2-19-2009

Relaxing Latent Ignorability in the ITT Analysis of Randomized Studies with Missing Data and Noncompliance

L Taylor
University of Washington, taylorl@u.washington.edu

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Suggested Citation
http://biostats.bepress.com/uwbiostat/paper257

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors
Relaxing Latent Ignorability in the ITT Analysis of Randomized Studies with Missing Data and Noncompliance

Leslie Taylor\textsuperscript{1} and Xiao-Hua Zhou\textsuperscript{2,1}

\textsuperscript{1}Department of Biostatistics, University of Washington
F-600 Health Sciences Building, 1705 NE Pacific Street, Seattle, WA 98195-7232
taylorl@u.washington.edu (L. Taylor)

\textsuperscript{2}Northwest HSR\textsuperscript{2}D Center of Excellence, VA Puget Sound Health Care System
Metropolitan Park West, 1100 Olive Way #1400, Seattle, WA 98101
azhou@u.washington.edu (X. Zhou)

Abstract:

In this paper we consider the problem in causal inference of estimating the local complier average causal effect (CACE) parameter in the setting of a randomized clinical trial with a binary outcome, cross-over noncompliance, and unintentional missing data on the responses. We focus on the development of a moment estimator that relaxes the assumption of latent ignorability and incorporates sensitivity parameters that represent the relationship between potential outcomes and associated potential response indicators. If conclusions are insensitive over a range of logically possible values of the sensitivity parameters, then the number of interpretations of the data is reduced, and causal conclusions are more defensible. We illustrate our methods using a randomized encouragement design study on the effectiveness of an influenza vaccine.

Keywords: Causal Inference; Complier Average Causal Effect; Encouragement Design
1. Introduction

Well-designed randomized clinical trials are a powerful tool for investigating causal relationships and producing valid estimates of a causal effect of treatment. But in trials involving human subjects there are oftentimes problems of noncompliance and missing data which standard analyses either ignore, which can lead to biased estimators, or account for in such a way that the estimand can no longer be considered a causal effect. Rubin developed an approach to causal inference using potential outcomes (Rubin (1974, 1978)) that has been referred to as the Rubin Causal Model (Holland (1986)). This model provides a framework for defining the parameters of interest and correctly attributing the data observed between different treatment groups to causal effects of the treatments.

Method-of-moment estimators are useful in understanding where information comes from within the observed data and what assumptions help to identify the estimands of interest. Frangakis and Rubin (1999) developed a moment estimator for the complier average causal effect (CACE) in a setting where there was unintentional missing data and only the intervention group could receive the new treatment. Zhou and Li (2006) later extended these moment methods to a setting of cross-over noncompliance (i.e. intervention and control subjects could receive the new treatment) and missing data. This paper extends the results of Zhou and Li (2006) by developing the asymptotic theory for their moment estimator and examining its performance in finite samples and under deviations from model assumptions in the setting of a binary response, cross-over noncompliance, and unintentional missing data on the responses. This paper also focuses on the development of a moment estimator that relaxes the assumption of latent ignorability and incorporates sensitivity parameters that represent the relationship between potential outcomes and associated potential nonresponse indicators. These parameters are assumed known, and are allowed to take on a plausible range of values in order to assess the sensitivity of the conclusions to varying assumptions regarding this relationship.

Sections 2-4 introduce the setting, notation, and assumptions. In Section 5 we give the results of Zhou and Li (2006) for a CACE estimator derived under the assumption of latent ignorability, and we extend their results by deriving the asymptotic distribution of their estimator. Section 6 provides simulation results that examine the finite sample properties of the estimator under conditions that follow the assumptions, and then under certain deviations from these assumptions. In Section 7 we derive the CACE estimator and its asymptotic distribution when the latent ignorability assumption is relaxed and sensitivity parameters are introduced. In Section 8 we illustrate our methods using a randomized encouragement design study on the
effectiveness of an influenza vaccine.

2. Setting and Notation

The setting consists of a clinical trial with \( N \) subjects assigned to treatment \( Z \), where \( Z \) is an \( N \)-vector of treatment assignments with \( i \)th element \( Z_i \). In this setting \( Z_i = 1 \) if subject \( i \) is assigned the new treatment, and \( Z_i = 0 \) if assigned the control. Let \( D_i \) be the treatment received under the observed treatment assignment, where \( D_i = 1 \) if subject \( i \) received the new treatment and \( D_i = 0 \) if subject \( i \) received the control. Then let \( D(Z) \) be the vector of potential treatment receipts given the vector of treatment assignments \( Z \) with \( i \)th element \( D_i(Z) \). Let \( Y_i \) be the binary outcome for subject \( i \) under the observed treatment assignment and let \( Y_i(Z) \) be the binary potential outcome given the vector for treatment assignments \( Z \).

Let \( R_i \) be the binary indicator for response under the assigned treatment, equal to 1 if response \( Y_i \) was observed for subject \( i \) and 0 otherwise. Then let \( R_i(Z) \) be the binary indicator for response equal to 1 if response \( Y_i(Z) \) was observed for subject \( i \) and 0 otherwise, for a given vector of treatment assignments \( Z \). Then a random subset of the \( N \) subjects are assigned to treatment arm \( Z \). Table 1 provides a summary of the notation used throughout the paper.

3. Definition of Causal Estimands

We make the stable unit treatment value assumption (SUTVA) which allows us to write the potential outcomes as functions of \( Z_i \) rather than of the entire vector \( Z \). Formally the SUTVA states that \( D_i(Z) \) equals \( D_i(Z') \), \( Y_i(Z) \) equals \( Y_i(Z') \), and \( R_i(Z) \) equals \( R_i(Z') \) if \( Z_i = Z'_i \) which means that we can write \( D_i(Z) \), \( Y_i(Z) \), and \( R_i(Z) \) as \( D_i(Z_i) \), \( Y_i(Z_i) \), and \( R_i(Z_i) \), respectively. Under the SUTVA we can define the intention-to-treat (ITT) causal effect of \( Z \) on \( D \) as \( E[D_i(1) - D_i(0)] \).

