COMMUN. STATIST.—THEORY METH., 23(8), 2379-2412 (1994)

CORRECTING FOR NON-COMPLIANCE IN RANDOMIZED TRIALS USING STRUCTURAL NESTED MEAN MODELS

James M. Robins

Departments of Epidemiology and Biostatistics
Harvard School of Public Health
677 Huntington Ave., Boston, MA 02115

Key Words and Phrases: Semiparametric models; Semiparametric efficiency; Causal inference; Longitudinal data; Sequential moment restrictions; Panel data; Observational studies; Time-dependent covariates.

ABSTRACT

In a randomized trial designed to study the effect of a treatment of interest on the evolution of the mean of a time-dependent outcome variable, subjects are assigned to a treatment regime, or, equivalently, a treatment protocol. Unfortunately, subjects often fail to comply with their assigned regime. From a public health point of view, the causal parameter of interest will often be a function of the treatment differences that would have been observed had, contrary to fact, all subjects remained on protocol. This paper considers the identification and estimation of these treatment differences based on a new class of structural models, the multivariate structural nested mean models, when reliable estimates of each subject's actual treatment are available. Estimates of "actual treatment" might, for example, be obtained by measuring the amount of "active drug" in the subject's blood or urine at each follow-up visit or by pill counting techniques. In addition, we discuss a natural extension of our methods to observational studies.

1. INTRODUCTION

In this paper we shall be concerned with randomized trials designed to study the effect of a treatment of interest on the evolution of the mean of a time-dependent outcome variable. For concreteness, we shall consider a randomized trial studying the effect of AZT treatment on the evolution of T4-lymphocyte count history among subjects with newly diagnosed infection with the human immunodeficiency virus (HIV). In such a randomized trial, subjects are assigned to an AZT treatment regime, or, equivalently, a treatment protocol. Thereafter, at regular intervals, e.g., monthly, subjects return to clinic to have their T4-count measured. Unfortunately, subjects often fail to comply with their assigned regime. As discussed in Robins (1989), Robins and Tsiatis (1991), and Mark and Robins (1993b), the causal parameter of interest will often be a function of the differences in T4-count evolution that would have been observed had, contrary to fact, all subjects remained on protocol, that is under complete compliance.

This paper considers methods for the estimation of differences in the evolution of AZT treatment-regime-specific mean T4-count had there been complete compliance, when reliable estimates of each subject's actual AZT treatment are available. In general, the differences in regime-specific mean T4-count that would have been observed at any occasion had all subjects remained on protocol will not be identifiable without making various assumptions that are themselves non-identifiable. The simplest, but least plausible assumption, is that subjects who comply with their assigned treatment protocol are exchangeable with (i.e., comparable to) those who fail to comply. In this paper we shall develop assumptions that, when true, are sufficient to identify differences in regime-specific mean T4-counts that would have been observed had all subjects remained on protocol and yet are more plausible than a simple exchangeability assumption. Specifically, in Section 4, we show that we can identify these treatment differences if the following two conditions are met. First, there is no-current-treatment-interaction as defined in Section 3. Informally, there is no current-treatment-interaction if, at each follow-up time t, the investigator has data available on the history of all covariates (e.g., hematocrit, weight, presence of fever) that, conditional on AZT treatment history up to t, predict both (a) increments in AZT dosage at t and (b) the magnitude of the differences in subsequent AZT regime-specific mean T4-counts. Second, the population follows a multivariate structural nested mean model (SNMM) introduced in Robins (1989). SNMMs are parametric models for the effect of an incremental "blip" of AZT treatment at time t on the subsequent evolution of mean T4-count. We have adopted terminology used in the social science and econometric literature and have called our models "structural models" because they directly model the regime-specific mean T4-count history that would have been observed had, contrary to fact, all subjects followed treatment with a prescribed regime (Rubin, 1978).

In Section 3c, we show that the approach presented in this paper for estimating the treatment differences that would have been observed had all subjects remained on protocol is consistent with the standard intention to treat analysis of a randomized trial in regards to whether the null hypothesis of no treatment effect on the evolution of mean T4-count is rejected. Thus our approach is an extension of and does not conflict with the standard analytical approach.

In Sections 4 and 5, we extend our results to observational studies and consider G-estimation of the parameters of our structural nested mean models. We also show that the asymptotic variance of the optimal estimator in our class attains the semiparametric efficiency bound for our model. This section extends to SNMM the G-estimation of structural nested failure time models discussed in Robins (1992, 1993ab), Robins et al. (1992), Robins and Greenland (1994), Mark and Robins (1993b) and of structural nested distribution models for repeated measure outcomes discussed in Robins et al. (1992 Appendix 2), and Robins (1993a). We also discuss the connection between our model and models proposed by Chamberlain (1992, 1993). In Sections 6 and 7, we extend our results to allow for missing data and for multiplicative effects of exposure on the mean of the response. In Section 4, Remark 1, we describe why it is preferable to estimate the effect of a time-dependent treatment using G-estimators of structural nested mean models rather than the G-computation algorithm estimator of Robins (1989, Secs. 5, 13).

Holland (1988), Efron and Feldman (1991), Heckman and Robb (1985), Angrist and Imbens (1992), Angrist, Imbens, and Rubin (1993), and Robins (1986, 1989, 1993a), Robins and Rotnitzky (1992, 1993), Robins and Greenland (1992), Mark and Robins (1993), Robins, Rotnitzky, and Zhao (1993), and Robins and Greenland (1994) have previously considered the problem of correcting for non-compliance in randomized trials. The Appendix to Robins (1993a) covers for failure time outcomes the same ground as this paper. In Section 3c, Holland's approach is a special case of the more general approach adopted in this paper. Mark and Robins (1993b) show that Efron's and Feldman's approach has little in common with the approach adopted here. The relationship between our methods and those of Heckman and Robb (1985) is considered in Robins (1989, Section 14).

2. A FORMALIZATION OF OUR PROBLEM

2a. The Observed Data

For pedagogic purposes we shall consider throughout a multi-armed double-blind randomized trial of the effect of AZT treatment on T4-count history conducted on asymptomatic subjects with newly diagnosed HIV infection. We shall assume that the study subjects are assigned to one of J treatment protocols. Until the discussion section, we assume there is no missing data.

Let $(r_i, \underline{a}_i, \bar{\ell}_i)$ denote the data collected on each study subject where $i \in \{1, \cdots, n\}$ indexes the n study subjects entered in the trial. r_i is a treatment indicator that takes the value j if subject i was assigned to the j[th] protocol: $\underline{a}_i$ records subject i's AZT dosage history; and $\bar{\ell}_i$ records data on subject i's T4-count history and on other pretreatment and post-treatment variables.

Formally, suppose that data on T4-count and other time-dependent covariates (e.g., hematocrit, weight, and temperature) were obtained at and only at prespecified times $(t_0, \cdots, t_K)$ where t_0 is time of randomization and t_K is end of follow-up. Data on additional pretreatment variables such as race, sex, and age at entry are also collected at t_0. Let $y_{k,i}$ be T4-count recorded at t_k, $t_k \epsilon (t_0,\cdots,t_K)$. Let $\ell_{k,i}$ be the vector consisting of $y_{k,i}$ and all other time-dependent covariates measurements obtained on subject i at t_k except $\ell_{0,i}$ also includes data on pretreatment variables such as race and sex. Further, we suppose that at each t_k, we obtain accurate measurements of the average AZT dosage rate in $(t_{k-1},t_k]$. We make the simplifying assumption that, for each subject, the AZT dosage rate in $(t_{k-1},t_k]$ is constant. Let $a_{k-1,i}$ be subject i's AZT dosage rate in $(t_{k-1},t_k]$. Thus $a_{k-1,i}$ is the AZT dosage rate subsequent to measuring $\ell_{k-1,i}$ but prior to measuring $\ell_{k,i}$. For notational convenience we define $a_K \equiv 0$.

We shall use overbars to denote histories of time-dependent covariates or treatments. Thus, for example, $\bar{a}_{k,i} = (a_{0,i},...,a_{k,i})$. Note $\bar{a}_{k,i}$ is subject i's AZT history through t_{k+1}. Any history that is not subscripted by a particular value of k will denote a history through end of follow-up. Therefore $\bar{a}_i \equiv \bar{a}_{K,i}$ and $\bar{\ell}_i \equiv \bar{\ell}_{K,i}$. In this notation the data obtained on each subject is $(r_i,\bar{a}_i,\bar{\ell}_i)$. Both k and m will index times of observation (i.e. clinic visits) while the subscript i will be reserved for study subjects. Thus $\bar{\ell}_i$ is subject i's ℓ-history through t_K, $\bar{\ell}_{k,i}$ is subject i's ℓ-history through t_k, and $\bar{\ell}_k$ is some particular ℓ-history through t_k.

We adopt the convention that if two covariate or treatment histories, say, $\bar{\ell}_k$ and $\bar{\ell}_m$, are used in the same expression or definition with $t_m < t_k$, then $\bar{\ell}_m$ is the initial segment of $\bar{\ell}_k$ through t_m. Furthermore, we shall use superscripted integers in parentheses when we wish to represent possibly distinct covariate histories. For example, if $\bar{\ell}_m^{(1)}$, $\bar{\ell}_k^{(1)}$, and $\bar{\ell}_k^{(2)}$ were used in the same expression, $\bar{\ell}_m^{(1)}$ would be the initial segment of $\bar{\ell}_k^{(1)}$ but not necessarily of $\bar{\ell}_k^{(2)}$. We suppose $a_{k,i}$ and $\ell_{k,i}$ lie in sets $\mathcal{F}_k$ and $\mathcal{L}_k$ of feasible a_k and ℓ_k values. Let $\bar{\mathcal{F}}_k$ and $\bar{\mathcal{L}}_k$ be the set of all vectors $(a_0,a_1,\cdots,a_k)$ and $(\ell_0,\ell_1,\cdots\ell_k)$ with $a_m \in \mathcal{F}_m$, $\ell_m \in \mathcal{L}_m$, $0 \le m \le k$.

2b. The Set of Feasible Treatment Regimes

A subject assigned to given regime may fail to comply and actually follow an AZT treatment history consistent with some other regime that may differ from any of the J assigned regimes. We shall restrict consideration to a set **G** of feasible treatment regimes, members of which will be denoted by G. The set of feasible regimes is formally defined below. First we provide an informal discussion.

The feasible treatment regimes will be classified into dynamic and non-dynamic regimes. A non-dynamic regime is, by definition, any regime that can be characterized by a single planned AZT treatment history. Specifically, the non-dynamic regime G characterized by a specific AZT history $\bar{a} \equiv (a_0,\cdots,a_K)$, and written $G=(\bar{a})$, is a regime in which a subject is assigned to take at time t, $t \in (t_{k-1},t_k]$ the AZT dose $a_{k-1} \in A_k$ consistent with $\bar{a}$. A dynamic protocol or regime is, by definition, a protocol in which the dosage of AZT a subject is assigned to take at some time t is not known at time of randomization because the assigned dose depends on evolution of a time-dependent covariate. An example of a dynamic regime is "take 1,000 milligrams of AZT daily in the interval $(t_{k-1},t_k]$ if the subject's hematocrit measured at t_{k-1} exceeds 30. Otherwise, take no AZT in that interval." Note that any regime that includes a provision for modifying the dose of AZT in subjects who develop drug toxicity is dynamic.

Throughout we shall restrict attention to regimes in which the assigned AZT dosage rate is constant in intervals $(t_{k-1},t_k]$. As discussed in the previous sub-section all information potentially available to the investigators conducting the trial on outcome and covariate history on subject i at t_k is given by $\bar{\ell}_{k,i}$. We have assumed that the results of measurements made at t_k are immediately available. Therefore, for any feasible regime, the AZT dosage assigned in $(t_{k-1},t_k]$ may depend on and only on $\bar{\ell}_{k-1,i}$.

Formally, any feasible regime G is identified with a function "G"$(t_k,\bar{\ell}_k)$, defined for each t_k and $\bar{\ell}_k \in \mathcal{L}_k$, $0 \leq k \leq K-1$, whose value is the AZT dosage rate $a_k \in \mathcal{F}_k$ in the interval $(t_k,t_{k+1}]$ assigned by the regime to a subject with the history $\bar{\ell}_k$ through t_k. Conversely, any function "G"$(t_k,\bar{\ell}_k)$ whose range is a possible AZT dosage rate corresponds to some feasible regime. We use " " to distinguish the function "G" from the regime G it represents. Given a function "G"$(t_k,\bar{\ell}_k)$ of the two arguments $(t_k,\bar{\ell}_k)$, we can define an equivalent function "G"$(\bar{\ell}_k) \equiv$ {"G"$(t_m,\bar{\ell}_m)$; $0 \leq m \leq k$} which takes values in the set of possible AZT histories $\bar{a}_k \in \bar{\mathcal{F}}_k$. Let **G** represent the set of all feasible regimes. Note **G** depends on the covariates in ℓ_k. Let $G=(j)$, $j\epsilon(1,\cdots,J)$, be the regimes that were actually assigned. They must obviously be feasible and thus contained in **G**. The non-dynamic regimes are those treatment regimes for which "G"$[\bar{\ell}^{(1)}]$ = "G"$[\bar{\ell}^{(2)}]$ for all $\bar{\ell}^{(1)},\bar{\ell}^{(2)}$ where $\bar{\ell}^{(1)},\bar{\ell}^{(2)}$ are any two distinct ℓ-histories through t_K. The non-dynamic regimes are precisely the regimes of the form $G=(\bar{a})$ discussed previously.

2c. The Counter-Factuals

For any two regimes $G \in$ **G** and $G^*\epsilon(1,\cdots,J)$, we assume the existence of the ℓ-history $\bar{\ell}_{i,G,r=G^*} \equiv (\ell_{0,i,G,r=G^*},\cdots,\ell_{K,i,G,r=G^*})$ and AZT-history $\bar{a}_{i,G,r=G^*}$ that would have been observed if, possibly contrary to fact, subject i had been assigned to treatment regime G* but actually followed an AZT treatment history consistent with regime G, where $\bar{\ell}_{i,G,r=G^*}$ determines $\bar{a}_{i,G,r=G^*}$ through the relationship

$$\text{"G"}[t_k,\bar{\ell}_{k,i,G,r=G^*}] = a_{k,i,G,r=G^*} \tag{1}$$

Eq. (1) formalizes the idea that subject i followed regime G.

In addition we shall suppose that for all $G \in$ **G**, G^*, $G^{**}\epsilon(1,\cdots,J)$

$$\bar{\ell}_{i,G,r=G^*} = \bar{\ell}_{i,G,r=G^{**}} \equiv \bar{\ell}_{i,G} \qquad (2)$$

Eq. (2) says that ℓ-history depends on the treatment regime actually followed and not on treatment assignment. In a double-blind trial, one would expect (2) to hold, since subjects remain ignorant of their assigned treatment. Eq. (2) does not rule out a "placebo effect" of being entered in the trial in the sense that we do not require that $\bar{\ell}_{i,G}$, as defined in (2) equals the ℓ-history that subject i would have had if (a) subject i had never been entered in the trial but (b) had followed a AZT treatment history consistent with regime G. Note by assumption $\bar{\ell}_{i,G}$ does not depend on the treatment that any other subject either was assigned or took (Rubin, 1978).

We now define the "counterfactual data" to be $\bar{\ell}_{i,\mathbf{G}} = \{\bar{\ell}_{i,G}\ ;\ G\epsilon\mathbf{G}\}$. We shall make two consistency assumptions that formalize the idea that a subject's ℓ-history through t_m depends only on AZT treatment received prior to t_m. Consistency assumption (a) serves to link the counterfactual data with the observed data. Consistency assumption (b) serves to link the counterfactual data associated with various G.

Consistency Assumption (a): For any regime G for which "G"$[\bar{\ell}_{m,i}] = \bar{a}_{m,i}$, we assume $\bar{\ell}_{m+1,i,G} = \bar{\ell}_{m+1,i}$.

It follows from consistency assumption (a) that, for subject i, the counterfactual data $\bar{\ell}_{i,G}$ is actually observed for those G such that "G"$(\bar{\ell}_i) = \bar{a}_i$. $\bar{\ell}_{m+1,i,G}$ is observed for those regimes G such that "G"$[\bar{\ell}_{m,i}] = \bar{a}_{m,i}$ holds. For any such G, $\bar{\ell}_{m+1,i,G} = \bar{\ell}_{m+1,i}$. In particular,

$$\ell_{0,i} = \ell_{0,i,G} \qquad (3)$$

for all regimes G.

Consistency Assumption (b): Given a regime G* and subject i, let $\bar{a}^{(1)}$ be an AZT history through t_K such that $\bar{a}_m^{(1)} = \bar{a}_{m,i,G^*}$. Then, by assumption, $\bar{\ell}_{m+1,i,G=(\bar{a}^{(1)})} = \bar{\ell}_{m+1,i,G^*}$.

This assumption implies that if a subject would have the same AZT history through t_{m-1} when following regime G* or $G = (\bar{a}^{(1)})$, then the subject would have the same ℓ-history through t_{m-1} under both regimes.

2d. The Introduction of Randomness

We shall assume that the $(r_i, \bar{\ell}_i, \bar{a}_i, \{\bar{\ell}_{i,G}, \bar{a}_{i,G}\ :\ G\epsilon\mathbf{G}\})$ are realizations of independent and identically distributed random vectors $(R_i, \bar{L}_i, \bar{A}_i, \{\bar{L}_{i,G}, \bar{A}_{i,G}\ :\ G\epsilon\mathbf{G}\})$. Therefore, for notational convenience, we shall often drop the subscript i when referring to these random variables. We shall suppose that the density $f_{\bar{A}_K,\bar{L}_K}(\bar{a}_K, \bar{\ell}_K) \neq 0$ for all $\bar{a}_K \epsilon \bar{\mathcal{A}}_K, \bar{\ell}_K \epsilon \bar{\mathcal{L}}_K$.

