

## A Note on the Construction of Counterfactuals and the G-computation Formula

Zhuo Yu\*

Mark J. van der Laan<sup>†</sup>

\*Bristol-Myers Squibb, zhuo.yu@bms.com

<sup>†</sup>Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/ucbbiostat/paper181>

Copyright ©2005 by the authors.

# A Note on the 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. In this paper we provide an explicit function that maps the observed data into a counterfactual variable which satisfies the consistency and randomization assumptions. This offers a practically useful imputation method for counterfactuals. Gill & Robins [2001]'s construction of counterfactuals can be used as an imputation method in principle, but it is very hard to implement in practice. Robins [1987] 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 & Robins [2001] prove the G-computation formula for continuous variables under some continuity assumptions and reformulation of the consistency and the randomization assumptions. We prove that if treatment is discrete (which deals with a less general case compared with Gill & Robins [2001]), then Robins' G-computation formula holds under the original consistency, randomization assumptions and a generalized version of the 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 have happened, if the subject had been (counterfactually) treated with  $a$  (instead of the treatment he actually received). 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  $k = 1, \dots, K$ , 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  $L_1, A_1, \dots, L_{k-1}, A_{k-1}$  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

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

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  $\bar{L}_k = (L_1, \dots, L_k)$ ,  $\bar{A}_k = (A_1, \dots, A_k)$ . 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 measurable function 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 measurable 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  $\bar{l}_k = (l_1, \dots, l_k)$  up till then. Define  $\bar{g}_k(\bar{l}_k) = (g_1(l_1), g_2(l_1, l_2), \dots, g_k(l_1, \dots, l_k))$ . 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:

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

Gill and Robins [2001] showed that counterfactual variables satisfying A1 and A2 can be constructed given the observed variables. The first contribution of this article is that we provide an explicit function which maps the observed data and some auxiliary uniform random variables into a counterfactual variable which satisfies the consistency and randomization assumptions A1 and A2. This offers a practically useful imputation method for counterfactuals. Since the construction given in Gill and Robins [2001] involves computing the conditional distribution of counterfactuals given the observed data (see section 6 in Gill and Robins [2001] for details), their construction is from a practical point of view less appealing than our construction in terms of conditional distributions of the observed data.

In the general setting which allows treatment and covariates to be continuous, in order to establish the G-computation formula, Gill and Robins [2001] reformulate and strengthen these original assumptions A1 and A2: see Assumptions **A1\*** and **A2\*** in Gill and Robins [2001]. They also assume some continuity assumptions regarding the joint law of the counterfactual variables and the factual variables: see Assumptions **C** and **Cg** in Gill and Robins [2001]. The continuity assumptions are empty if both the treatment and covariate variables are discrete.

The second contribution of this article is that, in the case that only the treatment value is discrete (the covariates can be continuous), we prove that 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\*, where A3\* is a generalized version of A3 specified below. We note that in this case, assumptions **C** and **Cg** in Gill and Robins [2001] are not empty, and their reformulated assumptions **A1\*** and **A2\*** are stronger than the original A1 and A2.

We 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. Our result also shows that the G-computation formula does not depend on the actual version of the regular conditional distributions.

**Organization of the paper.** Section 2 provides our function which maps the observed data into counterfactual variable 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 with the G-computation formula with conditional probabilities in (1) replaced by conditional distributions.

## 2 Construction of counterfactuals

Given observed variables defined on a given probability space, we provide an explicit construction of 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. Section 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  $L_k = (L_{k,1}, \dots, L_{k,p_k}) \in \mathbb{R}^{p_k}$ . Let  $\Delta_k \equiv (\Delta_{k,1}, \dots, \Delta_{k,p_k})$ ,  $k = 1, \dots, K + 1$ , where  $\Delta_{k,j}$  are all uniformly independently distributed on  $(0, 1]$  defined on the same probability space and independent of  $O$ . Let  $\bar{L}_{k,j} \equiv (\bar{L}_{k-1}, L_{k,1}, \dots, L_{k,j})$ , and  $\bar{L}_{k,0} \equiv \bar{L}_{k-1}$ . Similarly we define  $\bar{\Delta}_{k,j}$ . Let  $Q_{L_{k,j}|\bar{a}_{k-1}, \bar{l}_{k,j-1}}(\cdot)$  be a regular conditional distribution of  $L_{k,j}$  given  $(\bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_{k,j-1} = \bar{l}_{k,j-1})$ . 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  $k > 1$  and  $j$ , define  $L_{k,j}^{\bar{g}}$  recursively by: for  $k = 2, \dots, K + 1$ , for  $j = 1, \dots, p_k$

$$L_{k,j}^{\bar{g}} \equiv Q_{L_{k,j}|\bar{g}_{k-1}(\bar{L}_{k-1}), \bar{L}_{k,j-1}^{\bar{g}}}^{-1} \left( Q_{L_{k,j}|\bar{A}_{k-1}, \bar{L}_{k,j-1}}^{\Delta_{k,j}}(L_{k,j}) \right),$$

