we obtain

$$EE\big(E\big(\ldots E\big(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\big)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{k-1}))\ldots|\mathcal{F}_1^{g_1}\big)|L_1\big). \quad (13)$$

We have

$$\begin{aligned} &E\big(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\big)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{k-1}))\\ &\qquad = E(h(Y^{\bar{g}})|\mathcal{F}_K^{g_K}\big)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{k-1}))\\ &\qquad = E(h(Y^{\bar{g}})|\mathcal{F}_K\big)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{k-1}))\end{aligned}$$

The first equality is due to (4) of Lemma 3.1 and the second equality is due to (6) of Lemma 3.1 and Lemma 3.4. Now, plug the last term in (13). Application of (2) and (5) of Lemma 3.1 allows us to delete the indicator $I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{k-1}))$. We also note that conditioning on $\mathcal{F}_K$ and further conditioning on $\mathcal{F}_{K-1}^+$ is equivalent to conditioning on $\mathcal{F}_{K-1}^+$. We have that (13) is equal to

$$EE\big(E\big(\ldots E\big(E(h(Y^{\bar{g}})|\mathcal{F}_{K-1}^+)|\mathcal{F}_{K-1}^{g_{K-1}}\big)\ldots|\mathcal{F}_1^{g_1}\big)|L_1\big).$$

Set $K = K - 1$ and repeat the last procedures till we eventually obtain (11).

Application of (2) and (5) of Lemma 3.1 with $F = \{A_k = g_k(\bar{L}_k)\}$ shows that (11) is equal to

$$EE\left(E\left(\ldots E\left(E(h(Y^{\bar{g}})I(\bar{A}_K = \bar{g}_K(\bar{L}_K))|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\right)\ldots|\mathcal{F}_1^{g_1}\right)|L_1\right)$$

By the consistency assumption A1, the last equality is equal to

$$EE\left(E\left(\ldots E\left(E(h(Y)I(\bar{A}_K = \bar{g}_K(\bar{L}_k))|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\right)\ldots|\mathcal{F}_1^{g_1}\right)|L_1\right)$$

Again, application of (2) and (5) of Lemma 3.1 allows us to delete the indicator function $I(\bar{A}_K = \bar{g}_K(\bar{L}_K))$ which results in (12).

It remains to show that (12) yields the G-computation formula (8). Firstly, we write (12) as

$$EE\left(E\left(\ldots E\left(E(h(Y)|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\right)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{K-1}))\ldots|\mathcal{F}_1^{g_1}\right)|L_1\right). \quad (14)$$

We have (recall that $l_{K+1} \equiv y$)

$$\begin{aligned}
&E\big(E(h(Y)|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}\big)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{K-1})) \\
&\quad = E\left(\int_{l_{K+1}} h(y)Q_{L_{K+1}|\bar{A}_K,\bar{L}_K}(dl_{K+1};\bar{A}_K,\bar{L}_K)\middle|\mathcal{F}_K^{g_K}\right)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{K-1})) \\
&\quad = \int_{l_{K+1}} h(y)Q_{L_{K+1}|\bar{A}_K,\bar{L}_K}(dl_{K+1};g_K(\bar{L}_K),\bar{A}_{K-1},\bar{L}_K)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{K-1})) \\
&\quad = \int_{l_{K+1}} h(y)Q_{L_{K+1}|\bar{A}_K,\bar{L}_K}(dl_{K+1};\bar{g}_K(\bar{L}_K),\bar{L}_K)I(\bar{A}_{K-1} = \bar{g}_{K-1}(\bar{L}_{K-1}))
\end{aligned} \quad (15)$$

The first equality is due to (3). The second equality is due to Lemma 3.3 and Lemma 3.4. We plug (15) in (14) and delete the indicator function. Now, repeating the last procedure by sequentially conditioning on $\mathcal{F}_{K-1}^+$ and $\mathcal{F}_{K-1}^{g_{K-1}}, \ldots$ results in the G-computation formula (8). □

Remark: Since the conditional distributions $Q_{L_{k+1}|\bar{A}_k,\bar{L}_k}(dl_{k+1};\bar{a}_k,\bar{l}_k)$ are not unique, one might be concerned that different choices of conditional distributions would result in different answers. In the proof we show that the regular conditional distributions are just used to compute (12) which is a well defined quantity. Consequently the result of the G-computation formula doesn't depend on how one chooses the regular conditional distributions.