2e. The Maintained Assumption

Our maintained assumption will be that the counterfactual data are independent of treatment assignment, i.e.,

$$\bar{L}_G \amalg R \tag{4}$$

where $A \amalg B \mid C$ means A is conditionally independent of B given C (Dawid, 1979). In view of (3), (4) implies that $\bar{L}_G \amalg R \mid L_0$. Eq. (4) will hold since (1) the $\bar{\ell}_{i,G}$ are conceptually "pretreatment variables" in the sense that their value depends neither on the treatment to which subject i was assigned nor on the AZT treatment subject i actually received, and (2) if, as we shall assume, a completely randomized design was employed then the random variable R is independent of any pretreatment variable.

2f. Goals of analysis

Estimation: One estimation goal would be to identify and estimate the J regime-specific mean T4-count histories $E[\bar{Y}_{G=(j)}]$ from the observables $(R_i, \bar{L}_i, \bar{A}_i)$, where, by definition, a parameter is identifiable if it is a function of the joint distribution of the observables. A more ambitious goal is to identify and estimate $E[\bar{Y}_G]$ for all $G \in \mathbf{G}$.

In the absence of non-compliance, $E[\bar{Y}_{G=(j)}]$ is identifiable. This follows since, by the maintained assumption (4), $E[\bar{Y}_{G=(j)}] = E[\bar{Y}_{G=(j)} \mid R=j]$, and, in the absence of non-compliance, by consistency assumption (a), $E[\bar{Y}_{G=(j)} \mid R=j]$ equals the identifiable parameter $E[\bar{Y} \mid R=j]$. In the face of non-compliance, $E[\bar{Y}_{G=(j)} \mid R=j]$ will not be identifiable without further assumptions.

Testing: We define the sharp null hypothesis of no effect of AZT on evolution of T4-count history to be the hypothesis

$$\bar{Y}_{G^{(1)}} = \bar{Y}_{G^{(2)}} \text{ for all i and } G^{(1)}, G^{(2)} \in \mathbf{G} \tag{5}$$

The sharp null hypothesis implies

$$E\left[\bar{Y}_{G^{(1)}}\right] = E\left[\bar{Y}_{G^{(2)}}\right] \text{ for all } G^{(1)}, G^{(2)} \in \mathbf{G} \tag{6}$$

which we call the G-null mean hypothesis. Under consistency assumption (a) and maintained assumption (4), Eq. (5) also implies

$$E[\bar{Y} \mid R=j^{(1)}] = E[\bar{Y} \mid R=j^{(2)}] \text{ for all } j^{(1)}, j^{(2)} \tag{7}$$

which we call the "intention to treat null mean hypothesis" since it is the null hypothesis of interest in the usual intention to treat analysis of a trial designed to study the effect of AZT on mean T4 level. The standard intention-to-treat analysis of such a trial consists in constructing an asymptotically distribution-free test of Eq. (7). Neither Eq. (6) nor Eq. (7) implies the other. Our testing goals will be to derive asymptotically distribution free tests of these null hypotheses.

2g. The Data Recorded for Data Analysis

It is often the case that not all components of the vector ℓ_k of covariates measured at t_k are entered into a computer and recorded for subsequent data analysis. To allow for this possibility let z_k be a known subset of the components of ℓ_k. For reasons that will become clear in Section 3d, we suppose z_k may or may not contain T4-count y_k. We assume that for each k and subject i, $z_{k,i}$, $y_{k,i}$ are the only components of $\ell_{k,i}$ recorded for data analysis. Then we shall be interested in the question of identification and estimation of $E[\bar{Y}_G]$ from the observable random variables $(R_i, \bar{A}_i, \bar{Z}_i, \bar{Y}_i)$.

3. THE IDENTIFICATION AND ESTIMATION OF REGIME-SPECIFIC MEAN T4-HISTORY $E[\bar{Y}_G]$

3a. An Overview of Identification

The key to our approach to the identification and estimation of $E[\bar{Y}_G]$ from the observables $(R_i, \bar{A}_i, \bar{Z}_i, \bar{Y}_i)$ is a function which we shall call the <u>treatment mean transformation function with respect to rz</u>. To define this function we shall require additional notation.

We adopt the convention that "segmented" AZT-histories will be written "chronologically" in a single parenthesis separated by commas. Thus $(\bar{a}_{m-1}, 0)$ specifies a particular AZT history through t_m and no AZT exposure subsequent to t_m.

For any $k > m$, $\bar{a}_m$, $r, \bar{z}_m$, $G^* \in G$, <u>the treatment mean transformation function, $\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^*)$, with respect to rz is</u>

$$\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^*) \equiv E[Y_{k,G^*} \mid r, \bar{z}_m, \bar{a}_m] - E\left[Y_{k,G=(\bar{a}_{m-1},0)} \mid r, \bar{z}_m, \bar{a}_m\right]$$

where $E[Y_{k,G^*} \mid r, \bar{z}_m, \bar{a}_m] \equiv E[Y_{k,G^*} \mid (R, \bar{Z}_m, \bar{A}_m) = (r, \bar{z}_m, \bar{a}_m)]$. To clarify the meaning of $\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^*)$, consider the subset with history $r, \bar{z}_m, \bar{a}_m$ in the trial. Then $E\left[Y_{k,G=(\bar{a}_{m-1},0)} \mid r, \bar{z}_m, \bar{a}_m\right]$ is the mean T4-count at t_k if the subset received their observed AZT history up to t_m but no AZT from t_m onwards. $E[Y_{k,G^*} \mid r, \bar{z}_m, \bar{a}_m]$ is the mean T4-count at t_k for the same subset if the subset followed regime G^*. $\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^*)$ is the difference in these means.

For notational convenience, we define $\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^* = (\bar{a}_m, 0)) \equiv \gamma^{[rz]}(k, r, \bar{z}_m, \bar{a}_m)$. $\gamma^{[rz]}(k, r, \bar{z}_m, \bar{a}_m)$ is the effect of one final brief "blip" of AZT exposure a_m in the interval $(t_m, t_{m+1}]$ on mean T4-count at t_k for the subset of the superpopulation that would have had history $(r, \bar{z}_m, \bar{a}_m)$ if selected for the trial. Note that the magnitude of this final blip a_m is equal to the subset's observed AZT exposure at t_m. $\gamma^{[rz]}(k, r, \bar{z}_m, \bar{a}_m) = 0$ if $a_m = 0$.

Our key identification results can be described in terms of the functions $\phi^{[rz]}(k, r, \bar{z}_m, \bar{a}_m, G^*)$ and $\gamma^{[rz]}(k, r, \bar{z}_m, \bar{a}_m)$. Specifically, in Theorem 1 below, we prove that if the function $\gamma^{[rz]}(k, r, \bar{z}_m, \bar{a}_m)$ is identified, then $E[\bar{Y}_{G\equiv 0}]$ is identified where $G \equiv 0$ is the non-dynamic regime in which AZT is withheld at all times. The regime $G \equiv 0$ is the regime assigned in a "placebo" arm. Even when the actual trial does not contain a "placebo" arm, if $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ is identified, $E[\bar{Y}_{G\equiv 0}]$ is identified.

Unfortunately, it is straightforward to show that $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ is not identified from the observables $(R_i,\bar{A}_i,\bar{Z}_i,\bar{Y}_i)$ without additional assumptions beyond maintained assumption Eq. (4). We show in Section 3c that, under regularity conditions, the function $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ may be identified under the additional assumption that $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ is known up to a finite vector of unknown parameters. That is $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m) = \gamma(k,r,\bar{z}_m,\bar{a}_m,\psi_0)$ where $\gamma(\cdot,\cdot,\cdot,\cdot,\cdot)$ is a known function of five arguments and ψ_0 is an unknown vector of parameters ~~to be~~ estimated.

Even so, the identifiability of $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ plus the maintained assumption (4) is not sufficient to identify $E[\bar{Y}_G]$ for any G other than $G \equiv 0$. An additional non-identifiable assumption under which the identifiability of $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ implies the identification of $E[\bar{Y}_G]$ for certain G is the assumption of "no-current-treatment-interaction" with respect to rz as defined below. First let **G**(z) be the subset of **G** such that $G \in$ **G**(z) if and only if $"G"[\bar{\ell}^{(1)}] = "G"[\bar{\ell}^{(2)}]$ whenever $P_{\bar{z}}(\bar{\ell}^{(1)}) = P_{\bar{z}}(\bar{\ell}^{(2)})$ where $P_{\bar{z}}$ is the projection operator that selects from each ℓ_k those components that comprise z_k. Any $G \in$ **G**(z) can thus be represented by a function $"G"(\bar{z}_k)$ whose value is the AZT dosage history $\bar{a}_k$ associated with the z-history $\bar{z}_k$.

Next define $\phi^{|rz|}(k,r,\bar{z}_m,\bar{a}_{m-1},G^*) \equiv E[Y_{k,G^*} \mid r,\bar{z}_m,\bar{a}_{m-1}] - E\left[Y_{k,G=(\bar{a}_{m-1},0)} \mid r,\bar{z}_m,\bar{a}_{m-1}\right]$. Note $\phi^{|rz|}(k,r,\bar{z}_m,\bar{a}_{m-1},G^*)$ is defined just like $\phi^{|rz|}(k,r,\bar{z}_m,\bar{a}_m,G^*)$ but with $\bar{a}_{m-1}$ replacing $\bar{a}_m$.

Definition: We say that there is no-current-treatment-interaction with respect to rz for a particular $G^* \in$ **G**(z) if, for all m, $k > m, r, \bar{z}_m$,

$$\phi^{|rz|}(k,r,\bar{z}_m,"G^*"(\bar{z}_m),G^*) = \phi^{|rz|}(k,r,\bar{z}_m,"G^*"(\bar{z}_{m-1}),G^*) \tag{8}$$

To clarify the meaning of the assumption of no-current-treatment-interaction, consider the subset with observed history $[r,\bar{z}_m,"G^*"(\bar{z}_{m-1})]$. Then the assumption of no-current-treatment-interaction with respect to rz for regime G^* says that the effect, on a difference scale, of following regime G^* on mean T4-count at t_k compared to following regime $G=("G^*"(\bar{z}_{m-1}),0)$ is the same for the entire subset $[r,\bar{z}_m,"G^*"(\bar{z}_{m-1})]$ as for the sub-group of this subset defined by the AZT dosage $"G"(t_m,\bar{z}_m)$ at t_m.

In Theorem 4, we prove that if there is no-current-treatment-interaction with respect to rz for a particular $G \in$ **G**(z) and $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ is identified, then $E[\bar{Y}_G]$ is identified. It follows that an investigator can and should attempt to ensure no-current-treatment-interaction with respect to rz by recording in z_k all covariates that one believes could conceivably both predict (1) increments in AZT exposure at t_m given past AZT exposure and (2) the magnitude of regime-specific-treatment effects on a difference scale.

In summary, in order to identify $E[\bar{Y}_G]$ our task is to record in z_t data on a sufficient number of covariates to ensure no-current-treatment-interaction with respect to rz for all $G \in \mathbf{G}(z)$ and then to specify a realistic, flexible parametric model $\gamma(k,r,z_m,a_m,\psi_0)$ for the unknown function $\gamma^{(\alpha)}(k,r,z_m,a_m)$. If we succeed, $E[\bar{Y}_G]$ will be identified for $G \in \mathbf{G}(z)$ from the observables $(R_i, \bar{A}_i, \bar{Z}_i, \bar{Y}_i)$.

3b. Identification and Estimation of $E[\bar{Y}_{G=0}]$ when $\gamma^{(\alpha)}(k,r,z_m,a_m)$ Is Known

Our main result in this subsection is Theorem 1 below. Before stating the theorem we require the following additional definitions. Define $h_{k,i} \equiv y_{k,i} - \sum_{m=0}^{m=k-1} \gamma^{(\alpha)}(k,r_i,\bar{z}_{m,i},\bar{a}_{m,i})$. Define $\bar{h}_i = (h_{0,i},...,h_{K,i})$ with $\bar{H}_i$ the corresponding random variable. Note, if $\gamma^{(\alpha)}(k,r,z_m,a_m)$ is always zero, then $\bar{H}_i = \bar{Y}_i$.

Informally, $h_{k,i}$ can be interpreted as an "estimate" of $y_{k,i,G=0}$, since, in computing $h_{k,i}$ one subtracts from the observed $y_{k,i}$ the "average effect" of each of the observed treatments $(a_{0,i},...,a_{k-1,i})$. This interpretation is formalized in the following Theorem proven in Corollary A.1 of Appendix 1.

Theorem 1: $E[\bar{Y}_{G=0}] = E[\bar{H}]$.

Note Theorem 1 does not require that the maintained assumption (4) holds. If the function of $\gamma^{(\alpha)}(k,r,z_m,a_m)$ were known, then $\bar{H}_i$ would be an observable and $E[\bar{Y}_{G=0}]$ would be identified.

3c. Identification and Estimation of $\gamma^{(\alpha)}(k,r,z_m,a_m)$

We next show that, under regularity conditions, the function $\gamma^{(\alpha)}(k,r,z_m,a_m)$ is identified if the population follows the multivariate SNMM $\gamma(k,r,z_m,a_m,\psi_0)$ with respect to rz. By definition, the population follows the multivariate SNMM $\gamma(k,r,z_m,a_m,\psi_0)$ with respect to rz if $\gamma^{(\alpha)}(k,r,z_m,a_m) = \gamma(k,r,z_m,a_m,\psi_0)$ and the function $\gamma(\cdot,\cdot,\cdot,\cdot,\cdot)$ is a known function of 5 arguments, ψ_0 is an unknown V-dimensional vector of parameters, $\gamma(k,r,z_m,a_m,\psi) = 0$ if $\psi = 0$ or $a_m = 0$, and the ψ-derivatives of $\gamma(k,r,z_m,a_m,\psi)$ are continuous in ψ.

Example 1: $\gamma(k,r,\bar{z}_m,\bar{a}_m,\psi_0) = \psi_{0,1}a_m + \psi_{0,2}a_m y_m + \psi_{0,3}(r-1)a_m + \psi_{0,4}a_{m-1}a_m + \psi_{0,5}w_m a_m + \psi_{0,6}a_m(t_k-t_m)$ is a SNMM with respect to rz, where $\psi_0' = (\psi_{0,1},\psi_{0,2},\psi_{0,3},\psi_{0,4},\psi_{0,5},\psi_{0,6})$, and we suppose $z_m' = (y_m', w_m')$, w_m is hematocrit at t_m and $r \in \{1,2\}$. This model says that magnitude of the effect of a brief blip of exposure to AZT in $(t_m, t_{m+1}]$ on T4-count at t_k may depend on current T4-count and hematocrit, AZT exposure at t_{m-1}, treatment arm r, and $t_k - t_m$. If, as is usually the case, subjects with history (z_m, a_m) on treatment arm $r=1$ are not comparable to subjects with that same history on treatment arm $r=2$, one would expect that $\gamma(k,r,z_m,a_m,\psi_0)$ might, as in our example, depend on r.

Our main result is Theorem 2 below. Following the proof of the theorem we discuss its implications for identification of ψ_0 and thus of $\gamma^{|\alpha|}(k,r,z_m,a_m)$. Let $X_i \equiv (Y_{0,i},Z_{0,i})$ where $Y_{0,i},Z_{0,i}$ are the "pre-treatment" observables recorded for data analysis.

Theorem 2: Under maintained assumption Eq. (4),

$$E[\bar{H}_i \mid R_i,X_i] = E[\bar{H}_i \mid X_i] \tag{9}$$

Proof: In Theorem A. 1 of Appendix 1 we show that $E[\bar{Y}_{i,G=0} \mid R_i,X_i] = E[\bar{H}_i \mid R_i,X_i]$ regardless of whether Eq. (4) holds. If Eq. (4) holds, $E[\bar{Y}_{i,G=0} \mid R_i,X_i] = E[\bar{Y}_{i,G=0} \mid X_i]$ proving the theorem. Eq. (9) represents the sole restriction on the function $\gamma^{|\alpha|}(k,r,z_m,a_m)$ attributable to maintained assumption (4).

Corollary 1: If Eq. (4) holds, and $\gamma^{|\alpha|}(k,r,z_m,a_m) = \gamma(k,r,z_m,a_m,\psi_0)$, then (1)

$$E[\bar{H}_i(\psi_0) \mid R_i,X_i] = E[\bar{H}_i(\psi_0) \mid X_i] \quad . \tag{10}$$

and (2) $\psi_0 = 0$ implies the intentions to treat null mean hypothesis Eq. (7).

Proof: Conclusion (1) follows immediately from Eq. (9). Conclusion (2) follows from conclusion (1) and the fact that $\bar{H}_i(0) = \bar{Y}_i$. Eq. (10) is the sole equality restriction on the observable random variables implied by the supposition on Corollary 1. Eq. (4) implies additional inequality restrictions, but these do not provide information about ψ_0 asymptotically.