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

$$Q_{L_{k,j}|\bar{A}_{k-1}, \bar{L}_{k,j-1}}^{\Delta_{k,j}}(L_{k,j}) = \Delta_{k,j} Q_{L_{k,j}|\bar{A}_{k-1}, \bar{L}_{k,j-1}}(L_{k,j}) + (1 - \Delta_{k,j}) Q_{L_{k,j}|\bar{A}_{k-1}, \bar{L}_{k,j-1}}(L_{k,j-}).$$

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

**Remark:** Note that the function depends 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\*.

## 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)  $B \rightarrow Q_{Y|X}(B; x)$  is a probability measure for any fixed  $x$ .
- (b)  $x \rightarrow Q_{Y|X}(B; x)$  is a measurable function for any Borel set  $B$ ,

and

$$E(h(X, Y)|X) = \int_y h(X, y)Q_{Y|X}(dy; X), a.s. \quad (2)$$

for any bounded measurable function  $h(x, y)$ . We note that  $Q_{Y|X}(dy; x)$  can be defined arbitrarily outside the support of  $X$ . We denote  $Q_{Y|X}((-\infty, y] : x)$  with  $Q_{Y|X}(y; x)$ . It follows that  $(y, x) \rightarrow Q_{Y|X}(y; x)$  is a measurable function as well. To avoid cumbersome notation, we sometimes use  $Q_{Y|x}(dy)$  to denote  $Q_{Y|X}(dy; x)$ .

**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  $\delta$ . We define the support of  $X$ , denoted  $\text{Supp}(X)$  or  $\text{Supp}(F_X)$ , 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  $Z$ .  $X \perp (Y_t, t \in T) | Z$  for an arbitrary index set  $T$  means 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  $\text{Supp}'$  of a univariate distribution function  $F$ . We say  $x \in \text{Supp}'(F)$ , if for all  $\delta > 0$ ,  $F((x - \delta, x]) > 0$ . Let  $D(F)$  be all the continuity points,  $x$ , of  $F$  satisfying the following conditions: (1) for any  $x' > x$ ,  $F(x') > F(x)$  and (2) there exists  $x' < x$ , such that  $F(x') = F(x)$ . It is easy to see that  $D(F)$  is countable. Since  $D(F)$  only consists of continuity points of  $F(dy)$ , it has zero mass w.r.t.  $F(dy)$ . It is not hard to show  $\text{Supp}'(F) = \text{Supp}(F) \setminus D(F)$ . Therefore,

$$1 = F(\text{Supp}(F)) = F(\text{Supp}'(F)) = \int_x I(x \in \text{Supp}'(F)) F(dx) \quad (3)$$

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

$$F^{-1}(\delta F(y) + (1 - \delta)F(y-)) = y, \quad \text{for all } y \in \text{Supp}'(F). \quad (4)$$

**Proof:** If  $y$  is a discontinuity point of  $F(\cdot)$ , then certainly  $y \in \text{Supp}'(F)$  and (4) obviously holds. If  $y \in \text{Supp}'(F)$  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 (4) holds in this case.  $\square$

**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*

$$(x, y) \rightarrow I(y \in \text{Supp}'(Q_{Y|X}(\cdot; x)))$$

is a measurable function and

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

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

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

Thus

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

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

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

The second equality is due to (2) and the last equality is due to (3).  $\square$

**Lemma 2.3.** *Let  $Y$  be a univariate random variable with distribution  $F$ . Let  $\Delta$  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  $\text{Unif}(0, \frac{1}{n}]$ , 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  $y \in \mathbb{R}$ ,

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

Thus

$$F_n\left(Y + \frac{\Delta}{n}\right) \longrightarrow \Delta F(Y) + (1 - \Delta)F(Y-), \text{ a.s. for } n \rightarrow \infty. \quad (5)$$

The desired result follows now from (5) and the fact that  $F_n(Y + \Delta/n)$  is uniformly distributed on  $(0, 1]$ .  $\square$

## 2.4 Proof of Theorem 2.1

**Proof:** For simplicity, assume that all  $L_k$  are univariate, i.e. that all  $p_k = 1$ . The proof can be easily generalized for the case  $p_k > 1$ : it is the step from one  $k$  to the next that is important, not what happens when moving from one coordinate of a multivariate  $L_k$  to the next.