We assume that compliance is all-or-none, meaning that any switching of treatments was done soon after randomization so that the subject is assumed to have completely taken the new treatment or the control. We can stratify the population into four compliance principal strata (Frangakis and Rubin (1999)) as determined...
by the value of the vector \([D_i(0), D_i(1)]\), where

\[
C_i = \begin{cases} 
  n \text{ (never-taker)} & \text{if } D_i(0) = D_i(1) = 0 \\
  a \text{ (always-taker)} & \text{if } D_i(0) = D_i(1) = 1 \\
  c \text{ (complier)} & \text{if } D_i(0) = 0 \text{ and } D_i(1) = 1 \\
  d \text{ (defier)} & \text{if } D_i(0) = 1 \text{ and } D_i(1) = 0
\end{cases}
\]

Note that unlike membership to the observed compliance strata, membership to these principal compliance strata (referred to as compliance types for the remainder of the paper) is unaffected by assigned treatment and therefore can be considered as a baseline covariate (Frangakis and Rubin (2002)). For our setting we make the assumption of monotonicity (Imbens and Angrist (1994)), where \(D_i(1) \geq D_i(0)\) for all subjects (i.e. there are no defiers) where compliance type is observable when \(Z_i \neq D_i\). Here subjects with observed \(Z_i = D_i = 0\) are a mixture of compliers and never-takers, and subjects with observed \(Z_i = D_i = 1\) are a mixture of compliers and always-takers.

Let \(\psi_{td} = P[C_i = t|Z_i = z, D_i = d]\) be the probability of compliance type \(t\) given the assigned treatment \(z\) and received treatment \(d\), and let \(\eta_{zt} = E[Y_i(z)|Z_i = z, C_i = t]\) be the conditional expectation of the outcome given treatment assignment \(z\) and compliance type \(t\). Then, under the monotonicity assumption, we define the ITT effect as

\[
\text{ITT} = \sum_{t \in \{n,a,c\}} \omega_t \text{ITT}_t
\]

where \(\omega_t = P(C_i = t)\) and \(\text{ITT}_t = E[Y_i(1) - Y_i(0)|C = t]\) is the average ITT effect of \(Z\) on \(Y\) for the subpopulation of compliance type \(t\). Noncompliers (never-takers and always-takers), by definition, do not carry information about the comparison between treatments. Thus we focus on the the subpopulation of compliers and define the complier average causal effect (CACE) to be \(\text{ITT}_c\), or

\[
\text{CACE} = E[Y_i(1) - Y_i(0)|C_i = c] = \eta_{1c} - \eta_{0c},
\]

which is the treatment effect among the subpopulation of compliers and the focus of the remainder of the paper. Table 1 provides a summary of the notation used throughout the paper.

### 4. Additional Assumptions

In addition to the SUTVA and monotonicity, there are two assumptions that are sometimes plausible
and that help facilitate inference: the compound exclusion restriction for never-takers and always-takers (Frangakis and Rubin (1999)) which generalizes the standard exclusion restriction (Angrist, Imbens, and Rubin (1996); Imbens and Rubin (1997)); and latent ignorability (Frangakis and Rubin, 1999). The compound exclusion restriction states that among the subpopulation of never-takers or always-takers, treatment assignment does not affect potential outcomes or missing data distributions, or $P[Y_i(1), R_i(1)|C_i = n] = P[Y_i(0), R_i(0)|C_i = n]$ and $P[Y_i(1), R_i(1)|C_i = a] = P[Y_i(0), R_i(0)|C_i = a]$. Next we invoke a latent ignorability assumption which states that, within each latent compliance type, potential outcomes and associated potential response indicators are independent, or $P[R_i(1), R_i(0)|Y_i(1), Y_i(0), C_i] = P[R_i(1), R_i(0)|C_i]$. We make the assumption of latent ignorability here because it is more plausible than the assumption of standard ignorability (Rubin (1978); Little and Rubin (1987)).

5. Asymptotic Theory of the CACE Moment Estimator

Under the SUTVA, monotonicity assumption, latent ignorability, and the compound exclusion restriction for never-takers and always-takers, the CACE is identifiable and Zhou and Li (2006) derived the following moment estimators

$$
\hat{\eta}_{1c} = \frac{\sum Y_i R_i Z_i D_i}{\sum R_i Z_i D_i} - \frac{\sum Y_i R_i (1 - Z_i) D_i}{\sum R_i (1 - Z_i) D_i}, \quad \hat{\eta}_{0c} = \frac{\sum Y_i R_i (1 - Z_i) (1 - D_i)}{\sum R_i (1 - Z_i) (1 - D_i)} - \frac{\sum Y_i R_i Z_i (1 - D_i)}{\sum R_i Z_i (1 - D_i)}.
$$