3c.1 A Class of Estimators

Since $H_{0,i}(\psi)$ is the pretreatment variable $Y_{0,i}$, for all ψ, it will be convenient to redefine $\bar{H}_i(\psi)$ to be the K-vector $(H_{1,i}(\psi),\cdots,H_{K,i}(\psi))$. Let $\bar{q}(x)$ be a non-random function of x taking values in R^K. Let $g(r,x)$ be a non-random $K\times v$ dimensional matrix valued function of (r,x). Define $c(r,x) = (\bar{q}(x),g(r,x))$ and $c \equiv (\bar{q},g)$. Let $\hat{\psi}(\bar{q},g) \equiv \hat{\psi}(c)$ be a solution to the v-dimensional estimating equation $0 = n^{-1/2}S(\psi,c) \equiv n^{-1/2}\Sigma_i U_i(\psi,c)$ where

$$U_i(\psi,c) = d'(R_i,X_i)[\bar{H}_i(\psi) - \bar{q}(X_i)] \quad . \tag{11}$$

$$d(R_i,X_i) = g(R_i,X_i) - E[g(R_i,X_i) \mid X_i] \quad . \tag{12}$$

Since

$$p[R_i=j \mid X_i]=1/J,\ j\epsilon\{1,\cdots,J\} \quad . \tag{13}$$

we can explicitly calculate the expectations in Eq. (12). Therefore for each value of ψ, $n^{-1/2}\sum_i U_i(\psi,c)$ is an observable random variable.

Theorem 3: Under regularity conditions as in Chapters 7 and 8 of Manski, 1988, there exists a solution $\hat{\psi}(c)$ to $n^{-1/2}S(\psi,c) = 0$ such that $n^{1/2}(\hat{\psi}(c)-\psi_0)$ has a limiting normal distribution with mean 0 and asymptotic variance

$$\tau^{-1}\,\Omega(\psi_0,c)\tau'^{-1} \tag{14}$$

provided $\tau \equiv E[\partial U_i(\psi_0,c)/\partial\psi']$ is invertible where $\Omega(\psi_0,c)$ is $Var\{U_i(c)\}$ with $U_i(c) = U_i(\psi_0,c)$.

Proof: The theorem follows from general results on estimating equations since, by (10),

$$E[U_i(\psi_0,c) \mid X_i] = 0 \tag{15}$$

For use later, define

$$\bar{q}_{opt}(X_i) \equiv E[\bar{H}_i(\psi_0) \mid X_i] \ . \tag{16}$$

$\epsilon_i(\psi) = \bar{H}_i(\psi) - \bar{q}_{opt}(X_i)$, and $\epsilon_i = \epsilon_i(\psi_0)$. Therefore,

$$\bar{H}_i(\psi_0) - \bar{q}(X_i) = \epsilon_i + q_{opt}(X_i) - \bar{q}(X_i) \tag{17}$$

It follows from Eq. (10) that

$$E[\epsilon_i \mid R_i, X_i] = 0 \ . \tag{18}$$

Note that

$$Var[U_i(c)] = plim[n^{-1} \textstyle\sum_i \{U_i(\psi_0,c) \cdot U_i(\psi_0,c)'\}] \tag{19}$$

and

$$\tau \equiv E\left[\partial U_i(\psi_0,c)/\partial\psi'\right] = plim\left[n^{-1} \textstyle\sum_i d(R_i,X_i)'\partial\bar{H}_i(\psi_0)/\partial\psi'\right] \tag{20}$$

A consistent estimator of the asymptotic variance Eq. (14) is given by substituting $\hat{\psi}(c)$ for ψ_0 on the right-hand side of Eqs. (19) and (20) and ignoring the "plim."

3c.2 A Semiparametric Efficient Estimator

Define $w_{opt}(R_i,X_i) = \{E[\epsilon_i\epsilon_i' \mid R_i,X_i]\}^{-1}$ and $\mu_{opt}(R_i,X_i) = E\left[\partial\bar{H}_i(\psi_0)/\partial\psi' \middle| R_i,X_i\right]$; and $g_{opt}(R_i,X_i) \equiv w_{opt}(R_i,X_i)\{\mu_{opt}(R_i,X_i)-E[w_{opt}(R_i,X_i) \mid X_i]^{-1}E[w_{opt}(R_i,X_i)\mu_{opt}(R_i,X_i) \mid X_i]\}$. [Note: $E\{g_{opt}(R_i,X_i) \mid X_i\} = 0$.] Define $c_{opt} = (\bar{q}_{opt},g_{opt})$. Then $Var^{A}[n^{1/2}(\hat{\psi}(c_{opt})-\psi_0]$ attains the semiparametric efficiency bound in the sense of Begun et al., (1983) for estimating ψ_0 under the sole restriction Eq. (10). This follows because $n^{-1/2}S(\psi_0,c_{opt})$ is the "efficient score" for ψ_0 under the restriction Eq. (10) derived in Section 3 of Chamberlain (1988). The additional restriction that $p[R_i=j \mid X_i]=1/J,\ j\epsilon\{1,\cdots,J\}$ has no effect on the "efficient score" since it is a restriction on the joint distribution of (X_i,R_i) which are S-ancillary for ψ_0. It follows from Eq. (14) and the weak law of large numbers that the semiparametric efficiency bound is

$$Var^{A}[n^{1/2}(\hat{\psi}(c_{opt})-\psi_0)] = \left\{ E\left[\frac{\partial U_i(\psi_0,c_{opt})}{\partial\psi'} \right] \right\}^{-1} \tag{21}$$

(Chamberlain 1987, 1988). It follows from the theory of semiparametric efficiency bounds that no estimator which is regular under the sole restriction (10) can have asymptotic variance less than (21).

3c.3 Identification

Eq. (21) has implications for the identification of ψ_0. Under the regularity conditions in Chapters 7-8 of Manski (1988), a necessary and sufficient condition for ψ_0 be locally identified is that $E[\partial U_i(\psi_0,c_{opt})/\partial\psi']$ be invertible (Manski, 1988). if, in addition, $\bar{H}_i(\psi)$ is linear in ψ and $E[\partial U_i(\psi_0,c)/\partial\psi']$ is invertible, then ψ_0 is globally identified since, then, $E[n^{-1/2}S(\psi,c)] = 0$ only if $\psi = \psi_0$.

As an illustration suppose that the p^{th} column of $E[\partial\bar{H}_i(\psi_0)/\partial\psi' | R_i,X_i]$ does not depend on R_i. Then the p^{th} column of $g_{opt}(R_i,X_i)$ is identically zero, $E[\partial U_i(\psi_0,c_{opt})/\partial\psi']$ is not invertible and ψ_0 will not be identified. As second illustration, consider the special case in which there exists no data on pretreatment variables. (So X_i has a single realization.) Then it is easy to see that $E[\partial U_i(\psi_0,c_{opt})/\partial\psi']$ cannot be of full rank if the number of parameters v in ψ_0 is greater than the product of K and (J-1) where J is the number of treatments arms. Later, we consider additional restrictions under which ψ_0 is identified even if $v > (J-1)K$.

3c.4 Additional Remarks

Since $c_{opt}(R_i,X_i)$ depends on unknown population parameters, the estimator $\hat{\psi}(c_{opt})$ is not feasible. But a partially adaptive estimation is possible if we hypothesize that $c_{opt}(R_i,X_i)$ is known up to a finite vector of unknown parameters θ_0 and estimate θ_0. It follows from our results that, under Eq. (10), for any fixed c, an asymptotically distribution free test of the null hypothesis $\psi_0 = 0$ (and thus of the intention-to-treat null mean hypothesis) can be obtained by setting $\psi = 0$ and comparing $[n^{-1/2}S(\psi,c)']\{\hat{Var}[n^{-1/2}S(\psi,c)]\}^{-1}[n^{-1/2}S(\psi,c)]$ to a χ^2_v distribution with the variance estimator based on (19). In this sense, our approach is consistent with the usual-intention-to-treat analysis described in Section 2g. Since $\psi_0 = 0$ neither implies nor is implied by the G-null hypothesis Eq. (6), this test is not an asymptotically distribution free test of Eq. (6) without additional assumptions.

3c.5 An Expanded Class of Estimators

In this section, following Robins, Mark, and Newey (1992), we first show that, for a fixed c, given an inefficient estimator $\hat{\psi}(c)$ we can use models for the known probabilities $p[R=r | X_i]$ to produce a modified estimator $\tilde{\psi}(c)$ that is guaranteed to be at least as efficient as $\hat{\psi}(c)$. We now allow R to be continuous or discrete. We redefine $p[R=r | X_i]$ to be $f[r | X_i]$ where $f[r | X_i]$ is the density or probability function of the random variable R conditional on X_i. Then, let $f[r | X_i ; \alpha]$ be a parametric model for $f[r | X_i]$, such that for a unique α_0, we have $f[r | X_i;\alpha_0] = f[r | X_i]$. Note that in our randomized trial, this implies that $f[r | X_i;\alpha_0] = 1/J$. Let $\hat{\alpha}$ be the maximum likelihood estimator of α based on the observables (X_i,R_i) solving $\Sigma_i S_{\alpha,i}(\alpha) = 0$ with $S_{\alpha,i}(\alpha) = \partial \ell og f(R_i | X_i ; \alpha)/\partial\alpha$. For a given choice of c, define $\tilde{\psi}(c) \equiv \hat{\psi}(c,\hat{\alpha})$ to be a solution to $0=n^{-1/2}S(\psi,c,\hat{\alpha})$ where $S(\psi,c,\alpha)$ is defined like $S(\psi,c)$ except $d(R_i,X_i)$ is replaced by $d(R_i,X_i,\alpha) \equiv g(R_i,X_i)-\{\int g(r,X_i)\cdot dF(r | X_i,\alpha)\}$. Note $\hat{\psi}(c,\alpha_0) = \hat{\psi}(c)$. Thus one can view $\dot{\psi}(c,\dot{\alpha})$ as the estimator obtained when we ignore the fact that we know the truth α_0 and, instead, substitute $\dot{\alpha}$ for α_0.

It follows from Newey (1990) that, under the regularity conditions in Chapters 7-8 of Manski (1988), $n^{1/2}(\dot{\psi}(c,\dot{\alpha})-\psi_0)$ is asymptotically normal with mean 0 and $\text{Var}^{A}[n^{1/2}(\dot{\psi}(c,\dot{\alpha})-\psi_0)] = \tau^{-1}\Omega^*(\psi_0,c)\tau^{-1\prime}$ where $\Omega^*(\psi_0,c) = \text{Var}\{\text{Resid}(U_i(c),S_{\alpha,i})\}$, $S_{\alpha,i} = S_{\alpha,i}(\alpha_0)$, $\text{Resid}(A,B) = A-E(AB')\{E(BB')\}^{-1}B$. $\dot{\Omega}^*(\psi,c) \equiv n^{-1}\Sigma_i \dot{\text{R}}\text{esid}(U_i(\psi,c),S_{\alpha,i}(\dot{\alpha}))^{\otimes 2}$ is consistent for $\Omega^*(\psi_0,c)$ where $\dot{\text{R}}\text{esid}(A_i,B_i)$ is the residual for subject i from the least squares regression of A_i on B_i, $i = 1,\ldots,n$. It follows that $\dot{\psi}(c) \equiv \psi(c,\dot{\alpha})$ is always at least as efficient as $\psi(c) = \psi(c,\alpha_0)$.

Consider an observational study and suppose, based on subject matter knowledge, we are willing to assume that Eq. (10) held for some possibly continuous random variable R. Then, even though $f[r \mid X_i]$ will not be known, we will still be able to $n^{1/2}$-consistently estimate the parameters of our SNMM if our model for $f[r \mid X_i ; \alpha]$ is correctly specified. Indeed, as discussed in Newey (1990), if we can produce non-parametric density estimates $\hat{f}(r \mid X_i)$ of $f(r \mid X_i)$ such that the associated estimator $\hat{d}(R_i,X_i)$ converges at a rate of $n^{1/4}$ or greater to $d(R_i,X_i)$, then, under regularity conditions, the above procedure will produce $n^{1/2}$ consistent estimates of ψ_0.

3d. Identification and Estimation of $E[\bar{Y}_G]$

Our main result in this section is the following theorem proved in Appendix 1.

Theorem 4: If there is no-current-treatment-interaction with respect to rz for a particular $G \epsilon \mathbf{G}(z)$, then $E[Y_{k,G} \mid R=r] - E[Y_{k,G=0} \mid R=r] =$

$$\int\int\cdots\int \sum_{m=0}^{m=k-1} \gamma^{[rz]}(k,r,\bar{z}_m,\bar{a}_m) \cdot \prod_{m=0}^{m=k-1} dF[z_{m+1} \mid r,\bar{z}_m,\bar{a}_m] \cdot dF(z_0 \mid r) \qquad (22)$$

where $\bar{a}_m = \text{"}G\text{"}(\bar{z}_m)$.

It follows that if $\gamma^{[rz]}(k,r,z_m,a_m)$ is known then, under the supposition of Theorem 4, $E[Y_{k,G} \mid R=r] - E[Y_{k,G=0} \mid R=r]$ is identified based on the observables $(R_i,\bar{A}_i,\bar{Z}_i,\bar{Y}_i)$ (and, therefore, $E[\bar{Y}_G]-E[\bar{Y}_{G=0}]$ is also identified) even if the maintained assumption (4) did not hold. But, if $\gamma^{[rz]}(k,r,z_m,a_m)$ is known, $E[\bar{Y}_{G=0}]$ is identified by Theorem 1, and thus $E[\bar{Y}_G]$ will be identified. The assumption of no-current-treatment-interaction with respect to rz is, in principle, testable in the sense, that under maintained assumption (4), it places a further restrictions on the joint distribution of the observables beyond Eq. (9). Specifically, it implies that the entire integral expression (22) is not a function of r, although each term within the integral may depend on r. Thus, we would not expect $\dot{\psi}(c_{opt})$ to remain semiparametric efficient.

It follows from Eq. (22) that, with $z_t = \ell_t$, if there is no-current-treatment-interaction with regards to rz for all $G \epsilon \mathbf{G}$, then $\gamma^{[rz]}(k,r,z_m,a_m)$ being identically zero implies the G-null mean hypothesis Eq. (6), and thus the test of $\psi_n=0$ of subsection 3C.4 is an asymptotically distribution-free test of Eq. (6).

Even if the densities $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ and the function $\gamma^{[z]}(k, r, z_m, \bar{a}_m)$ were known for all j, Eq. (22) cannot be evaluated analytically. A consistent estimator is obtained by the following simple Monte Carlo algorithm.

For $v=\{1,\cdots,V\}$

Step 1: Draw $z_{0,v}$ from $f[z_0 \mid r]$

Step 2: For $m\epsilon\{1,\cdots,K\text{-}1\}$, recursively draw $z_{m,v}$ from

$f[z_m \mid r, z_{m-1,v}, \text{"}G\text{"}(z_{m-1,v})]$.

Then, for $k\epsilon\{1,\cdots,K\}$, consistent estimates of $E[Y_{k,G} \mid R=r] - E[Y_{k,G=0} \mid R=r]$ are obtained as $V\to\infty$ by

$$V^{-1} \sum_{v=1}^{V} \sum_{m=0}^{k-1} \gamma^{[z]}(k, r, \bar{z}_{m,v}, \bar{a}_{m,v}) \text{ where } \bar{a}_{m,v} = \text{"}G\text{"}(\bar{z}_{m,v})$$

Of course in practice the function $\gamma^{[z]}(k, r, z_m, \bar{a}_m)$ and the densities $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ are unknown and must be estimated. In practice, under maintained assumption 4, we will estimate $\gamma^{[z]}(k, r, z_m, \bar{a}_m)$ by specifying a SNMM $\gamma(k, r, z_m, \bar{a}_m, \psi_0)$ and estimating ψ_0 by $\hat{\psi}(c)$ as discussed in the preceding section. Suppose that, for all k, the dimension of z_k is large so that, in practice, our only choice is to sample from a parametric estimate of $f[z_{j+1} \mid r, z_j, \bar{a}_j]$. One approach to estimating the densities $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ is to specify a parametric model $f[z_{j+1} \mid r, z_j, \bar{a}_j\,;\, \theta_0]$ depending on a vector of unknown parameters θ_0 and estimate θ_0 by $\hat{\theta}$ maximizing $\prod_i \prod_{m=0}^{K-1} f[z_{m+1,i} \mid r_i, \bar{z}_{m,i}, \bar{a}_{m,i}\,;\,\theta] \cdot f[z_{0,i};\theta]$. However, we show in Appendix 2 that if z_k contains y_k, our parametric model for $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ may be incompatible with the joint assumption that (4) holds and our parametric model for $\gamma^{[z]}(k, r, z_m, \bar{a}_m)$ is correctly specified. Furthermore, even if z_k does not contain y_k, our parametric model for $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ may be incompatible with the joint assumption that (4) holds, the supposition of Theorem 4 is true, and our model for $\gamma^{[z]}(k, r, z_m, \bar{a}_m)$ is correctly specified, since this joint assumption implies (22) does not depend on r. Under additional assumptions given in Section 5 below, these incompatibility problems can be solved by using the "simulation algorithm" and the likelihood function $\mathcal{L}(\psi, \theta)$ provided in Appendix 4.