Define for  $k \geq 2$ ,

$$U_k \equiv Q_{L_k|\bar{A}_{k-1}, \bar{L}_{k-1}}^{\Delta_k}(L_k). \quad (6)$$

By Lemma 2.3  $U_k$  is uniform  $(0, 1]$  and independent of  $\bar{A}_{k-1}, \bar{L}_{k-1}$ .  $U_k$  is also a function of  $\bar{A}_{k-1}, \bar{L}_k$  and  $\Delta_k$  and since the  $\Delta_j$  are independent of the observation vector  $O = (\bar{A}_K, \bar{L}_K)$ , it follows that

$$U_k \text{ is independent of } \mathcal{F}_{k-1}$$

where  $\mathcal{F}_j$  is the  $\sigma$ -algebra generated by  $(\bar{A}_j, \bar{L}_j, \Delta_j)$ . This implies in particular that the  $U_k$  for  $k \geq 2$  are i.i.d. uniform  $(0, 1]$ .

*Proof of Consistency, A1*

Let

$$\Omega_0 \equiv \bigcap_k \left\{ \omega : L_k(\omega) \in \text{Supp}' \left( Q_{L_k|\bar{A}_{k-1}(\omega), \bar{L}_{k-1}(\omega)}(\cdot) \right) \right\}.$$

By Lemma 2.2,  $P(\Omega_0) = 1$ .

For  $\omega \in \{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(\bar{L}_K)\}$

$$L_2^{\bar{g}}(\omega) = Q_{L_2|A_1(\omega), L_1(\omega)}^{-1} \left( Q_{L_2|A_1(\omega), L_1(\omega)}^{\Delta_2(\omega)}(L_2(\omega)) \right) \quad (7)$$

is of the form  $F^{-1}(\delta F(y) + (1 - \delta)F(y-))$  with  $F = Q_{L_2|A_1(\omega), L_1(\omega)}$ ,  $\delta = \Delta_2(\omega)$  and  $y = L_2(\omega)$ . By Lemma 2.1 therefore,  $L_2^{\bar{g}}(\omega) = L_2(\omega)$  on  $\{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(\bar{L}_K)\}$

Now it can be verified by induction that for  $k = 2, \dots, K + 1$ ,

$$L_k = L_k^{\bar{g}}, \text{ on } \{\omega \in \Omega_0 : \bar{A}_K = \bar{g}_K(\bar{L}_K)\}$$

and the consistency assumption A1 follows.

*Proof of Randomization, A2*

Given a finite subset  $\mathcal{G}_0 \subset \mathcal{G}$  and a bounded measurable function  $h(Y^{\mathcal{G}_0})$  of  $Y^{\mathcal{G}_0} \equiv (Y^{\bar{g}}; \bar{g} \in \mathcal{G}_0)$ , we will show that  $E(h|\bar{A}_k, \bar{L}_k)$  is only a function of  $(\bar{A}_{k-1}, \bar{L}_k)$ .

By the definition of  $Y^{\bar{g}}$  we may write

$$h(Y^{\mathcal{G}_0}) = \psi(L_1, \bar{U}_{K+1})$$

Since  $U_{k+1}, \dots, U_{K+1}$  are i.i.d. uniform given  $\mathcal{F}_k$  and  $(L_1, \bar{U}_k)$  is  $\mathcal{F}_k$ -measurable, we find

$$E(h(Y^{\mathcal{G}_0})|\mathcal{F}_k) = \int_0^1 du_{k+1} \dots \int_0^1 du_{K+1} \psi(L_1, \bar{U}_k, u_{k+1}, \dots, u_{K+1}) = \rho(L_1, \bar{U}_k).$$

But  $\bar{U}_k$  is of the form  $\phi_k(\bar{A}_{k-1}, \bar{L}_k, \bar{\Delta}_k)$  so

$$E(h(Y^{\mathcal{G}_0})|\mathcal{F}_k) = \rho(L_1, \phi_k(\bar{A}_{k-1}, \bar{L}_k, \bar{\Delta}_k))$$

which because  $\bar{\Delta}_k$  is independent of  $\bar{A}_k, \bar{L}_k$  with uniform coordinates implies that

$$E(h(Y^{\mathcal{G}_0})|\bar{A}_k, \bar{L}_k) = \int_0^1 dz_1 \dots \int_0^1 dz_k \rho(L_1, \phi_k(\bar{A}_{k-1}, \bar{L}_k, \bar{z}_k)).$$

Thus

$$E(h(Y^{\mathcal{G}_0})|\bar{A}_k, \bar{L}_k) = E(h(Y^{\mathcal{G}_0})|\bar{A}_{k-1}, \bar{L}_k)$$

which proves the randomization requirement **A2**.  $\square$

### 3 G-computation formula

In section 2, we showed that we can map the observed data structure in the counterfactuals  $(Y^{\bar{g}} : \bar{g} \in \mathcal{G})$  satisfying A1 and A2. In this section we show that,

Research Archive

if  $\bar{A}$  is discrete valued, then the counterfactual distribution can be identified by the observed data distribution under A1, A2 and A3\*, where A3\* is a generalized version of A3 specified below. In other words, the counterfactual distribution is uniquely determined by A1, A2 and A3\*. The counterfactual distribution is indeed given by the general form of the G-computation formula (1) in which conditional probabilities are replaced by conditional distributions. The result of this generalized G-computation formula does not depend on the actual versions of the conditional distributions.