Then the estimator for the CACE computed by Zhou and Li (2006) is $\hat{\text{CACE}}^{LI} = \hat{\eta}_{1c} - \hat{\eta}_{0c}$. Note that in the first summation for $\hat{\eta}_{1c}$, contributions come from subjects with $Z_i = D_i = 1$, which consist of a mixture of compliers and always-takers. Since we are interested in the average among compliers, the averages for the always-takers (in the second term) are subtracted. Note that Imbens, Angrist, and Rubin (1996) develop an equivalent estimator for the case where there are no missing outcomes (i.e. $R_i = 1$ for all subjects) under the assumptions of SUTVA, monotonicity, and an exclusion restriction on outcomes only. Let $\pi_{zd} = P(R_i = 1, Z_i = z, D_i = d)$ denote the joint probability of observing the response with treatment assignment $z$ and treatment receipt $d$; and let $v_{zd} = P(Y_i = 1, R_i = 1, Z_i = z, D_i = d)$ denote the joint distribution of observing outcome $Y = 1$ with treatment assignment $z$ and treatment receipt $d$. The following theorem, proved in the online appendix (http://www.stat.sinica.edu.tw/statistica/submission) using the delta method, forms a basis for inference about the estimator.
Theorem 5.1. Under the assumptions of Section 3,

\[ \sqrt{n}(\hat{CACE}^{LI} - CACE) \rightarrow_d N(0, (V_0 + V_1)^\frac{1}{2}) \]

as \( n \rightarrow \infty \), where

\[ V_0 = \frac{A^2(3\pi_{10} - \pi_{00}) + A(\pi_{00} - \pi_{10} - 4v_{10}) + 2v_{10}}{(\pi_{00} - \pi_{10})^2}, \]
\[ V_1 = \frac{B^2(3\pi_{01} - \pi_{11}) + B(\pi_{11} - \pi_{01} - 4v_{01}) + 2v_{01}}{(\pi_{11} - \pi_{01})^2}, \]

for \( A = (v_{00} - v_{10})/(\pi_{00} - \pi_{10}) \) and \( B = (v_{11} - v_{01})/(\pi_{11} - \pi_{01}) \).

6. Simulation Study

In this section we examine some finite sample properties of this estimator, first under hypothetical conditions that follow the assumptions of latent ignorability and the compound exclusion restriction, and then under certain deviations from latent ignorability.

6.1. Numerical Results Under Latent Ignorability and the Compound Exclusion Restriction

The \( N = 300 \) subjects were randomized to the control or new treatment arm with \( P(Z_i = 1) = 0.5 \) where \( C_i \) was generated independently as a multinomial random variable. Subject outcomes \( Y_i \) were generated from a binomial distribution with a mean conditional upon treatment assignment \( Z_i \) and compliance type \( C_i \). We fixed average outcomes \( E[Y_i(1)|Z_i = 1, C_i = a] = E[Y_i(1)|Z_i = 1, C_i = n] = 0.5 \) for simplicity, which implies (by the compound exclusion restriction) that \( E[Y_i(0)|Z_i = 0, C_i = a] \) and \( E[Y_i(0)|Z_i = 0, C_i = n] \) equal 0.5 as well. We also fixed \( E[Y_i(1)|Z_i = 1, C_i = c] \) to be 0.5.

We varied the following parameters: proportions of compliance types, true CACE, and response probabilities for subjects. For the response probabilities, we either let the response probability for all compliance types equal 0.5, which gave us a missing at random (MAR) missing data mechanism, or we let the response prob-
abilities depend on latent compliance type (where $E(R_i(z)|Z_i = z, C_i = c) = E(R_i(z)|Z_i = z, C_i = a) = 0.5$ but $E(R_i(z)|Z_i = z, C_i = n) = 0.8$ for $z \in (0, 1)$), which gave us a not missing at random (NMAR) missing data mechanism since response probabilities depend on compliance type which is not observed for all subjects. Table 3 reports the coverage rates of nominal 95 percent confidence intervals and the bias for $C\hat{A}C_E^{LI}$. Note that when the response mechanism was MAR, $C\hat{A}C_E^{LI}$ performed well, giving good coverage and relatively little bias (Table 3). When the response mechanism was NMAR, the estimator continued to perform well. In both scenarios, true compliance type probabilities and true CACE values were not critically important in terms of the behavior of the estimator.

6.2. Numerical Results with Deviations from Latent Ignorability

Next we tested the sensitivity of our estimator to potential outcomes that are no longer independent of potential response indicators for subjects in the control arm. We defined the constant $f_{0t} = \frac{P(R_i(0)=1|Z_i=0,C_i=t,Y_i=0)}{P(R_i(0)=1|Z_i=0,C_i=t,Y_i=1)}$, where $f_{0t}$ represents the amount of dependence between potential outcomes and associated response indicators for subjects in the control arm. Note that $f_{0t} \equiv 1$ corresponds to having latent ignorability, and distance from $f_{0t}$ to 1 corresponds to the degree of dependence between outcomes and response indicators. We fixed the true CACE to zero and the response probabilities $P(R_i(z) = 1|Z_i = z, C_i = n) = P(R_i(z) = 1|Z_i = z, C_i = a) = 0.5$ and $P(R_i(z) = 1|Z_i = z, C_i = c) = 0.7$ for $z \in (0, 1)$.

We then varied the compliance type proportions and allowed $f_{0t}$ to vary between $\frac{1}{2}$ and 2 for all compliance types in the control arm (where $f_{0n} = f_{0c} = f_{0a}$). Table 4 reports the coverage rates of nominal 95 percent confidence intervals and the bias for $C\hat{A}C_E^{LI}$ When $f_{0t}$ was less than one, the estimator underestimated the true CACE, whereas for values greater than one, the estimator overestimated the true CACE. The further $f_{0t}$ was from one (meaning the more dependence between outcome and response), the worse the coverage probabilities. Higher proportions of compliers (relative to always-takers and never-takers) improved the bias somewhat but slightly worsened the coverage probabilities. Overall we see how sensitive our results can be when latent ignorability does not hold.