3e. Discarding unnecessary covariates

It may be of interest to reduce the dimensionality of the problem by finding a subset w_k of the covariates in z_k that are still sufficient to identify $E[\bar{Y}_{G=(\bar{a})}]$. Now, providing we specify a flexible SNMM, $E[\bar{Y}_{G=(\bar{a})}]$ will be identified from data in w_k provided there is no-current-treatment-interaction w.r.t. rw. Specifically, for each k, let w_k be a subset of the components of z_k. Let $\mathbf{G}(w)$, $P_{\bar{w}}$, $\gamma^{[rw]}(k,r,\bar{w}_m,\bar{a}_m)$, and "no-current-treatment-interaction with respect to rw" be defined as above with w replacing z. Note $\mathbf{G}(w) \subset \mathbf{G}(z)$ and any $G \in \mathbf{G}(w)$ can be represented by a function "G"$(\bar{w}_m)$. In Appendix 1, we prove

Theorem 5: If there is no-current-treatment-interaction with respect to rz for $G \in \mathbf{G}(w)$ and either

(1)

$$A_m \coprod \bar{Z}_m \mid \text{"}G\text{"}(\bar{W}_{m-1}), \bar{W}_m, R \quad , \tag{23}$$

or

2(a)

$$\gamma^{[rw]}(k,r,z_m^{(1)},\bar{a}_m) = \gamma^{[rw]}(k,r,z_m^{(2)},\bar{a}_m) \tag{24}$$

whenever $P_{\bar{w}}(z_m^{(1)}) = P_{\bar{w}}(z_m^{(2)})$ and

(2b)

$$W_{m+1} \coprod \bar{Z}_m \mid \text{"}G\text{"}(\bar{W}_m), \bar{W}_m, R \tag{25}$$

then there is no-current-treatment-interaction with respect to rw for G.

Note that Eqs. (23), (24) and (25) can be subjected to empirical tests.

4. ASSUMPTIONS GUARANTEEING NON-PARAMETRIC IDENTIFICATION AND OBSERVATIONAL ESTIMATES

To use the approach introduced in this paper, we must assume there be no-current-treatment-interaction with respect to rz. No matter how many covariates are recorded in z_k, one can never be sure that one has collected data on all the necessary covariates. Nevertheless, we would expect, that by polling a number of patients as to what factors influence their decision, we could obtain a near exhaustive list of all determinants of actual AZT treatment and try to record the requisite data.

An apparently more disturbing aspect of our approach is that even if the assumption of no-current-treatment-interaction w.r.t. rz is met, differences in regime-specific mean T4-count that would be observed under complete compliance and, more generally, $E[\bar{Y}_G]-E[\bar{Y}_{G=0}]$ for $G \in \mathbf{G}(z)$ are still not non-parametrically identified from the observables $(R_i, \bar{A}_i, \bar{Z}_i, \bar{Y}_i)$. Rather, these differences are identified only if the population is known to follow a multivariate SNMM $\gamma(k,r,\bar{z}_m,\bar{a}_m,\psi_0)$ with respect to rz. In fact, non-parametric identification of $E[\bar{Y}_G]-E[\bar{Y}_{G=0}]$ is possible under an additional plausible assumption. Suppose, as discussed above, we succeeded in recording in z_k data on essentially all determinants of AZT dosage. Then we would expect for all $a_m^{(1)}, a_m^{(2)}, r, \bar{z}_m, \bar{a}_{m-1}$

$$E\left[Y_{k,G=(\bar{a}_{m-1},0)} \mid r, \bar{z}_m, \bar{a}_{m-1}, a_m^{(1)}\right] = E\left[Y_{k,G=(\bar{a}_{m-1},0)} \mid r, \bar{z}_m, \bar{a}_{m-1}, a_m^{(2)}\right] \tag{26}$$

would be true. That is, the T4-count at t_k that would be observed in the absence of further AZT treatment after t_m is mean independent of a_m among subjects with history $r, \bar{z}_m, \bar{a}_{m-1}$. We then have

Theorem 6: If Eq. (26) is true and for all $G \in G(z)$, there is no current-treatment interaction w.r.t. rz, then $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ and $E[Y_G]$ for $G \in G(z)$ are identified from $(R_i,\bar{A}_i,\bar{Z}_i,\bar{Y}_i)$. Specifically, for $G \in G(z)$ and $m<k$, and all $r,\bar{z}_m,\bar{a}_m$ and with $\prod_a^b = 1$ if $b<a$.

$$E[Y_{k,G} \mid r,\bar{z}_m,\bar{a}_{m-1} = {}^{*}G^{*}(\bar{z}_{m-1}),a_m] = E[Y_{k,G} \mid r,\bar{z}_m,\bar{a}_{m-1} = {}^{*}G^{*}(\bar{z}_{m-1})] \tag{27}$$

$$E[Y_{k,G} \mid r,\bar{z}_m,\bar{a}_{m-1} = {}^{*}G^{*}(\bar{z}_{m-1})] = \int\int\cdots\int E[Y_k \mid r,\bar{z}_{k-1},{}^{*}G^{*}(\bar{z}_{k-1})] \prod_{j=m}^{j=k-2} dF[z_{j+1} \mid \bar{z}_j,{}^{*}G^{*}(\bar{z}_j),r] \tag{28}$$

$$E[H_k \mid \bar{z}_m,\bar{a}_m,r] = E[H_k \mid \bar{z}_m,\bar{a}_{m-1},r] \tag{29}$$

Proof: Eq. (27) is an immediate consequence of the supposition of Theorem 6. In Appendix 1, we prove (27) implies (28). (27) and (28) immediately imply that $E[Y_G]$ and $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m)$ are identified. Eq. (29) is an immediate consequence of (27) and Theorem (A.1). Eq. (28) is the G-computation algorithm formula given in Robins (1989, Sec. 13).

The equality restrictions on the observables implied by (4), (27), and our SNMM w.r.t. rz. are (i) Eq. (10), (ii) for $0 \le j<k$,

$$E[H_k(\psi_0) \mid \bar{z}_j,\bar{a}_j,r] = E[H_k(\psi_0) \mid \bar{z}_j,\bar{a}_{j-1},r] \quad , k = (1,\ldots,K) \tag{30}$$

and (iii) that Eq. (28) does not depend on r when $m = -1$.

<u>Remark 1</u>: <u>SNMM versus G-computation algorithm</u>

Since, in view of the G-computation algorithm formula Eq. (28), when the suppositions of Theorem 6 are true, one can consistently estimate $E[\bar{Y}_G]$, $G \in \mathbf{G}(z)$ based on correctly specified models for $f(Z_{j+1} \mid R,\bar{Z}_j,\bar{A}_j)$ (even if Eq. (4) is false), one might ask why, in this setting, use SNMMs at all. [Robins (1989, Sec.5) refers to this estimator as the Monte Carlo G-computation algorithm estimator.] We offer three reasons. First, if the effect of a final blip of treatment does not depend on $\bar{z}_m$ in the sense that $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m) = \gamma^{|rz|}(k,r,\bar{a}_m)$ independent of $\bar{z}_m$, it then follows from Eq. (22) that $E[Y_{k,G=\{\bar{a}_t\}} \mid r=R] - E[Y_{k,G=0} \mid R=r] = \sum_{m=0}^{k-1} \gamma^{|rz|}(k,r,\bar{a}_m)$ can be consistently estimated using SNMMs without having to perform the integration in (22) or (28). A second reason to use SNMMs has to do with the ability to specify SNMMs that incorporate specific biological assumptions. Suppose that it is known, based on biological considerations, that a latent interval of 3 periods must occur before a treatment applied at t_m can affect the outcome. This implies that $\gamma^{|rz|}(k,r,\bar{z}_m,\bar{a}_m) = 0$ for $k<m+3$, a restriction that can be straightforwardly incorporated in a SNMM. For example, consider the SNMM $\gamma(k,r,\bar{z}_m,\bar{a}_m,\psi_0) = \psi_{10}a_m I(k>m+3) + \psi_{20}a_m I(k \le m+3)$. Then there is a maximum latent interval of 3 periods (i.e. a latent interval of 3 but not 4 periods) if and only if $\psi_{20} = 0$ and $\psi_{10} \ne 0$. Similarly, for any SNMM, the G-null mean hypothesis is true if and only if $\psi_0 = 0$. The only simple restriction on $f(Z_{j+1} \mid R,\bar{Z}_j,\bar{A}_j)$ implied by a 3 period maximum latent interval is that $f(Y_{j+1} \mid R,\bar{Z}_j,\bar{A}_j)$ does not depend on A_j where, by a simple restriction, we mean a restriction not specified by an integral equation. But this is also the only simple restriction implied by a maximum latent interval of 1 period and by the G-null mean hypothesis. Thus, there is no natural way to specify a parametric model for $f(Z_{j+1} \mid R, \bar{Z}_j, \bar{A}_j)$ with parameter vector $\theta = (\theta_1', \theta_2')'$, such that $\theta_1 = 0$ and $\theta_2 \neq 0$ if and only if there is a maximum latent interval of 3 periods. Similarly there is no natural parametric model for which $\theta_1 = 0$ if and only if the G-null hypothesis is true. A third reason is related to robustness to model misspecification when the G-null mean hypothesis is true and the Y_m and Z_m are discrete, and is discussed in Sec. (13a) of Robins (1989).

When (26) is true, we can estimate ψ_0 by considering G-estimators $\hat{\psi}^{\Delta}(c^{\Delta})$ that solve $0 = n^{-1/2}S^{\Delta}(\psi, c^{\Delta}) = n^{-1/2}S(\psi, c) + n^{-1/2}S^{\dagger}(\psi, c^{\dagger})$ where $c^{\Delta} = (c, c^{\dagger})$, $S(\psi, c)$ and c are as defined previously and $n^{-1/2}S^{\dagger}(\psi, c^{\dagger}) = n^{-1/2}\Sigma_i U_i^{\dagger}(\psi, c^{\dagger}) = n^{-1/2}\sum_{i=1}^{n}\sum_{m=0}^{K-1} U_{m,i}(\psi, c_m)$ where $q^{\dagger} = (q_0, \ldots, q_{K-1})$, $g^{T} = (g_0, \ldots, g_{K-1})$, $c^{\dagger} = (q^{\dagger}, g^{\dagger}) = (c_0, \cdots, c_{K-1})$, $c_m \equiv c_m(a_m, x_m) \equiv (q_m, g_m) = \{q_m(x_m), g_m(a_m, x_m)\}, x_m = (r, \bar{z}_m, \bar{a}_{m-1})$, $q_m(\cdot)$ takes value in R^{K-m}, $g_m(\cdot,\cdot)$ is a $(K-m) \times V$ dimensional matrix-valued function, $U_{m,i}(\psi, c_m) = d_m'(A_{m,i}, X_{m,i})\{H_i^{(m)}(\psi) - q_m(X_{m,i})\}$ where $d'_m(A_{m,i}, X_{m,i}) = g_m'(A_{m,i}, X_{m,i}) - \{\int g'(a_m, X_{m,i})dF(a_m \mid X_{m,i})\}$ and $H_i^{(m)}(\psi) = (H_{m-1,i}(\psi), \cdots, H_{K,i}(\psi))'$. Since $f(a_m \mid X_{m,i})$ is unknown, we shall need to consider G-estimators $\hat{\psi}^{\Delta}(c^{\Delta}) = \hat{\psi}^{\Delta}(c^{\Delta} : \hat{\alpha}^{\Delta})$ where $\hat{\alpha}^{\Delta\prime} = (\hat{\alpha}', \hat{\alpha}^{\dagger\prime})$, $\hat{\alpha}$ is as defined previously, $\hat{\alpha}^{\dagger}$ solves $0 = \Sigma_i S_{\alpha i}^{\dagger}(\alpha^{\dagger}) = \partial \log \prod_i \prod_{m=0}^{m=K-1} f[A_{m,i} \mid X_{m,i} : \alpha^{\dagger}]/\partial\alpha^{\dagger}$ where

$$f[A_{m,i} \mid X_{m,i}] = f[A_{m,i} \mid X_{m,i} : \alpha_0^{\dagger}] \quad m = (0, \ldots, K\text{-}1) \tag{31}$$

for some $\alpha_0^{\dagger}$; and $\hat{\psi}^{\Delta}(c^{\Delta})$ solves $0 = S^{\Delta}(\psi, c^{\Delta}, \hat{\alpha}^{\Delta})$, and $S^{\Delta}(\psi, c^{\Delta}, \hat{\alpha}^{\Delta})$ is defined like $S^{\Delta}(\psi, c^{\Delta})$ except with $dF[r \mid X_i : \hat{\alpha}]$ and $dF[a_m \mid X_{m,i} : \hat{\alpha}^{\dagger}]$ replacing the true densities. By arguments analogous to those in Sec. 3c.1 and 3c.5, subject to regularity conditions, under (10), (30) and (31), $\hat{\psi}^{\Delta}(c^{\Delta})$ and $\hat{\psi}^{\Delta}(c^{\Delta}) \equiv \hat{\psi}^{\Delta}(c^{\Delta} : \hat{\alpha}^{\Delta})$ will be asymptotically linear with influence functions $-\tau(c^{\Delta})^{-1}U_i^{\Delta}(c^{\Delta})$ and $-\tau(c^{\Delta})^{-1}\text{Resid}(U_i^{\Delta}(c^{\Delta}), S_{\alpha i}^{\Delta})$ where $\tau(c^{\Delta}) = E[\partial U_i^{\Delta}(\psi_0, c^{\Delta})/\partial\psi']$, and $S_{\alpha i}^{\Delta} = (S_{\alpha,i}', S_{\alpha i}^{\dagger\prime})'$. Similarly, under (30) and (31), $\hat{\psi}^{\dagger}(c^{\dagger}) = \hat{\psi}^{\dagger}(c^{\dagger} : \hat{\alpha}^{\dagger})$ solving $0 = S^{\dagger}(\psi, c^{\dagger}, \hat{\alpha}^{\dagger})$ is asymptotically linear with influence function $-[\tau(c^{\dagger})]^{-1}\ \text{Resid}(U_i^{\dagger}(\psi_0, c^{\dagger}), S_{\alpha i}^{\dagger})$ where $-\tau(c^{\dagger}) = E\{\partial U^{\dagger}(\psi_0, c^{\dagger}) / \partial\psi'\}$, $S^{\dagger}(\psi, c^{\dagger}, \hat{\alpha}^{\dagger})$ is defined like $S^{\dagger}(\psi, c^{\dagger})$ except with $dF[a_m \mid X_m : \hat{\alpha}^{\dagger}]$ replacing the true density, and we have used the following definition.

An estimator $\hat{\psi}$ of ψ_0 is asymptotically linear with influence function B if

$$n^{1/2}(\hat{\psi} - \psi_0) = n^{-1/2}\Sigma_i B_i + o_p(1) \ ,$$

$E(B) = 0$, $E(B'B) < \infty$. If $\hat{\psi}$ is asymptotically linear, then, by the central limit theorem and Slutsky's Theorem, $n^{1/2}(\hat{\psi} - \psi_0)$ is asymptotically normal with mean 0 and variance $E[BB']$. Asymptotically linear estimators $\hat{\psi}^{(1)}$ and $\hat{\psi}^{(2)}$ with the same influence function are asymptotically equivalent in the sense that $n^{1/2}(\hat{\psi}^{(1)} - \hat{\psi}^{(2)}) = o_p(1)$ Conversely, two asymptotically linear estimators that are asymptotically equivalent must have the same influence function.

Note $\tau(c^{\Delta})$ and $\tau(c^{\dagger})$ do not depend on the functions $q(\cdot)$ or $q_m(\cdot)$. We call $\hat{\psi}^{\dagger}(c^{\dagger})$ an "observational estimate" since its consistency in no way depends on the fact that R_i was assigned by physical randomization. Further $-\tau(c^{\dagger}) = E[U^{\dagger}(\psi_0, c^{\dagger})S_{\psi}']$ by the generalized information equality (Newey, 1990), where S_{ψ} is the score for ψ at the truth in the model characterized by Eqs. (30) and (31).

5. THE ASSUMPTION OF NO UNMEASURED CONFOUNDERS FOR Z-HISTORY

5a. Some Implications

Henceforth, if Z_m and Y_m are disjoint, we redefine Z_m to be $(Z_m', Y_m)'$, so Z_m now includes Y_m.

Definition: There exists no unmeasured confounders for Z-history given r if for all $G \in \mathbf{G}(z)$ and all m

$$\bar{Z}_G \amalg A_m \mid \bar{Z}_m, R, \text{"}G\text{"}(\bar{Z}_{m-1}) \quad . \tag{32}$$

Eq. (32) is stronger than our previous assumptions in that Eq. (32) implies that Eqs. (26), (27) and the assumption of no current treatment interaction w.r.t. rz are all true but the converse is false. However, Robins (1993, Appendix 2) argues that it would rarely make substantive sense to believe that (26), (27), or even the assumption of no current treatment interaction w.r.t. rz were true but (32) false. Henceforth, we shall assume (32) is true. We then have

Theorem 7: If (4) and (32) are true, then

(a) $$f(Z_{m+1} \mid R, \bar{Z}_m, \bar{A}_m) = f(Z_{m+1} \mid \bar{Z}_m, \bar{A}_m) \tag{33}$$

(b) $$\gamma^{|rz|}(k, r, \bar{z}_m, \bar{a}_m) \equiv \gamma^{|rz|}(k, \bar{z}_m, \bar{a}_m) \text{ does not depend on } r \tag{34}$$

(c) $$\bar{Z}_G \amalg A_m \mid \bar{Z}_m, \text{"}G\text{"}(\bar{Z}_{m-1}) \tag{35}$$

(d) In the semiparametric model (a) characterized by (4), (32), (31) and a SNMM $\gamma^{|rz|}(k, r, \bar{z}_m, \bar{a}_m, \psi) \equiv \gamma^{|rz|}(k, \bar{z}_m, \bar{a}_m, \psi)$, semiparametric efficient estimation of ψ_0 and $E(\bar{Y}_G)$ does not depend on the data through R.