Section 3.1 states the theorem. The proof of the G-computation formula is provided 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,  $k = 1, \dots, K$ . 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  $k = 1, \dots, K$  (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  $l_{K+1} \equiv y$ )

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

**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  $(\bar{a}_{k-1}, \bar{l}_k) \in \text{Supp}(\bar{A}_{k-1}, \bar{L}_k)$ , 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  $(\bar{A}_{k-1}, \bar{L}_k)$ .

In the proof of the theorem, we actually only need the following weaker version of A2:

**A2\***  $A_k \perp Y^{\bar{g}} | (\bar{A}_{k-1}, \bar{L}_k)$  for each  $\bar{g}$  on the event  $\{\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1})\}$ .

### 3.2 Proof of Theorem 3.1

We first provide a definition which defines a conditional expectation conditioning on an event and a sub  $\sigma$ -field. Then we establish a Lemma which we need in the proof of Theorem 3.1. We will first recall the definition of conditional expectation. Let  $\mathcal{H} \subset \mathcal{F}$  be a sub  $\sigma$ -field of  $\mathcal{F}$ .  $E(X|\mathcal{H})$  is defined as the unique  $\mathcal{H}$ -measurable random variable  $\xi$  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 (in the a.s. sense) random variable which minimizes  $E(X - \xi)^2$  among  $\mathcal{H}$ -measurable random variable  $\xi$ .

In the remainder of this section, we make the following conventions. If we condition on a random variable, then we mean conditioning on the  $\sigma$ -field generated by the random variable. In addition, if we condition on an event and a random variable, then we mean conditioning on the event and the  $\sigma$ -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  $(F \equiv \{A_k = g_k(\bar{L}_k)\}, \mathcal{H} \equiv \sigma(\bar{A}_{k-1}, \bar{L}_k))$ . Finally, if we state that two random variables are equal (e.g.,  $X = Y$ ), then we mean that they are equal almost surely ( $X = Y$ , a.s.).

**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  $\sigma$ -field. We define  $E(Y|F, \mathcal{H})$  as follows:

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

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

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

(3)  $E(Y|F, \mathcal{H}) = YI(P(F|\mathcal{H}) > 0)$  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  $\sigma$ -fields, then

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

(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  $E(X|F, \mathcal{H}) = E(X|\mathcal{H})I(P(F|\mathcal{H}) > 0)$ .

**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

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

where  $X, Y$  and  $Z$  are random variables and  $h$  is a measurable function.  $\square$

**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  $h$

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

**Proof:** The result follows directly from definition 3.1.  $\square$

**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

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

Suppose (9) is incorrect. That is,

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

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

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

We note that  $\{\bar{A}_{k-1} = \bar{g}_{k-1}(\bar{L}_{k-1}), F_k\}$  is an element of the  $\sigma$ -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.  $\square$

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

$$E(h(Y^{\bar{g}})) = EE(E(\dots E(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}) \dots |\mathcal{F}_1^{g_1}) |L_1) \quad (11)$$

$$\equiv EE(E(\dots E(E(h(Y)|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K}) \dots |\mathcal{F}_1^{g_1}) |L_1), \quad (12)$$

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

$$EE(E(\dots E(E(E(h(Y^{\bar{g}})|\mathcal{F}_K^+)|\mathcal{F}_K^{g_K})|\mathcal{F}_{K-1}^+) \\ I(A_{K-1} = g_{K-1}(\bar{L}_{k-1}))|\mathcal{F}_{K-1}^{g_{K-1}}) \dots I(A_1 = g_1(L_1))|\mathcal{F}_1^{g_1})|L_1)$$

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

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

We have

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

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

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

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

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

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

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

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

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

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

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

The first equality is due to (2). 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  $\mathcal{F}_{K-1}^{g_{K-1}}, \dots$  results in the G-computation formula (8).  $\square$

**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.

## Acknowledgement.

We would like to thank the referees for their helpful contributions. This research was funded by an NIH grant R01 GM67233.

## 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 Methodology: 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.