7. Relaxing the Latent Ignorability Assumption

7.1. Defining the Causal Parameters

Once again, for binary outcome $Y$, we let $\psi_{td} = P[C_i = t|Z_i = z, D_i = d]$ and $\eta_{zd} = E[Y_i(z)|Z_i = z, C_i = c] = \ldots$
as in Section 3, where we focus on the CACE (1). We invoke the SUTVA, monotonicity assumption, and the compound exclusion restriction for never-takers and always-takers. Letting \( \phi_{zt} = P(R_i(z) = 1|Z_i = z, C_i = t, Y_i(z) = y) \), we relax the assumption of latent ignorability and incorporate sensitivity parameters that represent the relationship between the potential outcomes and associated response indicators, where the sensitivity parameters \( f_{zt} \) are defined as

\[
f_{zt} = \frac{\phi_{zt0}}{\phi_{zt1}},
\]

and represent the ratio of response probabilities between subjects with outcome \( Y = 0 \) versus those with outcome \( Y = 1 \) (for a given assigned treatment \( z \) and compliance type \( t \)). These parameters are assumed known and are allowed to take on a plausible range of values in order to assess the sensitivity of the conclusions of a study to various assumptions regarding the relationship between outcomes and response indicators. If conclusions are insensitive over a range of logically possible values for \( f_{zt} \), then the number of interpretations of the data is reduced, and causal conclusions are more defensible.

Under monotonicity, there are no defiers, and under the compound exclusion restrictions, \( \eta_{1a} = \eta_{0a} \) for always-takers and \( \eta_{1n} = \eta_{0n} \) for never-takers. Letting \( \eta_{1a} = \eta_{0a} \equiv \eta_a \) and \( \eta_{1n} = \eta_{0n} \equiv \eta_n \), note that \( \eta_{zt} \) can be specified in terms of \( \eta_{0c}, \eta_{1c}, \eta_a, \) and \( \eta_n \). Next we note that \( \psi_{a0} = \psi_{a00} = 0 \) since one cannot be an always-taker if one receives the control, and \( \psi_{n1} = \psi_{n01} = 0 \) since one cannot be a never-taker if one receives the new treatment. Similarly \( \psi_{01} = \psi_{c10} = 0 \) since, for compliers, \( Z_i = D_i \). Since only always-takers have \( D_i = 1 \) with \( Z_i = 0 \) and only never-takers have \( D_i = 0 \) then \( Z_i = 1 \), then \( \psi_{a0} = \psi_{n01} = 1 \). Also note that \( \psi_{c00} + \psi_{n00} = \psi_{c11} + \psi_{a11} = 1 \). Then letting \( \psi_a \equiv \psi_{a11} \) and \( \psi_n \equiv \psi_{n00} \), note that \( \psi_z \) can be specified in terms of \( \psi_a \) and \( \psi_n \). And since the compound exclusion restriction implies that response probabilities for always-takers (or never-takers) do not depend on treatment assignment, \( \phi_{0n1}(\eta_n + f_{0n}(1-\eta_n)) = \phi_{1n1}(\eta_n + f_{1n}(1-\eta_n)) \) and \( \phi_{0a1}(\eta_a + f_{0a}(1-\eta_a)) = \phi_{1a1}(\eta_a + f_{1a}(1-\eta_a)) \), so that \( \phi_{zt} \) can be specified in terms of \( \phi_{0n1}, \phi_{1n1}, \phi_{0a1}, \) and \( \phi_{1a1} \). Then let \( \theta = (\psi_a, \psi_n, \eta_a, \eta_n, \phi_{a11}, \phi_{0n1}, \phi_{0a1}, \phi_{1c1}, \phi_{0c1}, \phi_{1c1}, \eta_{0c}, \eta_{1c}) \).

7.2. Estimation

Let \( \xi_{zd} = P(Z_i = z, D_i = d) \), \( \pi_{zd} = P(R_i = 1, Z_i = z, D_i = d) \), and \( v_{zd} = P(Y_i = 1, R_i = 1, Z_i = z, D_i = d) \). Then with \( N \) subjects in the study, let \( \hat{\xi}_{zd} = \frac{1}{N} \sum_{i=1}^{n} I[Z_i = z, D_i = d] \), \( \hat{\pi}_{zd} = \frac{1}{N} \sum_{i=1}^{n} R_i I[Z_i = z, D_i = d] \), and \( \hat{v}_{zd} = \frac{1}{N} \sum_{i=1}^{n} Y_i R_i I[Z_i = z, D_i = d] \) be unbiased estimators for \( \xi_{zd}, \pi_{zd}, \) and \( v_{zd} \), respectively. The following
result, proved in the online appendix (http://www.stat.sinica.edu.tw/statistica/submission), defines the moment estimators when latent ignorability is relaxed.

**Result 7.1.** Under the assumptions of Section 3, the estimators for the parameters in the always-taker and never-taker subpopulations are:

\[
\hat{\psi}_a = \frac{\hat{\xi}_{01}}{\xi_{11}}; \quad \hat{\psi}_n = \frac{\hat{\xi}_{1n}}{\xi_{01}}
\]

\[
\hat{\eta}_a = \frac{f_{0a}\hat{\xi}_{01}}{\pi_{01} + (f_{0a} - 1)\pi_{01}}; \quad \hat{\eta}_n = \frac{f_{1n}\hat{\xi}_{10}}{\pi_{10} + (f_{1n} - 1)\pi_{10}}
\]

\[
\hat{\phi}_{0a1} = \frac{\hat{\phi}_{01}}{\xi_{01}\eta_a}; \quad \hat{\phi}_{1n1} = \frac{\hat{\nu}_{10}}{\xi_{10}\eta_n}
\]

\[
\hat{\phi}_{1a1} = \frac{\hat{\phi}_{0a1}(\eta_a + f_{0a}(1 - \eta_a))}{\eta_a + f_{1a}(1 - \eta_a)}; \quad \hat{\phi}_{0n1} = \frac{\hat{\phi}_{1n1}(\eta_n + f_{1n}(1 - \eta_n))}{\eta_n + f_{0n}(1 - \eta_n)}
\]