Proof: By a proof analogous to that of Theorem (4.1) in Robins (1986), (32) implies $f(Z_{m+1} \mid R, \bar{Z}_m, \bar{A}_m) = f(Z_{G,m+1} \mid \bar{Z}_{G,m}, R)$ for any G s.t. "G"$(\bar{Z}_m) = \bar{A}_m$. But, by (4), $f(\bar{Z}_{G,m+1} \mid \bar{Z}_{G,m}, R)$ does not depend on R. Eq. (34) follows from Eq. (28) and Eq. (33). Eq. (35) follows from Eq. (33), Eq. (4) and Eq. (32) by a proof similar to those in the Appendix of Robins (1989a). Part (d) follows by using (33) to write the likelihood for a single subject i as

$$f(R)\, f(Z_0) \prod_{m=0}^{K-1} f(Z_{m+1} \mid \bar{Z}_m, \bar{A}_m)\, f(A_m \mid \bar{A}_{m-1}, \bar{Z}_m, R) \tag{36}$$

and noting that Eqs. (28), (34), and (36) imply, for each parametric submodel in which each term in likelihood (36) has variation independent parameters, the maximum likelihood estimator of ψ_0 and $E(\bar{Y}_G)$, $G \in \mathbf{G}(z)$, does not depend on the data through R.

Part (d) implies that, in semiparametric model (a), efficient inference concerning ψ_0 and $E(\bar{Y}_G)$ is unchanged if data on R is not recorded for data analysis or if R was recorded as 1 for all subjects. That is, given (32) it does not matter that R was assigned at random, or even that it was recorded for data analysis. *Henceforth, for notational convenience, in order to allow reuse of our previous formula, we shall assume that R was recorded as 1 for each subject so that R = 1 with probability 1 (w.p.1) in the recorded data.* That is, we can regard the data as obtained from an observational study. Thus, all our subsequent results also apply to observational studies.

5b. Efficiency Considerations

We next characterize all regular, asymptotically linear (RAL) estimators of ψ_0 in semiparametric model (a) of Theorem (7d) (when R = 1 w.p.1). We first note that the restrictions on the joint distribution of the observables implied by model (a) are Eqs. (30) and (31). In Appendix 3, we prove

Theorem 8: (i) The influence function of any RAL estimator $\hat{\psi}$ of ψ_0 in the model (a) characterized by restriction Eqs. (30) and (31) equals $-\{\tau(c^\dagger)\}^{-1}\text{Resid}(U^\dagger(c^\dagger), S^\dagger_\alpha)$ for some $c^\dagger$ and, thus, $\hat{\psi}$ is asymptotically equivalent to $\hat{\psi}^\dagger(c^\dagger)$. (ii) Further there exists $c^\dagger_{eff}$ such that $U^\dagger(c^\dagger_{eff})$ is the efficient score and $\{\text{Var}[U^\dagger(c^\dagger_{eff})]\}^{-1}$ is the semiparametric variance bound (Newey, 1990) for model (a), where $U^\dagger(c^\dagger) \equiv U^\dagger(\psi_0, c^\dagger)$. In addition, $-\tau(c^\dagger_{eff}) = \text{Var}[U^\dagger(c^\dagger_{eff})] = \text{Var}[\text{Resid}(U^\dagger(c^\dagger_{eff}), S^\dagger_\alpha)]$ so that the asymptotic variance of $\hat{\psi}^\dagger(c^\dagger_{eff})$ attains the bound.

In the following Theorem proved in Appendix 3, we provide a formula for $c^\dagger_{eff}$. Define $\dot{H}^{(m)}(\psi) = H^{(m)}(\psi) - q_{eff,m}(X_m)$ where $q_{eff,m}(X_m) \equiv E[H^{(m)}(\psi_0) \mid X_m]$. Recursively define $H^{*(K-1)}(\psi) = \dot{H}^{(K-1)}(\psi)$, $H^{*(m)}(\psi) = \dot{H}^{(m)}(\psi) - \sum_{j=m+1}^{K-1} \gamma_{mj} H^{*(j)}(\psi)$ with $\gamma_{mj} = \{E[\dot{H}^{(m)} H^{*(j)\prime} \mid A_j, X_j] - E[\dot{H}^{(m)} H^{*(j)\prime} W^{*(j)} \mid X_j]\, E[W^{*(j)} \mid X_j]^{-1}\} W^{*(j)}$ with $W^{*(j)} = \{\text{Var}(H^{*(j)} \mid A_j, X_j)\}^{-1}$ and, for example, $\dot{H}^{(m)} = \dot{H}^{(m)}(\psi_0)$. Define $\rho^*_m = \{E[\partial H^{*(m)}(\psi_0)/\partial\psi' \mid A_m, X_m] - E[\{\partial H^{*(m)}(\psi_0)/\partial\psi'\} W^{*(m)} \mid X_m]\, E[W^{*(m)} \mid X_m]^{-1}\} W^{*(m)}$. Recursively define $\rho_1 = \rho^*_1$, $\rho_m = \rho^*_m - \sum_{j=0}^{m-1} \rho_j \gamma_{jm}$. Note $E[\rho_m \mid X_m] = 0$ since $E[\rho^*_m \mid X_m] = 0$, $E[\gamma_{jm} \mid X_m] = 0$ and ρ_j is fixed given X_m for $j < m$. Let $c^\dagger_{eff,m} = (q_{eff,m}, g_{eff,m})$ with $g_{eff,m} = \rho_m$.

Theorem 9: The semiparametric efficient score S_{eff} in model (a) is $S_{eff} = \sum_{m=0}^{K-1} \rho^*_m H^{*(m)} = U^\dagger(c^\dagger_{eff})$ and $\rho^*_m H^{*(m)}$ and $\rho^*_k H^{*(k)}$ are uncorrelated if $m \neq k$. Further, the asymptotic variance of $\hat{\psi}^\dagger(c^\dagger_{eff})$ attains the semiparametric variance bound.

Corollary A3.2: If $E(H_k H_j \mid A_m, X_m) = E[H_k H_j \mid X_m]$ for $k>m$, $j>m$, $m = 0,\ldots,K-1$, then (1) $\gamma_{mj} = 0$ for $j = (m+1,\ldots,k-1)$; (2) $H^{*(m)}(\psi) = \dot{H}^{(m)}(\psi)$; (3) $\rho_m = \rho^*_m = \{E[\partial \dot{H}^{(m)}(\psi_0)/\partial\psi' \mid A_m, X_m] - E[\partial \dot{H}^{(m)}(\psi_0)/\partial\psi' \mid X_m]\}\{\text{Var}[\dot{H}^{(m)} \mid X_m]\}^{-1} \dot{H}^{(m)}$.

Corollary A3.3: If $H^{(m)} \coprod A_m \mid X_m$, then the conclusions of Corollary (A3.2) are true, and our structural nested mean model $\gamma^{(m)}(k, \bar{z}_m, \bar{a}_m, \psi)$ is a structural nested distribution model as defined in Appendix 2 of Robins et. al. (1992).

Theorem 9 is related to efficiency results independently obtained by Chamberlain (1992, 1993) although the method of proof is quite different. Specifically, if we impose the assumption that

$$E[H_k \mid X_j] = t_{jk}(X_j ; \omega_0),\ 0 \leq m < k,\ k = 1, \ldots, K \tag{37}$$

where $t_{jk}(\cdot,\cdot)$ is a known function and ω_0 is an unknown finite dimensional parameter, then the model characterized by restrictions (30) and (37) (i) is a special case of the sequential conditional moment restriction model discussed in Sec. 2 of Chamberlain (1992) and (ii) has the same efficiency bound as the model characterized by restriction (30), (31), and (37). However, for arbitrary chosen functions $t_{jk}(\cdot,\cdot)$, the model characterized by (30) may be incompatible with the further restriction (37) in the sense there exists no distribution for which (30) and (37) are both true. In contrast, the model defined by (30) is always compatible with the additional restriction (31).

The model characterized by (30) differs from the model studied by Chamberlain (1993) in that there exist a number of unknown functions $E[H_k \mid X_j]$ and each depends only on the data X_j rather than on all the data (A_K, X_K).

Finally, one may wish to estimate the entire law of the vector-valued random variable $\bar{Y}_G$ rather than simply the mean. Robins et al. (1992, Appendix 2) develop methods for doing so based on multivariate structural nested distribution models.

6. MISSING DATA

Throughout we have assumed the data on the observables $(R_i, \bar{A}_i, \bar{Z}_i, \bar{Y}_i)$ were available for each study subject. In practice, one would expect some missing data and loss to follow-up. In the set up of Sec. 5, let $C^* = K+1$ if a subject is without missing data and let $C^* = m$, $m \leq K$, if t_m is the first time any component of (A_m, X_m) is missing. Define $R_k^* = 1$ for $k < C^*$, $R_k^* = 0$ for $k \geq C^*$. Assume the data is missing at random (Rubin, 1976) in the sense that

$$\pi_k \equiv p[R_k^* = 1 \mid R_{k-1}^* = 1, X_K, A_K] = p[R_k^* = 1 \mid R_{k-1}^* = 1, X_{k-1}, A_{k-1}] \tag{38}$$

Set $\bar{\pi}_{mk} = \prod_{j=m+1}^{k} \pi_j, k > m$. Assume $\pi_k = \pi_k(\omega_0)$ where $\pi_k(\omega) \equiv \pi_k(A_{k-1}, X_{k-1}, \omega)$ is a known function, and ω_0 is an unknown parameter. Let $\hat{\omega}$ solve $0 = \Sigma_i S_{\omega,i}(\omega)$, $S_\omega(\omega) = \frac{\partial}{\partial \omega}\{\Sigma_k R_k^* \pi_k(\omega) + (1 - R_k^*) R_{k-1}^* [1 - \pi_k(\omega)]\}$. Let $\Delta_m(\omega) = \text{Diag}\{R_k^*/\bar{\pi}_{mk}\}$, $k = m+1, \ldots, K$. Then, under model (a) of Theorem 7d and (38), by arguments similar to those in Robins and Rotnitzky (1992) and Robins, Rotnitzky, and Zhao (1993), $\hat{\psi}^\dagger(c^\dagger ; \hat{\alpha}^\dagger, \hat{\omega})$ solving $0 = \Sigma_i U_i(\psi, c^\dagger, \hat{\alpha}^\dagger, \hat{\omega}) = \Sigma_i \{\Sigma_m R_{mi}' d_m'(A_{m,i}, X_{m,i} ; \hat{\alpha}^\dagger) \Delta_{mi}(\hat{\omega}) [H_i^{(m)}(\psi) - q_m(X_{m,i})]\}$ is asymptotically linear with influence function $[-\tau(c^\dagger)]^{-1}\text{Resid}\{U(\psi_0,c^\dagger,\alpha_0^\dagger,\omega_0),(S_\alpha^{\dagger\prime},S_\omega^\prime)^\prime\}$. Further $E[S_\alpha S_\omega^{\dagger\prime}] = 0$. Missing data may be due to loss to follow-up (censoring) or just failure to record the relevant variables. By arguments similar to those in Robins and Rotnitzky (1993), there exists a $c^\dagger$ and a model $\pi_k(\omega)$ such that $\hat{\psi}^\dagger(c^\dagger:\hat{\alpha}^\dagger,\hat{\omega})$ is semiparametric efficient.

7. MULTIPLICATIVE STRUCTURAL NESTED MEAN MODELS

Redefine $\gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m) = \log\left\{E\left[Y_{k,G=(\bar a_m,0)}\middle|r,\bar z_m,\bar a_m\right]/E\left[Y_{k,G=(\bar a_{m-1},0)}\middle|r,\bar z_m,\bar a_m\right]\right\}$ and $h_{k,i} = y_{k,i}\exp\left\{\sum_{m=0}^{k-1}\gamma^{|\bar z|}(k,r_i,\bar z_{m,i},\bar a_{m,i})\right\}$ with $h_{k,i}(\psi)$ similarly redefined. $\gamma(k,r,\bar z_m,\bar a_m,\psi)$ is now referred to as a multiplicative SNMM. Then it can be shown that Theorem 8 remains true and, under its suppositions, $E\left[Y_{k,G=\bar a}\right] = E\left[Y_{k,G=(0)}\right]\exp\left[\sum_{m=0}^{k-1}\gamma^{|\bar z|}(k,r,\bar a_m)\right]$ whenever $\gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m)$ does not depend on $\bar z_m$. If $\gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m)$ does depend on $\bar z_m$, then $E\left[Y_{k,G=\bar a}\right]$ can be estimated using the obvious generalization of the simulation algorithm of Appendix 4. Note that no analog of Eq. (22) is available.

Appendix 1

For $0 \le m \le k\text{-}1$, define the random variable $H_{(k,m)} = Y_k - \sum_{j=m}^{j=k-1}\gamma^{|\bar z|}(k,R,\bar Z_j\bar A_j)$. Note $H_{(k,0)} = H_k$ where H_k is as defined previously.

Theorem A.1: $E[H_{(k,m)}\mid r,\bar z_m,\bar a_m] = E\left[Y_{k,G=(\bar a_{m-1},0)}\middle|r,\bar z_m,\bar a_m\right]$

Proof: Case 1: $m=k\text{-}1$ $E\left[Y_{k,G=(\bar a_{m-1},0)}\middle|r,\bar z_m,\bar a_m\right] = E\left[Y_{k,G=(\bar a_m,0)}\middle|r,\bar z_m,\bar a_m\right] - \gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m) = E[Y_k\mid r,\bar z_m,\bar a_m]-\gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m) = E[H_{(k,m)}\mid r,\bar z_m,\bar a_m]$, where the first equality is by the definition of $\gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m)$, the second follows from consistency assumption (a) of Section 2c, and the third by the definition of $H_{(k,m)}$.

Case 2: $m<k\text{-}1$: We proceed by induction and assume the Theorem is true with $m+1$ replacing m and show it is true for m, which together with Case 1 will prove the theorem. Again, by definition,

$$E\left[Y_{k,G=(\bar a_{m-1},0)}\middle|r,\bar z_m,\bar a_m\right] = E\left[Y_{k,G=(\bar a_m,0)}\middle|r,\bar z_m,\bar a_m\right] - \gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m) \tag{A.1}$$

But, $E\left[Y_{k,G=(\bar a_m,0)}\middle|r,\bar z_m,\bar a_m\right] = \int dF(a_{m+1},z_{m+1}\mid r,\bar z_m,\bar a_m)\cdot E\left[Y_{k,G=(\bar a_m,0)}\middle|a_{m+1},z_{m+1},r,\bar z_m,\bar a_m\right] = \int dF(a_{m+1}z_{m+1}\mid r,\bar z_m,\bar a_m)\cdot E[H_{(k,m+1)}\mid a_{m+1},z_{m+1},r,\bar z_m,\bar a_m]$, by the induction assumption. Thus, the right hand side of Eq. (A.1) equals $E[H_{(k,m+1)}\mid r,\bar z_m,\bar a_m] - \gamma^{|\bar z|}(k,r,\bar z_m,\bar a_m)$, which, by definition, equals $H_{(k,m)}$ when $(R,\bar Z_m,\bar A_m)=(r,\bar z_m,\bar a_m)$ proving the Theorem.

Corollary A.1: $E[H_k] = E[Y_{k,G=0}]$.

Proof: If $m=0$, $E[H_k \mid r, z_0, \bar{z}_0] = E[Y_{k,G=0} \mid r, z_0, \bar{z}_0]$ by Theorem A.1. Now take unconditional expectations.

Theorem A.2: If there is no-current-treatment-interaction with respect to rz for a particular $G^* \epsilon G(z)$, then

$$\phi^{|rz|}(k,r,z_m,\text{``}G^*\text{''}(z_{m-1}),G^*) =$$

$$\int\int\cdots\int \left\{ \sum_{j=m}^{j=k-1} \gamma^{|rz|}\left(k,r,\bar{z}_j,\text{``}G^*\text{''}(\bar{z}_j)\right) \right\} \cdot \left\{ \prod_{j=m}^{j=k-1} [dF[z_{j+1} \mid r,\bar{z}_j,\text{``}G^*\text{''}(\bar{z}_j)] \right\} \quad \text{(A.3)}$$

Proof: Case 1: $m=k-1$: The right-hand side of Eq. (A. 3) equals $\gamma^{|rz|}(k,r,z_m,\text{``}G^*\text{''}(z_m))$ by direct calculation. But, $\phi^{|rz|}(k,r,z_m,\text{``}G^*\text{''}(z_{m-1}),G^*) = E[Y_{k,G^*} \mid r,z_m,\text{``}G^*\text{''}(z_m)] -$
$E\left[Y_{k,G=(\text{``}G^*\text{''}(\bar{z}_{m-1}),0)} \middle| r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] = E\left[Y_{(k,G=\text{``}G^*\text{''}(\bar{z}_m),0)} \middle| r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] -$
$E\left[Y_{k,G=\text{``}G^*\text{''}(\bar{z}_{m-1}),0)} \middle| r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] = \gamma^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m))$ where the first equality is by the definition Eq. (8), the second follows from consistency assumption (b) of Section 2c, and the third by the definition of $\gamma^{|rz|}(k,r,z_m,\text{``}G^*\text{''}(z_m))$.

Case 2: $m<k-1$: We proceed by induction and show that if the theorem is true with $m+1$ replacing m then it is true for m. In conjunction with Case 1 this will prove the theorem.