4 Statistical methods based on our construction of counterfactuals

Theorem 2.1 provides a function which maps the observed variable O and its distribution to the treatment specific counterfactuals. Suppose we have n I.I.D.

observations of O. We can obtain a MLE of the partial likelihood (2) by assuming a parametric or semiparametric model for $Q_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j};\bar{A}_{k-1},\bar{L}_{k,j-1})$. For each subject, we could estimate his/her counterfactuals following the procedure described in section 2.1. Now we can estimate a function of counterfactual distribution using the estimated counterfactuals. For example, we can use the sample mean and the sample median of the estimated counterfactuals to get the counterfactual mean and median. The bootstrap can be used to assess the variability of the resulting estimate of $P(Y^{\bar{g}} \in \cdot)$. An alternative way to estimate the counterfactual distribution is to use the G-computation formula directly. This can be implemented by a Monte-Carlo experiment which is described in Gill and Robins [2001]. There are two important gains from the construction approach. First, we can estimate a function of the counterfactual distribution without calculating multivariate integrals (analytically or by Monte-Carlo simulation) which is computationally easier. Second, we can use the constructed counterfactuals to do model selection on the partial likelihood (2), in the manner discussed next.

We notice that for a given estimate $Q^n_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(\cdot;\bar{A}_{k-1},\bar{L}_{k,j-1})$, the constructed counterfactual variable $\hat{Y}^{\bar{g}}$ satisfies the consistency assumption A1 as long as $L_{k,j}$ lies in the Supp' of $Q^n_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(\cdot;\bar{A}_{k-1},\bar{L}_{k,j-1})$. Thus under the additional identifiability assumption A3* , the true counterfactual distribution is uniquely identified by A2 (see Theorem 3.1). We can use a correctly specified model for the treatment mechanism to directly fine-tune the model for $Q_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j};\bar{A}_{k-1},\bar{L}_{k,j-1})$ so that $\hat{Y}^{\bar{g}}$ satisfies the randomization assumption A2. For example the treatment is a $0,1$ random variable and satisfies $P(A_k = 1|\bar{A}_{k-1},\bar{L}_k) = expit(\alpha^T W_k)$, $k = 1,\ldots,K$, where α is an unknown parameter vector, W_k is a known function of $\bar{A}_{k-1},\bar{L}_k$, and $expit(x) = e^x/(1+e^x)$. Let $\hat{Y}^{\bar{g}}$ be the counterfactual variable calculated based on a partial likelihood estimate $Q^n_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j};\bar{A}_{k-1},\bar{L}_{k,j-1})$. We estimate θ with the MLE in the extended model $P(A_k = 1|\bar{A}_{k-1},\bar{L}_k,\hat{Y}^{\bar{g}}) = expit(\alpha^T W_k + \theta\hat{Y}^{\bar{g}})$, $k = 1,\ldots,K$. Since $A_k \perp Y^{\bar{g}}|\bar{A}_{k-1},\bar{L}_k$, if the partial likelihood is estimated well, one would expect that θ is close to zero. Suppose we have several candidate models for $Q_{L_{k,j}|\bar{A}_{k-1},\bar{L}_{k,j-1}}(L_{k,j};\bar{A}_{k-1},\bar{L}_{k,j-1})$. We first calculate the counterfactual variable $\hat{Y}^g$ based on each model and choose the model which minimizes $\hat{\theta}$.

5 Discussion

Our q-q function given by Lemma 2.4 can be used to generalize Robins' Structural Nested Mean (SNM) models to both discrete and continuous outcome variables. It is of interest to understand the precise conditions under which the counterfactual distribution of $Y^{\bar{g}}$ can be identified from the observed data distribution in general, that is, without assuming treatment is discrete. We were not able to settle this completely yet.

Firstly, we shall realize that A1, A2 and A3 themselves can not guarantee the correctness of the G-computation formula. Here is a simple example with no covariates. Let $\Omega = [0,1]$, $a \in \mathcal{A} = [0,1]$, $Y^a(\omega) = I(a = \omega)$ and $A(\omega) = \omega$. Obviously, $A \perp (Y^a; a \in \mathcal{A})$ since the σ-field generated by $(Y^a; a \in \mathcal{A})$ is trivial. Let $Y(\omega) = Y^A(\omega)$. We have $Y \equiv 1$. The variables defined above satisfy A1 and A2. That is, $Y^a = Y$ on $\{A = a\}$ and $A \perp (Y^a; a \in \mathcal{A})$. But certainly $law(Y^a) \neq Q_{Y|A}(dy; a)$. Gill and Robins [2001] provide another counterexample.

References

R. Gill and J. Robins. Causal inference in complex longitudinal studies: continuous case. *Ann. Stat.*, 29(6), 2001.

J. M. Robins. Errata to: "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect" [Math. Modelling **7** (1986), no. 9-12, 1393–1512; MR 87m:92078]. *Comput. Math. Appl.*, 14(9-12):917–921, 1987. ISSN 0898-1221.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Modelling*, 7(9-12):1393–1512, 1986. ISSN 0270-0255. Mathematical models in medicine: diseases and epidemics, Part 2.

J.M. Robins. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A Mulley, editors, *Health Service Research Methdology: A Focus on AIDS*, pages 113–159. NCHSR, U.S. Public Health Service, Dordrecht, 1989.

J.M. Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality (Los Angeles, CA, 1994)*, pages 69–117. Springer, New York, 1997.