The estimators for the parameters in the complier subpopulation are:

\[
\hat{\phi}_{0c1} = \frac{\hat{\phi}_{0n1}\hat{\psi}_n\hat{\xi}_{00}(f_{0c}\hat{\eta}_n + f_{0n}(1 - \hat{\eta}_n)) + (1 - f_{0c})\hat{\psi}_n - \hat{\pi}_{00}}{f_{0c}\hat{\xi}_{00}(\hat{\psi}_n - 1)}
\]

\[
\hat{\phi}_{1c1} = \frac{\hat{\phi}_{1a1}\hat{\psi}_a\hat{\xi}_{11}(f_{1a}\hat{\eta}_a + f_{1a}(1 - \hat{\eta}_a)) + (1 - f_{1c})\hat{\psi}_1 - \hat{\pi}_{11}}{f_{1c}\hat{\xi}_{11}(\hat{\psi}_a - 1)}
\]

\[
\hat{\eta}_{1c} = \frac{f_{1c}(f_{1a}\hat{a}_{01} + f_{0a}\hat{a}_{11})}{(f_{1c} - 1)(f_{1a}\hat{a}_{01} + f_{0a}\hat{a}_{11}) + f_{1a}\hat{b}_{01} + f_{0a}\hat{c}_{01}}
\]

\[
\hat{\eta}_{0c} = \frac{f_{0c}(f_{1n}\hat{a}_{10} + f_{0n}\hat{a}_{00})}{(f_{0c} - 1)(f_{1n}\hat{a}_{10} + f_{0n}\hat{a}_{00}) + f_{1n}\hat{b}_{10} + f_{0n}\hat{c}_{10}}
\]

\[
\hat{a}_{zd} = \hat{v}_{zd}(\hat{\psi}_{(1-z)d} - \hat{\pi}_{zd})
\]

\[
\hat{b}_{zd} = \hat{v}_{zd}(\hat{\pi}_{(1-z)d} - \hat{\pi}_{zd})
\]

\[
\hat{c}_{zd} = (\hat{\pi}_{zd} - \hat{v}_{zd})(\hat{\pi}_{(1-z)d} - \hat{\pi}_{zd})
\]

The estimator for the CACE derived without the LI assumption, but under known (fixed) sensitivity parameters, is \(\text{CAE}^{LI} = \hat{\eta}_{1c} - \hat{\eta}_{0c}\).

Note that the parameter estimates from the never-taker and always-taker subpopulations generally involve summations over subjects with observed \(Z_i \neq D_i\); the parameter estimate from the complier subpopulation, \(\hat{\eta}_{1c}\), incorporates a mixture of summations across subjects with observed \(Z_i = D_i = 1\) (which consist of a mixture of compliers and always-takers) and subjects with \(Z_i = 0\) and \(D_i = 1\) (the ob-
served always-takers); similarly $\tilde{\eta}_0c$ incorporates a mixture of summations across subjects with observed $Z_i = D_i = 0$ (which consist of a mixture of compliers and never-takers) and subjects with observed $Z_i = 1$ and $D_i = 0$ (the observed never-takers). Note that if $f_{1a} = f_{0a}$ (or $f_{1n} = f_{0n}$), then neither contribute to the estimator $\hat{CACE}$. Since moment estimators are non-parametric, they can unfortunately be outside the (-1,1) range of the estimand of interest. The following theorem, proved in the online appendix (http://www.stat.sinica.edu.tw/statistica/submission) using the delta method, forms a basis for inference about $\hat{CACE}$.

**Theorem 7.1** Under the assumptions of Section 3

$$\sqrt{n}(\hat{CACE} - CACE) \rightarrow_d N(0, \delta^t V_0 \delta + \beta^t V_1 \beta)$$

as $n \rightarrow \infty$, for

$$V_0 = \begin{pmatrix}
\pi_{00}(1 - \pi_{00}) & -\pi_{00}\pi_{10} & (1 - \pi_{00})\pi_{00} & -\pi_{00}\pi_{10} \\
-\pi_{10}\pi_{00} & \pi_{10}(1 - \pi_{10}) & -\pi_{10}\pi_{00} & (1 - \pi_{10})\pi_{10} \\
\pi_{00}(1 - \pi_{00}) & -\pi_{00}\pi_{10} & \pi_{10}(1 - \pi_{00}) & -\pi_{00}\pi_{10} \\
-\pi_{10}\pi_{00} & \pi_{10}(1 - \pi_{10}) & -\pi_{10}\pi_{00} & \pi_{10}(1 - \pi_{10})
\end{pmatrix},$$

$$V_1 = \begin{pmatrix}
\pi_{11}(1 - \pi_{11}) & -\pi_{11}\pi_{01} & (1 - \pi_{11})\pi_{11} & -\pi_{11}\pi_{01} \\
-\pi_{01}\pi_{11} & \pi_{01}(1 - \pi_{01}) & -\pi_{01}\pi_{11} & (1 - \pi_{01})\pi_{01} \\
-\pi_{11}\pi_{01} & \pi_{11}(1 - \pi_{11}) & -\pi_{11}\pi_{01} & \pi_{11}(1 - \pi_{11}) \\
-\pi_{01}\pi_{11} & \pi_{01}(1 - \pi_{01}) & -\pi_{01}\pi_{11} & \pi_{01}(1 - \pi_{01})
\end{pmatrix}.$$}