$\phi^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_{m-1}),G^*) = E\left[Y_{k,G^*} - Y_{k,G=(\text{``}G^*\text{''}(\bar{z}_{m-1}),0)} \middle| r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] =$

$E\left[Y_{k,G^*} - Y_{k,G=(\text{``}G^*\text{''}(\bar{z}_m),0)} \middle| r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] + \gamma^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)) =$

$\int E\left[Y_{k,G^*} - Y_{k,G=(\text{``}G^*\text{''}(\bar{z}_m),0)} \middle| \bar{z}_{m+1},r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)\right] \cdot dF[z_{m+1} \mid r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)] =$

$\gamma^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)) = \int \phi^{|rz|}(k,r,\bar{z}_{m+1},\text{``}G^*\text{''}(\bar{z}_m),G^*) \cdot dF[z_{m+1} \mid r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m)] +$

$\gamma^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m))$, where the first equality is by the definition Eq. (8), the second by the definition of $\gamma^{|rz|}(k,r,\bar{z}_m,\text{``}G^*\text{''}(\bar{z}_m))$, the third from the laws of probability, and the fourth by our induction assumption. But, by a straightforward calculation, the right-hand side of the last equality equals the right-hand side of Eq. (A.3) proving the theorem.

Proof of Theorem 4: Theorem 4 in the text follows by setting $m = 0$ and then integrating out z_0.

Proof of Theorem 5: By the definition Eq. (8), it is sufficient to show

$$E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \middle| \bar{w}_m,\text{``}G^*\text{''}(\bar{w}_{m-1}),r,a_m\right] = E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \middle| \bar{w}_m,\text{``}G^*\text{''}(\bar{w}_{m-1}),r\right]$$

for $a_m = \text{``}G^*\text{''}(t_m,\bar{w}_m)$. But $E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \middle| \bar{w}_m,\text{``}G^*\text{''}(\bar{w}_{m-1}),r,a_m\right] =$

$\int E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \middle| \bar{w}_m,\bar{z}_m,\text{``}G^*\text{''}(\bar{w}_{m-1}),r,a_m\right] \cdot dF[\bar{z}_m \mid \bar{w}_m,\text{``}G^*\text{''}(\bar{w}_{m-1}),r,a_m]$

Now the first term in the integral does not depend on a_m by the assumption of no-current-treatment-interaction with respect to rz. If Eq. (18) holds, then the second term also does not depend on a_m which proves the theorem in this case.

Furthermore, the first term in the integral is given by Eq. (A.3) since z_m determines w_m. But, by direct calculation, when Eqs. (19) and (20) hold, Eq. (A.3) only depends on $\bar{w}_m$, and not further on $\bar{z}_m$. Thus $E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \mid \bar{w}_m, \bar{z}_m, \text{"}G^{**}\text{"}(\bar{w}_{m-1}), r, a_m\right] =$ $E\left[Y_{k,G^*} - Y_{k,G=(\bar{a}_{m-1},0)} \mid \bar{w}_m, \text{"}G^{**}\text{"}(\bar{w}_{m-1}), r\right]$ which proves the theorem in this case.

<u>Theorem A.3</u>: Eq. (27) implies Eq. (28):

Proof: Case 1: $m = k-1$. This case is an immediate consequence of consistency assumption a.

Case 2: $m<k-1$: We proceed by induction and assume the Theorem is true with $m+1$ replacing m and show it is true for m, which together with Case 1 will prove the theorem.

$$E[Y_{k,G} \mid r, z_m, \bar{a}_{m-1} = \text{"}G\text{"}(\bar{z}_{m-1})] =$$

$$\int E\left[Y_{k,G}^{*} \mid r, \bar{z}_{m+1}, \bar{a}_{m-1} = \text{"}G\text{"}(\bar{z}_{m-1}), a_m\right] dF\left[a_m, z_{m+1} \mid r, \bar{z}_m, \bar{a}_{m-1} = \text{"}G\text{"}(\bar{z}_{m-1})\right] \quad . \qquad \text{(A.4)}$$

But $\int E\left[Y_{k,G} \mid r, \bar{z}_{m+1}, \bar{a}_{m-1} = \text{"}G\text{"}(\bar{z}_{m-1}), a_m\right] = E\left[Y_{k,G} \mid r, \bar{z}_{m+1}, \bar{a}_m = \text{"}G\text{"}(\bar{z}_m)\right]$ by (27). So the R.H.S. of (A.4) equals $\int E\left[Y_{k,G} \mid r, \bar{z}_{m+1}, \bar{a}_m = \text{"}G\text{"}(\bar{z}_m)\right] dF\left[Z_{m+1} \mid r, \bar{z}_m, \bar{a}_{m-1} = \text{"}G\text{"}(\bar{z}_{m-1})\right]$. Now apply the induction assumption to $E\left[Y_{k,G} \mid r, \bar{z}_{m+1}, \bar{a}_m = \text{"}G\text{"}(\bar{z}_m)\right]$.

<u>Appendix 2</u>

We first show that parametric models for $\gamma^{(m)}(k, r, z_m, \bar{a}_m)$ and $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ are compatible under Eq. (4) or, subsequently, Eq. (9) when z_k and y_k are disjoint. Let w_k represent all components z_k other than y_k. Then we can write the observables $(R_i, \bar{Y}_i, \bar{Z}_i, \bar{A}_i)$ as $(R_i, \bar{Y}_i, \bar{W}_i, \bar{A}_i)$. Consider the one to one transformation mapping $(\bar{h}_i, \bar{w}_i, r_i, \bar{a}_i)$ into $(\bar{y}_i, \bar{w}_i, r_i, \bar{a}_i)$ where $\bar{y}_i$ is recursively defined in terms of $(\bar{h}_i, \bar{w}_i, r_i, \bar{a}_i)$ as follows:

$Y_{0,i} = h_{0,i}, y_{k,i} = h_{k,i} + \sum_{m=0}^{m=k-1} \gamma^{(m)}(k, r_i, \bar{z}_{m,i}, \bar{a}_{m,i})$, where the $\bar{z}_{m,i}$ are recursively known for $m < k$. It follows from this recursion formula that $\partial\bar{y}_i / \partial\bar{h}_i'$ is a lower triangular matrix with ones along the diagonal. Thus the Jacobian determinant of the transformation is one. It follows that any function $\gamma^{(m)}(k, r, z_m, \bar{a}_m)$ and any joint distribution for $(\bar{H}_i, \bar{W}_i, R_i, \bar{A}_i)$ specifies a unique joint distribution for $(R_i, \bar{Y}_i, \bar{Z}_i, \bar{A}_i)$. Further any density $f(\bar{w}_i, \bar{a}_i \mid r_i)$ is compatible with the restriction Eq. (12). [Of course, Eq. (9) plus a particular density $f(\bar{w}_i, \bar{a}_i \mid r_i)$ implies a restriction on the density $f(\bar{h}_i \mid \bar{w}_i, \bar{a}_i, r)$.] It follows that any function $\gamma^{(m)}(k, r, z_m, \bar{a}_m)$ and any density $f[z_{j+1} \mid r, z_j, \bar{a}_j]$ are compatible under restriction (9) if $z_k = w_k$ for all k. Indeed, since the likelihood can be factored as $f(\bar{h}_i \mid w_i, \bar{a}_i, r_i)\, f(w_i, \bar{a}_i \mid r_i)\, p(r_i)$ and the parameter ψ_0 of a SNMM enters the likelihood only through the first term, the data $(w_i, \bar{a}_i, r_i)$ are S-ancillary for ψ_0. Thus, when $z_k = w_k$, specification of parametric models for $f[z_{j+1} \mid r, \bar{z}_j, \bar{a}_j]$ do not aid in the estimation of ψ_0.

In order to show that parametric models for $\gamma^{|\bar{a}|}(k, r, \bar{z}_m, \bar{a}_m)$ and $f[z_{j+1} \mid r, \bar{z}_j, \bar{a}_j]$ may be incompatible under Eq. (9) when y_k and z_k are not disjoint, it is sufficient to show that there exists some function $\gamma^{|\bar{a}|}(k, r, \bar{z}_m, \bar{a}_m)$ and some density $f[z_{j+1} \mid r, \bar{z}_j, \bar{a}_j]$ for which there is no joint density for the observables $(R_i, \bar{Y}_i, \bar{Z}_i, \bar{A}_i)$ satisfying Eq. (9) that is consistent with this function $\gamma^{|\bar{a}|}(k, r, \bar{z}_m, \bar{a}_m)$ and density $f[z_{j+1} \mid r, \bar{z}_j, \bar{a}_j]$. Choose the function $\gamma^{|\bar{a}|}(k, r, \bar{z}_m, \bar{a}_m)$ to be identically 0 and assume y_1 is contained in z_1. Then, by Eq. (9), $E[Y_1 \mid z_0, R=r] = E[Y_1 \mid z_0, R=r^*]$. Thus, it cannot be the case that $E[Y_1 \mid z_0, R=r, a_0] > E[Y_1 \mid z_0, R=r^*, a_0]$ for all a_0 . But, it is clear there exists densities $f[y_1 \mid z_0, r, a_0]$ for which this inequality is true.

Appendix 3

Proof of Theorems 8 and 9

We first provide a characterization of the nuisance tangent space for our model (Newey, 1990). To do so, let Eq. (30m) be defined like Eq. (30) but only for $k = m$ rather than for $k = (1,\ldots,K)$. Let model a_m be the semiparametric model characterized by Eq. (30m) and (31). Let models b and b_m be characterized by Eq. (30) and Eq. (30m) respectively. Since model (a) is the model defined by models a_m being true, $m = 1,\ldots,K$, the nuisance tangent space Λ^a for model a is the intersection of the nuisance tangent spaces Λ_m^a for the models a_m i.e.,

$$\Lambda^a = \Lambda_1^a \cap \Lambda_2^a \ldots \cap \Lambda_K^a \tag{A3.1}$$

Likewise

$$\Lambda^b = \Lambda_1^b \cap \Lambda_2^b \ldots \cap \Lambda_K^b \tag{A3.2}$$

where Λ^b and Λ_m^b are the nuisance tangent spaces of the models b and b_m respectively. In all four models, the parametric component is the parameter ψ_0. We then have by standard Hilbert space results

Lemma A3.1: $\Lambda^{a,\perp} = \ell s\{\Lambda_m^{a,\perp} : m = 1,\ldots,K\}$ and $\Lambda^{b,\perp} = \ell s\{\Lambda_m^{b,\perp} : m = 1,\ldots,K\}$ where $\perp$ stands for the orthogonal complement of the subspace and $\ell s\{\cdot\}$ refers to the closed linear span of the subspaces $\cdot$. The next theorem characterizes $\Lambda_k^{a,\perp}$ and $\Lambda_k^{b,\perp}$. It is proved in Appendix 4.

Theorem A3.1: (i) $\Lambda_k^{b,\perp} = \{U^k(\psi_0, q_{eff}^k, g^k)\}$ and (ii) $\Lambda_k^{a,\perp} = \{\mathrm{Resid}[U^k(\psi_0, q^k, g^k) + M^{(k)}(q^{(k)}, g^{(k)}), S_a^{\dagger}]\}$ where $g^k = (g_0,\ldots,g_{k-1})$, $g^{(k)} = (q_k,\ldots,g_K)$, $q^{(k)} = (q_k,\ldots,q_K)$, $q^k = (q_0,\ldots,q_{k-1})$, $g_m(A_m, X_m)$ is an arbitrary v - vector valued function of (A_m, X_m), $q_m(X_m)$ is an arbitrary real valued function of $X_m, q_{eff,m}^k = E[H_k \mid X_m]$,

$U^k(\psi_0,q^k,g^k) = \sum_{m=0}^{k-1} U_m^k(\psi_0,q_m^k,g_m^k)$, $U_m^k(\psi_0,q_m,g_m) = d_m(A_m,X_m)'\{H_k(\psi_0) - q_m(X_m)\}$ with $d_m(A_m,X_m) = g_m(A_m,X_m) - E[g_m(A_m,X_m) \mid X_m]$, and $M^{(k)}(q^{(k)},g^{(k)}) \equiv \sum_{m=k}^{K} d_m(A_m,X_m)'q_m(X_m)$.

The following corollary is an easy consequence of Lemma (A3.1) and Theorem (A3.1).

<u>Corollary (A3.1)</u>: $\Lambda^{b,\perp} = \{U^\dagger(\psi_0,q_{eff}^\dagger,g^\dagger)\}$ and $\Lambda^{a,\perp} = \{\text{Resid}[U^\dagger(\psi_0,q^\dagger,g^\dagger),S_\alpha^\dagger]\}$.

<u>Proof of Theorem (8i)</u>:

According to Theorem (2.2) in Newey (1990), the influence function of any RAL estimator of ψ in a semiparametric model with nuisance tangent space Λ and finite dimensional parameter ψ is of the form $E[AS_\psi']A$ for some $A \in \Lambda^\perp$ where S_ψ is the score with respect to ψ evaluated at the truth. Theorem (8i) then follows from the characterization of $\Lambda^{a,\perp}$ provided in Corollary (A3.1) and the generalized information equality (Newey, 1990) identity $-E[\partial U^\dagger(\psi_0,q^\dagger,g^\dagger) / \partial\psi'] = E[U^\dagger(\psi_0,q^\dagger,g^\dagger)S_\psi']$.

<u>Proof of Theorem (8ii) and Theorem 9</u>:

In Lemma A4.1 below, we prove that the efficient score S_{eff} in models (a) and (b) is the same. Now by definition, in model (b), $S_{eff} = \Pi[S_\psi \mid \Lambda^{b,\perp}]$ where Π is the Hilbert space projection operator. Thus, by Lemma (A3.1), $S_{eff} = \Pi[S_\psi \mid \ell s\{\Lambda_k^{b,\perp}; k = 1,\ldots,K\}]$. A Gram-Schmidt orthogonalization can be used to show that $\ell s\{\Lambda_k^{b,\perp}; k = 1,\ldots,K\} = \ell s\{\Lambda_k^*; k = 1,\ldots,K\}$, where $\Lambda_{k+1}^* = \{d(V_k)H^{*(k)}; E[d(V_k) \mid X_k] = 0\}$ with $V_k = (\bar{A}_k,\bar{Z}_k)$. Since $\Lambda_k^* \perp \Lambda_m^*$ for $m \neq k$, it follows that $\Lambda^{b,\perp} = \Lambda_1^* \oplus \Lambda_2^* \oplus \cdots \oplus \Lambda_K^*$ where $\oplus$ denotes a direct sum. Thus, $S_{eff} = \sum_{k=0}^{K-1} \Pi[S_\psi \mid \Lambda_{k+1}^*]$. Now it is straightforward to show that for any $B = b(V_K)$, $\Pi[B \mid \Lambda_{k+1}^*] = E[BH^{*(k)\prime} \mid V_k] - E[BH^{*(k)\prime}W^{*(k)} \mid X_k]\, E[W^{*(k)} \mid X_k]^{-1}W^{*(k)}H^{*(k)}$. Hence, by the generalized information equality $-E[\partial H(\psi_0)^{*(k)}/\partial\psi' \mid V_k] = E[H^{*(k)}S_\psi' \mid V_k]$, we obtain $\Pi[S_\psi \mid \Lambda_{k+1}^*] = \rho_k^* H^{*(k)}$. This proves Theorem (9) and (8ii).

<u>Appendix 4</u>: <u>Proof of Theorem (A3.1)</u>

Before proving Theorem (A3.1), we recast models a, a_k, b, and b_k in a way that makes it easy to simulate data from the model, to write down the likelihood function, and to characterize the nuisance tangent space. Write $Y_k = H_k + \sum_{m=0}^{k-1} \gamma^{(m)}(k,R,\bar{Z}_m,\bar{A}_m;\psi_0), k = 1,\ldots,K$ and consider the identity $H_k = H_k - E(H_k \mid V_{k-1}) + \sum_{m=0}^{k-1} E(H_k \mid V_m) - E(H_k \mid V_{m-1}) + E(H_k) \equiv \sigma_k + \sum_{m=0}^{k-1} \dot{q}_{mk}(V_m) + \beta_{0k}$ where, for $k = 1,\ldots,K$, $V_k \equiv (\bar{A}_k,\bar{Z}_k) \equiv (A_k,X_k) \equiv (\bar{A}_k,\bar{Y}_k,\bar{W}_k)$, $\sigma_k \equiv H_k - E(H_k \mid V_{k-1})$, $\beta_{0k} = E(H_k)$, for $0 \le m < k$, $\dot{q}_k(V_m) \equiv Q_{mk} \equiv E(H_k \mid V_m) - E(H_k \mid V_{m-1})$ and, by convention, $\sum_{m=0}^{k-1} S_m \equiv 0$ for any S_m when $k = 0$. Define $\sigma_0 = Y_0 - \beta_{00}$ with $\beta_{00} \equiv E(Y_0)$. We shall consider the 1-1 where $\eta_k = \{\psi, \Theta_{6m}, \Theta_{5m}, \Theta_{4m}, \Theta_{2m}\,; m = 0,1,\ldots,k-1\}$ is restricted only by

$\int t\, dF(t \mid V_{k-1}\,; \Theta_{2k}) = 0.$

Theorem A4.1: For any $B = b(V_K)$, $\Pi[B \mid \Lambda_k^{b,\perp}] = \Pi[B_k \mid \Lambda_k^{b,\perp}] + \sum_{m=0}^{k-1} \Pi[B_m \mid \Lambda_k^{b,\perp}]$ where $B_k \equiv E[B \mid \sigma_k, V_{k-1}] - E[B \mid V_{k-1}]$, $B_m = E(B \mid X_m) - E(B \mid V_{m-1})$, $\Pi[B_k \mid \Lambda_k^{b,\perp}] = \sum_{m=1}^{k} \{R_{(m-1)k} - T_{(m-1)k}^{-1} E[R_{(m-1)k} W_{(m-1)k} \mid X_{m-1}]\} W_{(m-1)k}\, \epsilon_{mk}$, and, for $m < k$, $\Pi[B_m \mid \Lambda_k^{b,\perp}] = \sum_{j=1}^{m} \{R^*_{(j-1)mk} - T_{(j-1)k} E[R^*_{(j-1)mk} W_{(j-1)k} \mid X_{j-1}]\} W_{(j-1)k}\, \epsilon_{jk}$. Here, we recursively define, for $m = k-1,\ldots,0$, $W_{mk}^{-1} = \mathrm{Var}[\epsilon_{(m+1)k} \mid V_m]$, $T_{mk} = E[W_{mk} \mid X_m]$ with $\epsilon_{kk} = \sigma_k$, $\epsilon_{mk} \equiv W_{mk}\, \epsilon_{(m+1)k} T_{mk}^{-1} + Q_{mk} = \dot{H}_{mk} + \sum_{j=m+1}^{k-1} F_{jk}\, \epsilon_{(j+1)k}$, $F_{jk} = -1 + W_{jk} T_{jk}^{-1}$, $\dot{H}_{kk} = \sigma_k$, $\dot{H}_{mk} \equiv H_k - E[H_k \mid X_{m-1}] = \sigma_k + \sum_{j=m}^{k-1} Q_{jk}$; $R_{(k-1)k} \equiv E[B\sigma_k \mid V_{k-1}]$, $R_{(m-1)k} = E[R_{mk} W_{mk} T_{mk}^{-1} \mid V_{m-1}]$ and, for $j = 0,\ldots,m-1$, $R^*_{(m-1)mk} = E[BQ_{mk} \mid V_{m-1}]$, $R^*_{(j-1)mk} = E[R^*_{jmk} W_{jk} T_{jk}^{-1} \mid V_{j-1}]$.