In this result, $\delta = (\delta_1, \ldots, \delta_4)^t$ is defined as follows:

$$\delta_1 = D_0^{-2}f_{0c}[f_{1n}\pi_{00}(\pi_{00} - \pi_{10}) + f_{0n}\pi_{00}(\pi_{10} - \pi_{10}) - f_{1n}\pi_{10}];$$

$$\delta_2 = D_0^{-2}f_{0c}[f_{1n}\pi_{00}(\pi_{00} - \pi_{10}) + f_{0n}\pi_{00}(\pi_{10} - \pi_{10}) - f_{1n}\pi_{10} + D_0^{-2}f_{0c}f_{1n}\pi_{10}^2(f_{1n} - f_{0n})(\pi_{10} - \pi_{00});$$

$$\delta_3 = D_0^{-2}f_{0c}[f_{0n}(\pi_{10} - \pi_{10}) + f_{1n}\pi_{10}]^2(\pi_{00} - \pi_{10});$$

$$\delta_4 = D_0^{-2}f_{0c}f_{1n}f_{0n}\pi_{10}^2(\pi_{10} - \pi_{00});$$

for $D_0 = (f_{0c} - 1)[f_{1n}\pi_{10}(\pi_{00} - \pi_{10}) + f_{0n}\pi_{00}(\pi_{10} - \pi_{10})] + (\pi_{00} - \pi_{10})[f_{1n}\pi_{10} + f_{0n}(\pi_{10} - \pi_{10})]$;

$\beta = (\beta_1, \ldots, \beta_4)^t$ is defined as follows:

$$\beta_1 = D_1^{-2}f_{1c}[f_{1a}\pi_{01}(\pi_{11} - \pi_{01}) + f_{0a}\pi_{11}(\pi_{01} - \pi_{01})](f_{0a}(\pi_{01} - \pi_{01}) - f_{1a}\pi_{01});$$

$$\beta_2 = D_1^{-2}f_{1c}[f_{1a}\pi_{01}(\pi_{11} - \pi_{01}) + f_{0a}\pi_{11}(\pi_{01} - \pi_{01})](f_{0a}(\pi_{01} - \pi_{01}) - f_{1a}\pi_{01}) + D_1^{-2}f_{1c}f_{1a}\pi_{01}^2(f_{1a} - f_{0a})(\pi_{11} - \pi_{01});$$

$$\beta_3 = D_1^{-2}f_{1c}[f_{0a}(\pi_{01} - \pi_{01}) + f_{1a}\pi_{01}]^2(\pi_{11} - \pi_{01});$$

$$\beta_4 = D_1^{-2}f_{1c}f_{1a}f_{0a}\pi_{01}^2(\pi_{01} - \pi_{11});$$

for $D_1 = (f_{1c} - 1)[f_{1a}\pi_{01}(\pi_{11} - \pi_{01}) + f_{0a}\pi_{11}(\pi_{01} - \pi_{01})] + (\pi_{11} - \pi_{01})[f_{1a}\pi_{01} + f_{0a}(\pi_{01} - \pi_{01})]$.
7.3. Simulation Study Results

In Table 4, data was generated under the model described in section 6 and therefore $\hat{CACE}$ was estimated under the assumed known sensitivity parameters. As expected, $\hat{CACE}$ performed well in this scenario with decent coverage and relatively little bias.

8. Influenza Vaccination Study

Among patients who are older or have a high risk of pulmonary disease, observational studies and experimental evidence suggest that those vaccinated with an influenza vaccine have better outcomes (McDonald, Hui, and Tierney (1992)). A controlled clinical trial to confirm these results has never been performed because of the ethical problems that arise from withholding the vaccine from patients in the control arm. A solution to this problem involves performing a controlled clinical trial where the intervention arm increases the use of the influenza vaccine without changing its use in the control arm. McDonald, Hui, and Tierney (1992) used this method to study the effects of computer-generated reminders of the influenza vaccine on flu-related hospitalizations in patients having a high risk for pulmonary disease. For doctors in the intervention arm, computer reminders were sent out when a patient with a scheduled visit was eligible for a flu shot. Since the study did not maintain records on the clustering of patients by doctor, we ignore this for the purposes of illustrating our methods. In this analysis we want to estimate the effect of the flu vaccine on flu-related hospitalizations (where $Y_i = 1$ if subject $i$ had a flu-related hospitalization and $Y_i = 0$ otherwise). There were missing outcomes, but no information was given on how the data came to be missing. The data are provided in Table 2.

Under latent ignorability, (where the sensitivity parameters equal 1 for all compliance types and treatment groups), the estimate of the CACE is 0.01 with 95% confidence interval (-0.25, 0.26) indicating that there was no significant decrease in hospitalizations as a result of receiving the flu vaccine. We illustrate the application of the proposed methods by presenting a sensitivity analysis where the CACE is estimated under differing assumptions regarding the dependence between outcomes and response indicators. Since no information was given on how the data came to be missing, we considered four scenarios for testing the sensitivity of our estimator to deviations in latent ignorability across assigned treatment group and compliance type. Results from the following scenarios are found in Figure 1 (a-d) where point estimates and 95% confidence intervals for the CACE are displayed, assuming specified values of the sensitivity parameters. Note that,
although we use 2-dimensional plots to illustrate the sensitivity analysis, 3-dimensional plots would allow the user to vary two sensitivity parameters simultaneously.

8.1. Scenario I

To see how sensitive the estimator is to deviations in latent ignorability among compliers in the control arm, we let $f_{0n} = f_{1n}$ and $f_{0a} = f_{1a}$, fixed $f_{1c} = 1$, and allowed $f_{0c}$ to vary between 1/2 and 2. Results are displayed in Figure 1a. Note that when the sensitivity parameter equals 1 for compliers in the control arm, latent ignorability is assumed. The estimate of the CACE did not change much as the sensitivity parameters for control compliers were varied.

8.2. Scenario II

To see how sensitive the estimator is to deviations in latent ignorability among never-takers in the control arm, we let $f_{0c} = f_{1c} = f_{0a} = f_{1a} = f_{1n} = 1$, and allowed $f_{0n}$ to vary between 1/2 and 2. Results are displayed in Figure 1b. The estimate of the CACE changed as the sensitivity parameters were varied, although there remained no significant decrease in hospitalizations as the confidence intervals all contain zero.