Proof Deferred

As a corollary, we obtain the following characterization of the efficient score in model b_k.

Corollary A4.1:

In model b_k, the efficient score for ψ, $S^k_{eff} \equiv \Pi[S^k_\psi \mid \Lambda_k^{b,\perp}] = \Pi[B_k \mid \Lambda_k^{b,\perp}]$ with $B_k = E[S^k_\psi \mid \sigma_k, V_{K-1}] - E[S^k_\psi \mid V_{k-1}]$ where $S^k_\psi = \partial \log \mathcal{L}_k(\psi,\Theta)/\partial\psi$ evaluated at the truth and $R_{(k-1)k} = -E[\partial H_k(\psi_0)/\partial\psi' \mid V_{k-1}]$.

Proof:

$B_m = 0$ since $E[S^k_\psi \mid V_{k-1}] = 0$ by the conditional mean property of scores. Further, $R_{(k-1)k} = -E[\partial H_k(\psi_0)/\partial\psi' \mid V_{k-1}]$ by the generalized information equality.

Proof of Theorem (A3.1i):

$\Lambda_k^{b,\perp} = \left\{\dot{D} \equiv \sum_{m=1}^{k} d(V_{m-1})\, \epsilon_{mk}\,; E[d(V_{m-1}) \mid X_{m-1}] = 0\right\}$ since, by Theorem A4.1, $\Pi[B \mid \Lambda_k^{b,\perp}]$ is of the form $\dot{D}$ and, using the characterization of Λ_k^b given in the proof of Theorem A4.1 below, one can show by direct calculation that for any $\dot{D}$, $\dot{D} \perp \Lambda_k^b$. Thus,

$\Lambda_k^{b,\perp} \subseteq \{U^k(\psi_0, q^k_{eff}, g^k)\} \equiv \left\{\sum_{m=1}^{k} d(V_{m-1})\, \dot{H}_{mk}\,; E[d(V_{m-1}) \mid X_{m-1}] = 0\right\}$ since $d(V_{m-1})\, \epsilon_{mk} = \sum_{j=1}^{k-m+1} d^*_j(V_{k-j})\, \dot{H}_{(k-j+1)k}$ where $E[d^*_j(V_{k-j}) \mid X_{k-j}] = 0$ with $d^*_j(V_{k-j}) = d(V_{m-1}) \{F_{(k-j)k}\}^{I(k \geq m-j)} \prod_{t=1}^{t=k-m-1} [1 + F_{(k-j-t)k}]^{I(k > m-t)}$ and $\prod_{t=a}^{b} \equiv 1$ if $b < a$. A similar argument shows that $\{U^k(\psi_0, q^k_{eff}, g^k)\} \subseteq \Lambda_k^{b,\perp}$ proving Theorem (A3.1i).

Proof of Theorem A4.1: By formally differentiating the likelihood $\mathcal{L}_k(\psi,\Theta)$ with respect to Θ, it follows that the nuisance tangent space corresponding to Θ is transformation of the observables $(\bar{Y}_K, \bar{W}_K, \bar{A}_K)$ to $(\bar{\sigma}_K, \bar{W}_K, \bar{A}_K)$ with Jacobian determinant of one (see Appendix 2). Note for any ψ_0, $E[\sigma_k \mid V_{k-1}] = 0$ and $E[Q_{mk} \mid V_{m-1}] = 0$ without any restrictions on the joint distribution of the data. Restriction (30m) implies $Q_{jm} = \dot{q}_m(Z_j, V_{j-1}) \equiv \dot{q}_m(X_j)$ is not a function of A_j. Conversely $Q_{jm} = \dot{q}_m(X_j)$ implies (30m) holds. Hence model (b) is characterized by

$$E(\sigma_k \mid V_{k-1}) = 0,\ E(Q_{jk} \mid V_{j-1}) = 0,\ Q_{jk} = \dot{q}_k(X_j),\ 0 \le j < k,\ k = 1,\dots,K \qquad (A4.1)$$

Model (a) is characterized by Eq. (A4.1) and Eq. (31). Model b_m is characterized by (A4.1m) where (A4.1m) is (A4.1) but for $k = m$ only. Model a_m is characterized by Eq. (A4.1m) and Eq. (31).

Simulation Algorithm: To produce a random draw, V_K from a joint distribution satisfying the restrictions of model b, choose a SNMM $\gamma^{|z|}(k, \bar{z}_m, \bar{a}_m, \psi)$ and, for $m = 0,\dots,K$, densities $f(\sigma_m \mid V_{m-1})$ satisfying $E(\sigma_m \mid V_{m-1}) = 0$, $f(A_m \mid X_m)$, $f(W_m \mid \sigma_m, V_{m-1})$, constants β_m and functions $q^{\dagger}_{km}(\sigma_k, W_k, V_{k-1})$, $k = 0,\dots,m-1$. Then, beginning with $m = 0$,

Step (1): Draw σ_m from $f(\sigma_m \mid V_{m-1})$.

Step (2): Compute $Y_m = \sigma_m + \beta_m + \sum_{k=0}^{m-1} q^{\dagger}_{km}(\epsilon_k, W_k, V_{k-1}) + \gamma^{|z|}(m, \bar{Z}_k, \bar{A}_k, \psi)$ with

$q^{\dagger}_{km}(\sigma_k, W_k, V_{k-1}) = q^{\dagger}_{km}(\sigma_k, W_k, V_{k-1}) - \int\int q^{\dagger}_{km}(\sigma_k, w_k, V_{k-1})dF(w_k \mid \sigma_k, V_{k-1})dF(\sigma_k \mid V_{k-1})$.

Step (3): Draw W_m from $f(W_m \mid \sigma_m, V_{m-1})$.

Step (4): Draw A_m from $f(A_m \mid Y_m, W_m, V_{m-1})$.

Step (5): If $m < K$, increment m by 1 and return to Step 1; else stop.

Likelihood Functions: Later we shall need the likelihood function $\mathcal{L}_k(\psi, \Theta)$ for model b_k where $\mathcal{L}_k(\psi, \Theta) =$

$f(V_K \mid Y_k, V_{k-1}; \Theta_1) f[\sigma_k(\psi, \Theta_4, \Theta_5, \Theta_6) \mid V_{k-1}; \Theta_2] \prod_{m=0}^{k-1} \{f(A_m \mid X_m; \Theta_3) f(Z_m \mid V_{m-1}; \Theta_4)\}$, $\Theta = (\Theta_1, \dots, \Theta_6)$, and $\sigma_k(\psi, \Theta_4, \Theta_5, \Theta_6) =$

$Y_k - \Theta_6 - \sum_{m=0}^{k-1} \gamma^{|z|}(k, \bar{Z}_m, \bar{A}_m, \psi) - \sum_{m=0}^{k-1} q^{*}_{mk}(X_m; \Theta_5) - \int q^{*}_{mk}(z_m, V_{m-1}; \Theta_5) dF(z_m \mid V_{m-1}; \Theta_4)$ where $\Theta_1, \Theta_2, \Theta_3, \Theta_4$ and Θ_5 are infinite dimensional nuisance parameters. Θ_6 is a one dimensional nuisance parameter. $q^{*}_{mk}(\cdot; \Theta_5)$ is an arbitrary function indexed by the parameter Θ_5, and the only restriction is $\int t dF(t \mid V_{k-1}; \Theta_2) = 0$ so that $E[\sigma_k \mid V_{k-1}] = 0$.

The likelihood function for model b is

$$\mathcal{L}(\psi, \Theta) = \prod_{k=0}^{K} f[\sigma_k(\eta_k) \mid V_{k-1}; \theta_{2k}]\, f[W_k \mid Y_k, V_{k-1}; \Theta_k] f[A_k \mid Y_k, W_k, V_{k-1}; \Theta_{3k}]$$

where

$$\sigma_k(\eta_k) = Y_k - \Theta_{6k} - \gamma^{|z|}(k, \bar{Z}_m, \bar{A}_m, \psi) - \sum_{m=0}^{k-1} q^{*}_{mk}(Y_m, W_m, V_{m-1}; \Theta_{5m}) -$$
$$\int q^{*}_{mk}(y_m, w_m, V_{m-1}; \Theta_{5m}) f[w_m \mid y_m, V_{m-1}; \Theta_{4m}] f[\sigma_m(\eta_m) \mid V_{m-1}; \Theta_{2m}] dy_m dw_m$$

$\Lambda_k^b = \Lambda_k^1 + \Lambda_k^2 + \Lambda_k^3 + \Lambda_k^4 + \Lambda_k^5 + \Lambda_k^6$ where $\Lambda_k^1 = \{A_k^1 = a_k^1(V_K) ; E[A_k^1 \mid Y_k, V_{k-1}] = 0\}$; $\Lambda_k^2 = \{A_k^2 = a_k^2(\sigma_k, V_{k-1}) ; E[A_k^2 \mid V_{k-1}] = 0$ and $E[A_k^2\sigma_k \mid V_{k-1}] = 0\}$ since $E[\sigma_k \mid V_{k-1}] = 0$. $\Lambda_k^3 = \sum_{m=0}^{k-1} \Lambda_m^3$, $\Lambda_k^4 = \sum_{m=0}^{k-1} \Lambda_{mk}^4$, and $\Lambda_k^5 = \sum_{m=0}^{k-1} \Lambda_{mk}^5$. Λ_k^j is the contribution to the nuisance tangent space corresponding to the nuisance parameter Θ_j. Here $\Lambda_m^3 = \{A_m^3 = a_m^3(V_m) ; E[A_m^3 \mid X_m] = 0\}$, $\Lambda_m^* = \{A_m^* = a_m^*(Z_m, V_{m-1}) ; E[A_m^* \mid V_{m-1}] = 0\}$. $\Lambda_{mk}^4 = \{A_{mk}^4 \equiv A_{mk}^4(A_m^*) \equiv S_k A_m^* ; A_m^* \in \Lambda_m^*\}$. $\Lambda_{mk}^5 = \{A_{mk}^5 \equiv A_{mk}^5(A_m^*) \equiv A_m^* + S_k E[Q_{mk} A_m^* \mid V_{m-1}] : A_m^* \in \Lambda_m^*\}$, $\Lambda_k^6 = \{A_k^6 \equiv aS_k ; a \in R^v\}$, and $S_k = \partial \log f[\sigma_k \mid V_{k-1} ; \Theta_2]/\partial\sigma_k$ evaluated at the truth.

Λ_k^1 and Λ_m^3 are each orthogonal to all other subspaces in Λ_k^b. For each $B = b(V_K)$, we need to compute $\Pi[B \mid \Lambda_k^b]$ so that we may characterize $\Lambda_k^{b,\perp}$. However, the Λ_k^2, Λ_k^6, Λ_{mk}^4, and Λ_{mk}^5 are not orthogonal. But let

$$T_{mk}^* = E\left[T_{mk}^{-1} \middle| V_{m-1}\right]$$

and define $\tilde{\Lambda}_{mk}^4 = \{\tilde{A}_{mk}^4 \equiv \tilde{A}_{mk}^4(A_m^*) \equiv A_m^* W_{mk} \epsilon_{(m+1)k} ; A_m^* \in \Lambda_m^*\}$, $\tilde{\Lambda}_{mk}^5 = \{\tilde{A}_{mk}^5 \equiv \tilde{A}_{mk}^5(A_m^*) \equiv A_m^* - E[Q_{mk} A_m^* \mid V_{m-1}](T_{mk}^*)^{-1} T_{mk}^{-1} W_{mk} \epsilon_{(m+1)k} ; A_m^* \in \Lambda_m^*\}$, $\tilde{\Lambda}_k^6 = \{a\epsilon_{0k} ; a \in R^v\}$. We then have $\Lambda_k^b = \Lambda_k^1 \oplus \Lambda_k^2 \oplus \Lambda_k^3 \oplus \tilde{\Lambda}_k^4 \oplus \tilde{\Lambda}_k^5 \oplus \tilde{\Lambda}_k^6$ where $\tilde{\Lambda}_k^j = \tilde{\Lambda}_{0k}^j \oplus \tilde{\Lambda}_{1k}^j, \ldots, \oplus \tilde{\Lambda}_{(k-1)k}^j$, $j = 4.5$, since

(i) $\tilde{\Lambda}_{(k-1)k}^4 = \{\Pi[A_{(k-1)k}^4 \mid \Lambda_k^{2,\perp}] ; A_{(k-1)k}^4 \in \Lambda_{(k-1)k}^4\}$ by $\Pi[A_{mk}^4(A_m^*) \mid \Lambda_k^{2,\perp}] = A_m^* W_{(k-1)k} \epsilon_{kk}$ for $m < k$;

(ii) $\tilde{\Lambda}_{(k-1)k}^5 = \{\Pi[A_{(k-1)k}^5 \mid (\Lambda_k^2 \oplus \tilde{\Lambda}_{(k-1)k}^4)^\perp] ; A_{(k-1)k}^5 \in \Lambda_{(k-1)k}^5\}$ by (a) $\Pi[A_{mk}^5(A_m^*) \mid \Lambda_k^{2,\perp}] = \dot{A}_{mk}^5(A_m^*) \equiv A_m^* - E[Q_{mk} A_m^* \mid V_{m-1}] W_{(k-1)k} \epsilon_{kk}$ and (b) $\Pi[A_m^* - E[Q_{mk} A_m^* \mid V_{m-1}] W_{mk} \epsilon_{(m+1)k} \mid \tilde{\Lambda}_{mk}^{4,\perp}] = \tilde{A}_{mk}^5(A_m^*)$ with $m < k$ by Theorem (A4.2c);

(iii) $\tilde{\Lambda}_{(k-2)k}^4 = \{\Pi[A_{(k-2)k}^4 \mid (\Lambda_k^2 \oplus \tilde{\Lambda}_{(k-1)k}^4 \oplus \tilde{\Lambda}_{(k-1)k}^5)^\perp] ; A_{(k-2)k}^4 \in \Lambda_{(k-2)k}^4\}$ by (i) above and. for $m < k-1$, $\Pi[A_m^* W_{(k-1)k} \epsilon_{kk} \mid (\tilde{\Lambda}_{(k-1)k}^4 \oplus \tilde{\Lambda}_{(k-1)k}^5)^\perp] = A_m^* W_{(k-2)k} \epsilon_{(k-1)k}$ by Theorem (A4.2b).

(iv) $\tilde{\Lambda}_{(k-2)k}^5 = \{\Pi[A_{(k-2)k}^5 \mid (\Lambda_k^2 \oplus \tilde{\Lambda}_{(k-1)k}^4 \oplus \tilde{\Lambda}_{(k-1)k}^5 \oplus \tilde{\Lambda}_{(k-2)k}^4)^\perp] ; A_{(k-2)k}^5 \in \Lambda_{(k-2)k}^5\}$ by applying ii(a) above, then Theorem (A4.2b) and finally (iib) above: etc. and

(v) $\tilde{\Lambda}_k^6 = \{\Pi[A_k^6 \mid (\Lambda_k^2 \oplus \tilde{\Lambda}_k^4 \oplus \tilde{\Lambda}_k^5)^\perp] ; A_k^6 \in \Lambda_k^6\}$ by $\Pi[A_k^6(a) \mid \Lambda_k^{2,\perp}] = aW_{(k-1)k} \epsilon_{kk}$, and then applying Theorem (A4.2b) recursively.

To compute $\Pi[B \mid \Lambda_k^{b,\perp}]$, write $B = b(V_k)$ as $B_k^1 + B_k + \sum_{m=0}^{k-1} B_m + \sum_{m=0}^{k-1} B_m^3$ where B_k and B_m are as defined before and $B_k^1 = B - E[B \mid \sigma_k, V_{k-1}]$, $B_m^3 = E[B \mid V_m] - E[B \mid X_m]$. Since $B_m^3 \in \Lambda_m^3$ and $B_k^1 \in \Lambda_k^1$, it follows that $\Pi[B \mid \Lambda_k^{b,\perp}] = \Pi[B_k^1 \mid \Lambda_k^{b,\perp}] + \sum_{m=0}^{k-1} \Pi[B_m \mid \Lambda_k^{b,\perp}]$ which can be shown to be as given in Theorem A4.1 by repeated application of Theorem (A4.2a) and (A4.2b) below.