8.3. Scenario III

To see how sensitive the estimator is to deviations in latent ignorability among always-takers in the control arm, we let $f_{0c} = f_{1c} = f_{0a} = f_{1a} = f_{1n} = 1$, and allowed $f_{0a}$ to vary between 1/2 and 2. Results are displayed in Figure 1c. The estimate of the CACE did not change much as the sensitivity parameters were varied.

8.4. Scenario IV

To see how sensitive the estimator is to deviations in latent ignorability among all subjects in the control arm, we fixed the sensitivity parameters to 1 for those in the treatment arm ($f_{1n} = f_{1c} = f_{1a} = 1$) and varied the sensitivity parameters between 1/2 and 2 for those in the control arm ($1/2 < f_{0n} = f_{0c} = f_{0a} < 2$). Results are displayed in Figure 1d. For patients in the control arm, when the probability of observing an outcome given a flu-related hospitalization differed from the probability of observing the outcome given no flu-related hospitalization (i.e. $f_{0n} = f_{0c} = f_{0a} \neq 1$), the estimate of the CACE varied considerably, although standard errors increased as the dependence increased between the outcome and reponse indicators. In fact,
it was only when the probability of observing an outcome given a flu-related hospitalization was two times the probability of observing the outcome given no flu-related hospitalization, that the interval estimate of the CACE excluded zero: -0.56 (95% CI: -0.92 to -0.20). Thus we find that the CACE point estimate is somewhat sensitive to reasonable deviations in latent ignorability across treatment group, although the CACE was not significantly different from zero.

8.5. Summarizing Results

Another way in which a sensitivity analysis can be summarized is a 95% sensitivity interval (Rosenbaum (1999)) defined here to be the union of all 95% confidence intervals for the CACE for varying values of \( f_{zt} \) that we are confident contain the true \( f_{zt} \). It has a similar property to the confidence interval in that if the assumption about the range in which the sensitivity parameter lies is correct, then it will contain the true parameter of interest at least 95% of the time (Rosenbaum, 1999). In the flu vaccine example, there is no information on how the data came to be missing, but typically this information could help determine an accurate range for the sensitivity parameters.

9. Discussion

There were some limitations in applying these methods to the flu vaccination study. One may question the validity of the compound exclusion restriction used to identify the causal parameter of interest, particularly for the always-takers. While it may make sense to assume that treatment assignment had no direct effect on outcome (given treatment received) for never-takers, it may not make sense to assume that treatment assignment had no direct effect on outcome for always-takers. Never-takers may be the healthier patients since their doctors might not encourage the vaccination under either treatment assignment. For these patients, the assignment to treatment should not lead doctors to take other measures that could directly affect outcome. On the other hand, always-takers may be the sicker patients because their doctor might encourage them to get the flu vaccination regardless of assigned treatment. For these patients, the added impact of being assigned to the encouragement arm may lead the doctor to encourage other precautionary measures beyond the flu vaccination which could directly affect the patient’s outcome. The setting or application will determine which values of the sensitivity parameters are considered plausible. In the case of the flu vaccine study, no information was given on how the data came to be missing, so we used a wide range of values for the
sensitivity parameters.

Future research topics could include methods that incorporate baseline covariates which are often collected in a randomized clinical trial, referencing work by Levy, O’Malley, and Normand (2004), as well as methods that incorporate clustering effects commonly found in encouragement design studies.

Acknowledgement

Leslie Taylor, Ph.D., is presently a Mathematical Statistician at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. Xiao-Hua Zhou, Ph.D., is presently a Core Investigator and Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. This work was supported in part by NIH/NHLBI grant R01HL62567. We wish to thank the referees for their careful review of the manuscript that helped produce a much improved paper.

Figure 1: Sensitivity analysis for the influenza vaccination study (point estimates and 95% confidence intervals for the CACE are displayed): (a) Scenario I: no LI for control compliers; (b) Scenario II: no LI for control never-takers; (c) Scenario III: no LI for control always-takers; (d) Scenario IV: no LI for all control subjects.
### Table 1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Specifics</th>
<th>General Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_i$</td>
<td>1 if $i$ assigned treatment, 0 if $i$ assigned control</td>
<td>Treatment assignment indicator</td>
</tr>
<tr>
<td>$D_i(Z_i)$</td>
<td>1 if $i$ received treatment under assignment $Z_i$, 0 if $i$ received control under assignment $Z_i$</td>
<td>Potential outcome formulation of treatment receipt</td>
</tr>
<tr>
<td>$D_i$</td>
<td>Treatment receipt indicator under observed assignment</td>
<td></td>
</tr>
<tr>
<td>$C_i$</td>
<td>n if $D_i(0) = 0$ and $D_i(1) = 0$, c if $D_i(0) = 0$ and $D_i(1) = 1$, a if $D_i(0) = 1$ and $D_i(1) = 1$, d if $D_i(0) = 1$ and $D_i(1) = 0$</td>
<td>Compliance type principal stratum: n=never-taker; c=complier; a=always-taker; d=defier</td>
</tr>
<tr>
<td>$Y_i(Z_i)$</td>
<td>Binary outcome of interest under assignment $Z_i$</td>
<td>Potential outcome formulation of the outcome of interest</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>Binary outcome of interest under observed assignment</td>
<td></td>
</tr>
<tr>
<td>$R_i(Z_i)$</td>
<td>1 if $Y_i(Z_i)$ would be observed, 0 if $Y_i(Z_i)$ would not be observed</td>
<td>Response indicator for $Y_i(Z_i)$</td>
</tr>
<tr>
<td>$R_i$</td>
<td>Response indicator for $Y_i$ under observed assignment</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Influenza Vaccine Data