<u>Proof of Theorem (A3.1ii)</u>: It is easy to show that the nuisance tangent space in model a_k is $\tilde{\Lambda}_k^1 + \Lambda_k^2 + \tilde{\Lambda}^3 + \Lambda_k^4 + \Lambda_k^5 + \Lambda_k^6$ where $\tilde{\Lambda}^3 = \{\tilde{A}^3 \equiv aS_a^\dagger\,;\, a$ any constant matrix such that $aS_a^\dagger \in R^v\} \subseteq \Lambda_k^3$, and $\tilde{\Lambda}_k^1 = \{A_k^1 + A_k^3 - \tilde{A}^3\,;\, A_k^1 \in \Lambda_k^1,\ A_k^3 \in \Lambda_k^3,\ \tilde{A}^3 \in \tilde{\Lambda}^3\}$. Therefore, $\{\Pi(B \mid \Lambda_k^{a,\perp})\} = \Pi(B \mid \Lambda_k^{b,\perp}) + \Pi(B \mid \tilde{\Lambda}^{3,\perp})$. But $\Pi(B \mid \tilde{\Lambda}^{3,\perp}) = \Pi\left[\sum_{m=0}^{k-1} B_m^3 \mid \tilde{\Lambda}^{3,\perp}\right] = \text{Resid}\left(\sum_{m=0}^{k-1} B_m^3, S_a^\dagger\right)$ where B_m^3 is defined above. But the set $\left\{\sum_{m=0}^{k-1} B_m^3\right\}$ equals $\left\{\sum_{m=0}^{k-1} \{g(A_m,X_m) - E[g(A_m,X_m) \mid X_m]\}q(X_m)\right\}$.

<u>Theorem A4.2</u>: Let $G_m = g(V_m)$, $D_j = d(X_j)$ with $E[D_j \mid V_{j-1}] = 0$, and $D_m^* = d^*(X_m)$. Then

(a) $\Pi[D_j \mid \tilde{\Lambda}_{mk}^{5,\perp}] = D_j$ if $j<m$ and $\Pi[D_j \mid \tilde{\Lambda}_{mk}^{5,\perp}] = E[Q_{mk}D_j \mid V_{m-1}]W_{(m-1)k}\,\epsilon_{mk}$ if $j = m$;

(b) $\Pi[G_m W_{mk}\,\epsilon_{(m+1)k} \mid (\tilde{\Lambda}_{mk}^4 + \tilde{\Lambda}_{mk}^5)^\perp] = \mathcal{L}_1 + \mathcal{L}_2$ where $\mathcal{L}_1 = \{G_m - T_{mk}^{-1}E[G_m W_{mk} \mid X_m]\}W_{mk}\,\epsilon_{(m+1)k}$ and $\mathcal{L}_2 = G_{(m-1)k}W_{(m-1)k}\,\epsilon_{mk}$ with $G_{(m-1)k} = E[G_m W_{mk} T_{mk}^{-1} \mid V_{m-1}]$. [Note $\mathcal{L}_1 = 0$ if $g(V_m)$ does not depend on A_m.]

(c) $\Pi[D_m^* T_{mk}^{-1} W_{mk}\,\epsilon_{(m+1)k} \mid \tilde{\Lambda}_{mk}^{4,\perp}] = E[D_m^* T_{mk}^{-1} \mid V_{m-1}](T_{mk}^* T_{mk})^{-1} W_{mk}\,\epsilon_{(m+1)k}$.

<u>Proof</u>: We first prove Theorem (A4.2a). $\tilde{\Lambda}_{mk}^5 = \{O(B) = O_2 O_1(B)\,;\, B = b(V_K)\}$ where $O_1(B) = E[B \mid X_m] - E[B \mid V_{m-1}]$ and $O_2(B) = \tilde{A}_{mk}^5(B)$. Then $O^\dagger(B) = O_1^\dagger O_2^\dagger(B)$ where "$\dagger$" denotes an operator adjoint. $O_1^\dagger(B) = O_1(B)$ and $O_2^\dagger(B) = B - E[B(T_{mk}^* T_{mk})^{-1} W_{mk}\,\epsilon_{(m+1)k} \mid V_{m-1}]Q_{mk}$. Hence $O^\dagger(D_j) = 0$ if $j<m$ and $O^\dagger(D_j) = D_j$ if $j = m$. Now, by standard Hilbert space results, $\Pi[D_j \mid \tilde{\Lambda}_{mk}^5] = \tilde{A}_{mk}^5(A_m^*)$ where A_m^* satisfies the "normal equation" $O^\dagger O(A_m^*) = O^\dagger(D_j)$. Now

$$O^\dagger O(A_m^*) = A_m^* + \nu_{m-1}(T_{mk}^*)^{-1}Q_{mk} \tag{A4.2}$$

where $\nu_{m-1} = E[Q_{mk}A_m^* \mid V_{m-1}]$. Thus $A_m^* = 0$ if $j<m$. Further, for $j = m$,

$$D_j = A_m^* + \nu_{m-1}(T_{mk}^*)^{-1}Q_{mk} \quad . \tag{A4.3}$$

Upon multiplying both sides of (A4.3) by Q_{mk} and taking conditional expectations given V_{m-1}, we obtain $E[Q_{mk}D_j \mid V_{m-1}] = \nu_{m-1} + \nu_{m-1}(T_{mk}^*)^{-1}\text{var}[Q_{mk} \mid V_{m-1}]$. Hence $\nu_{m-1} = E[Q_{mk}D_j \mid V_{m-1}]T_{mk}^* W_{(m-1)k}$ since $\{T_{mk}^* + \text{var}(Q_{mk} \mid V_{m-1})\}^{-1} = \{\text{var}[\epsilon_{mk} \mid V_{m-1}]\}^{-1} = W_{(m-1)k}$. Now $\tilde{A}_{mk}^5(A_m^*) \equiv A_m^* - \nu_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk}\,\epsilon_{(m+1)k} = D_j - \nu_{m-1}(T_{mk}^*)^{-1}[T_{mk}^{-1} W_{mk}\,\epsilon_{(m+1)k} + Q_{mk}]$ by (A4.2). Thus $\Pi[D_j \mid \tilde{\Lambda}_{mk}^{5,\perp}] = D_j - \tilde{A}_{mk}^5(A_m^*) = \{E[Q_{mk}D_j \mid V_{m-1}]T_{mk}^* W_{(m-1)k}\}(T_{mk}^*)^{-1}\epsilon_{mk}$, proving part (a).

We now turn to part (b). Given $G_m = g(V_m)$ and $\tilde{\Lambda}_{mk}^4 \equiv \{\tilde{A}_{mk}^4(B_m) = B_m W_{mk}\,\epsilon_{(m+1)k}\,;\, B_m = b(X_m)\}$, $\Pi[G_m W_{mk}\,\epsilon_{(m+1)k} \mid (\tilde{\Lambda}_{mk}^5 + \tilde{\Lambda}_{mk}^4)^\perp] = \mathcal{L}_1^* + \mathcal{L}_2^*$ where

$\mathcal{L}_1^* = \Pi[G_m W_{mk} \epsilon_{(m+1)k} \mid \hat{\Lambda}_{mk}^{4,\perp}]$, $\mathcal{L}_2^* = \Pi[\Pi[\Pi[G_m W_{mk} \epsilon_{(m+1)k} \mid \hat{\Lambda}_{mk}^4 \mid \tilde{\Lambda}_{mk}^{4,\perp} \mid \tilde{\Lambda}_{mk}^{5,\perp}]]]$ since $\tilde{\Lambda}_{mk}^4 \subset \hat{\Lambda}_{mk}^4$ and one can check that $\tilde{\Lambda}_{mk}^{4,\perp} \perp \tilde{\Lambda}_{mk}^5$. Now it is easy to show that

$$\Pi[G_m W_{mk} \epsilon_{(m+1)k} \mid \hat{\Lambda}_{mk}^4] = E[G_m W_{mk} \mid X_m] T_{mk}^{-1} W_{mk} \epsilon_{(m+1)k} \tag{A4.4}$$

so $\mathcal{L}_1 = \mathcal{L}_1^*$. After applying Theorem (A4.2c), it only remains to prove for any $G_{m-1} = g(V_{m-1})$,

$$\Pi[G_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k} \mid \tilde{\Lambda}_{mk}^{5,\perp}] = G_{m-1} W_{(m-1)k} \epsilon_{mk} \ . \tag{A4.5}$$

Combining (A4.4) and (A4.5), we obtain $\mathcal{L}_2 = \mathcal{L}_2^*$ proving part (b). It remains to show (A4.5). Now the L.H.S. of (A4.5) equals

$$G_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k} - \tilde{\Lambda}_{mk}^5(A_m^*) \tag{A4.6}$$

where A_m^* satisfies $O^\dagger O(A_m^*) = O^\dagger[G_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k}]$. Now

$O^\dagger[G_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k}] = -G_{m-1}(T_{mk}^*)^{-1} E[T_{mk}^{-1} W_{mk} T_{mk}^{-1} \mid V_{m-1}] T_{mk}^{*-1} Q_{mk} = -G_{m-1}(T_{mk}^*)^{-1} Q_{mk}$.

Hence, by (A4.2),

$$-G_{m-1}(T_{mk}^*)^{-1} Q_{mk} = A_m^* + \nu_{m-1}(T_{mk}^*)^{-1} Q_{mk} \ . \tag{A4.7}$$

By multiplying (A4.7) by Q_{mk} and taking conditional expectations given V_{m-1}, we obtain

$$\nu_{m-1} = -G_{m-1}(T_{mk}^*)^{-1} \text{var}[Q_{mk} \mid V_{m-1}] T_{mk}^* \{T_{mk}^* + \text{var}[Q_{mk} \mid V_{m-1}]\}^{-1} \ . \tag{A4.8}$$

Now (A4.6) equals, by (A4.7), $G_{m-1}(T_{mk}^*)^{-1}(T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k} + G_{m-1}(T_{mk}^*)^{-1} Q_{mk} + \nu_{m-1}(T_{mk}^*)^{-1} Q_{mk} + \nu_{m-1}(T_{mk}^* T_{mk})^{-1} W_{mk} \epsilon_{(m+1)k} = (T_{mk}^*)^{-1}[G_{m-1} + \nu_{m-1}] T_{mk}^{-1} W_{mk} \epsilon_{(m+1)k} + (G_{m-1} + \nu_{m-1})(T_{mk}^*)^{-1} Q_{mk} = (G_{m-1} + \nu_{m-1})(T_{mk}^*)^{-1} \epsilon_{mk} =$
$G_{m-1}(T_{mk}^*)^{-1}[1 - \text{var}(Q_{mk} \mid V_{m-1})/\{\text{var}(Q_{mk} \mid V_{m-1}) + T_{mk}^*\}] \epsilon_{mk} = G_{m-1} W_{(m-1)k} \epsilon_{mk}$ where the second to last equality uses (A4.8) and the last uses $W_{(m-1)k} = [\text{var}(Q_{mk} \mid V_{m-1}) + T_{mk}^*]^{-1}$.

To prove Part (c), note the L.H.S. of Part (c) equals

$$D_m^* T_{mk}^{-1} W_{mk} \epsilon_{(m+1)k} - A_m^* W_{mk} \epsilon_{(m+1)k} \tag{A4.9}$$

where A_m^* satisfies

$$O_1^\dagger O_3^\dagger O_3 O_1(A_m^*) = O_1^\dagger O_3^\dagger[D_m^* T_{mk}^{-1} W_{mk} \epsilon_{(m+1)k}] \tag{A4.10}$$

since $\bar{\Lambda}_{mk}^4 = \{O_3 O_1(B)\}$ where $O_3(B) = \bar{\Lambda}_{mk}^4(B)$.

Evaluating both sides of (A4.10), we obtain $A_m^* W_{mk}^2 \epsilon_{(m+1)k}^2 - E[A_m^* T_{mk} \mid V_{m-1}] = D_m^* T_{mk}^{-1} W_{mk}^2 \epsilon_{(m+1)k}^2 - E[D_m^* \mid V_{m-1}]$. Taking conditional expectations w.r.t. X_m, we obtain $A_m^* T_{mk} - E[A_m^* T_{mk} \mid V_{m-1}] = D_m^* - E[D_m^* \mid V_{m-1}]$ and thus that

$$A_m^* = T_{mk}^{-1}\{E[A_m^* T_{mk} \mid V_{m-1}] + D_m^* - E[D_m^* \mid V_{m-1}]\} \tag{A4.11}$$

Taking conditional expectations w.r.t. V_{m-1} and solving for $E[A_m^* T_{mk} \mid V_{m-1}]$ we obtain

$$E[A_m^* T_{mk} \mid V_{m-1}] = E[D_m^* \mid V_{m-1}] - (T_{mk}^*)^{-1} E[D_m^* T_{mk}^{-1} \mid V_{m-1}] \tag{A4.12}$$

Substituting (A4.12) into (A4.11), we obtain

$$A_m^* = -E[D_m^* T_{mk}^{-1} \mid V_{m-1}](T_{mk}^* T_{mk})^{-1} + D_m^* T_{mk}^{-1} \qquad \text{(A4.13)}$$

Substituting (A4.13) into (A4.9) proves Part (c).

ACKNOWLEDGEMENTS

This work was supported in part by NIH grants 5R01ESO3405, R01AI32475, 5KO4ESOO180 and 5P30ES00002.

BIBLIOGRAPHY

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1993). Identification of causal effects using instrumental variables. Technical Report # 136, Department of Economics, Harvard University.

Chamberlain, G. (1987). "Asymptotic efficiency in estimation with conditional moment restrictions." *J. Econometrics*. 34, 305-334.

Chamberlain, G. (1988). Efficiency Bounds for Semiparametric Regression, Technical Report. Department of Economics, University of Wisconsin.

Chamberlain, G. (1992). "Comment: Sequential Moment Restrictions in Panel Data." *J. Business & Economic Statistics*, v. 10, no. 1, 20-26.

Chamberlain, G. (1993). Feedback in Panel Data Models (unpublished).

Dawid, A.P. (1979). "Conditional independence in statistical theory (with Discussion)." *J. Roy. Stat. Soc.*, B41, 1-31.

Efron, B. and Feldman, D. (1991). "Compliance as an explanatory variable in clinical trials." *J. Am. Statist. Assoc.*, 86, 9-26.

Heckman, J.J. and Robb, R. (1985). "Alternative methods for evaluating the impact of interventions." Longitudinal analysis of labor market data (J.J. Heckman and B. Singer, Eds.). London: Cambridge University Press, 156-246.

Holland, P. (1986). "Statistics and causal inference." *J. Am. Statist. Assoc.*, 81, 945-960.

Holland, P. (1988). "Causal inference, path analysis, and recursive structural equation models." Sociological Methodology (C. Clogg, Ed.). Washington, DC: American Sociological Association, 449-484.

Manski, C.F. (1988). Analog estimation methods in econometrics. NY: Chapman and Hall.

Mark, S.D. and Robins, J.M. (1993a). "Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model." *Statistics in Medicine*. (In press).

Mark, S.D. and Robins, J.M. (1993b). "A method for the analysis of randomized trials with compliance information: An application to the multiple risk factor intervention trial." Controlled Clinical Trials. NY: Elsevier. 14, 79-97.

Newey, W.K. (1990). "Semiparametric Efficiency Bounds," *J. Applied Econometrics*, 5, 99-135.

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," *Math Modelling*, 7, 1393-1512.

Robins, J.M. (1988). "Confidence intervals for causal parameters," *Statist. Med.*, 7, 773-785.

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

Robins, J.M., Tsiatis, A.A. (1991). "Correcting for non-compliance in randomized trials using rank preserving structural failure time models," *Commun. Statist.-Theory.*, 20, 2609-2631.

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

Robins, J. (1992). "Estimation of the time-dependent accelerated failure time model in the presence of confounding factors," *Biometrika*, 79, 321-34.

Robins, J.M. (1993a). "Analytic methods for HIV treatment and cofactor effects." Methodological Issues of AIDS Behavioral Research (D.G. Ostrow and R. Kessler, Eds.). NY: Plenum Press, 213-290.

Robins, J.M. (1993b). "Estimating the causal effect of a time-varying treatment on survival using a new class of failure time models," *Communications in Statistics*. (In Press).

Robins, J.M., Mark, S.D. and Newey, W.K. (1992). "Estimating exposure effects by modelling the expectation of exposure conditional on confounders," *Biometrics*, 48, 479-495.

Robins, J.M., Greenland, S. (1992). "Identifiability and exchangeability for direct and indirect effects," *Epidemiology*, 3, 143-155.

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

Robins, J.M., Rotnitzky, A., and Zhao, L-P. (1994). "Analysis of semiparametric regression models for repeated outcomes under the presence of missing data." *J. Am. Statist. Assoc.* (To appear).

Robins, J.M. & Rotnitzky, A. (1994). "Semiparametric efficiency in multivariate regression models with missing data." *J. Am. Statist. Assoc.* (To appear).

Robins, J.M., Rotnitzky, A., Zhao, L-P., and Lipsitz, S. (1994). "Estimation of regression coefficients when some regressors are not always observed," *J. Am. Statist. Assoc.* (To appear).

Robins, J.M. and Greenland, S. (1994). "Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment in an AIDS randomized trial," *J. Am. Statist. Assoc. Applications and Case Studies*. (To appear).

Rubin, D.B. (1978). "Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, 6, 34-58.

Received March, 1991; Revised March, 1994.