<table>
<thead>
<tr>
<th></th>
<th>$R=1, Z=0$</th>
<th>$Y=0$</th>
<th>$Y=1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D=0$</td>
<td>573</td>
<td>49</td>
<td>622</td>
</tr>
<tr>
<td></td>
<td>$D=1$</td>
<td>143</td>
<td>16</td>
<td>159</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>716</td>
<td>65</td>
<td>781</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$R=1, Z=1$</th>
<th>$Y=0$</th>
<th>$Y=1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D=0$</td>
<td>499</td>
<td>47</td>
<td>546</td>
</tr>
<tr>
<td></td>
<td>$D=1$</td>
<td>256</td>
<td>20</td>
<td>276</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>755</td>
<td>67</td>
<td>822</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$R=0, Y=$</th>
<th></th>
<th>$D=0$</th>
<th>$D=1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Z=0$</td>
<td>492</td>
<td>17</td>
<td>509</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$Z=1$</td>
<td>497</td>
<td>9</td>
<td>506</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>989</td>
<td>26</td>
<td>1015</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Simulation results: N=300 with 5,000 replications of the data

<table>
<thead>
<tr>
<th>CACE ( (n, c, a) )</th>
<th>MAR</th>
<th>NMAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coverage</td>
<td>Bias</td>
</tr>
<tr>
<td>0</td>
<td>94.8</td>
<td>0.002</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>95.6</td>
<td>0.002</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>96.5</td>
<td>0.003</td>
</tr>
<tr>
<td>(0.25,0.5,0.25)</td>
<td>94.9</td>
<td>0.002</td>
</tr>
<tr>
<td>0.2</td>
<td>95.5</td>
<td>0.005</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>96.3</td>
<td>0.006</td>
</tr>
<tr>
<td>(0.25,0.5,0.25)</td>
<td>95.4</td>
<td>0.002</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>95.8</td>
<td>0.007</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>96.6</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 4: Simulation Results: N=300 with 5,000 replications of the data \( (CACE = 0; \hat{CACE}^{LI} = \text{estimator assuming latent ignorability}; \hat{CACE}^{LTI} = \text{estimator assuming no latent ignorability}) \)

<table>
<thead>
<tr>
<th>( f_{00} ) ( (n, c, a) )</th>
<th>( \hat{CACE}^{LI} )</th>
<th>( \hat{CACE}^{LTI} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coverage</td>
<td>Bias</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>35.4</td>
<td>-0.220</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>38.4</td>
<td>-0.249</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>39.7</td>
<td>-0.292</td>
</tr>
<tr>
<td>(0.25,0.5,0.25)</td>
<td>82.7</td>
<td>-0.093</td>
</tr>
<tr>
<td>( \frac{1}{4} )</td>
<td>84.8</td>
<td>-0.105</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>85.8</td>
<td>-0.125</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>94.8</td>
<td>-0.001</td>
</tr>
<tr>
<td>(0.25,0.5,0.25)</td>
<td>95.4</td>
<td>-0.002</td>
</tr>
<tr>
<td>1</td>
<td>95.9</td>
<td>-0.001</td>
</tr>
<tr>
<td>( \frac{1}{4} )</td>
<td>83.4</td>
<td>0.095</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>84.0</td>
<td>0.109</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>83.9</td>
<td>0.127</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>35.6</td>
<td>0.218</td>
</tr>
<tr>
<td>(0.15,0.7,0.15)</td>
<td>36.4</td>
<td>0.250</td>
</tr>
<tr>
<td>(0.2,0.6,0.2)</td>
<td>40.0</td>
<td>0.292</td>
</tr>
</tbody>
</table>
Appendix

1. Proof of Theorems 5.1 and 7.1

In order to derive the asymptotic distributions for $CACE^L$ and $CACE^T$ let $X_i \equiv (Y_i R_i (1-Z_i)(1-D_i), Y_i R_i Z_i (1-D_i), Y_i R_i Z_i D_i, R_i (1-Z_i)(1-D_i), R_i Z_i (1-D_i), R_i Z_i D_i)$ and let $p \equiv (\nu_0, \nu_1, \nu_{10}, \nu_{11}, \pi_0, \pi_1, \pi_{10}, \pi_{11})$ where $X_i$ is a vector of bernoulli random variables with mean $p$. If $X_1, \ldots, X_n$ are i.i.d. random vectors in $R^8$ with mean $p$ and covariance matrix $\Sigma = E(X-p)(X-p)$ (where $E(X^T X) < \infty$), then from the multivariate central limit theorem, $\sqrt{n}(X-p) \rightarrow_d N_d(0, \Sigma)$ where $\Sigma = (\nu_0 (1-\nu_0), -\nu_0 \nu_1, -\nu_0 \nu_{10}, -\nu_0 \nu_{11}, \nu_0 (1-\nu_0), -\nu_0 \nu_1, -\nu_0 \nu_{10}, -\nu_0 \nu_{11})$

1.1 Proof of Theorem 5.1

Using the multivariate delta method, the asymptotic distributions for $\hat{\eta}_0c$ and $\hat{\eta}_1c$ are:

$\sqrt{n}(\hat{\eta}_1c - \eta_1c) \rightarrow_d N(0, \delta' \Sigma \delta)$ where $\delta' = (0, \frac{-1}{\pi_{11}-\pi_{01}}, 0, \frac{1}{\pi_{11}-\pi_{01}}, 0, \frac{\nu_1-\nu_{10}}{(\pi_{11}-\pi_{01})}, 0, \frac{\nu_0-\nu_{11}}{(\pi_{11}-\pi_{01})})$ and $\sqrt{n}(\hat{\eta}_0c - \eta_0c) \rightarrow_d N(0, \beta' \Sigma \beta)$ where $\beta' = (\frac{-1}{\pi_{10}-\pi_{00}}, 0, \frac{1}{\pi_{10}-\pi_{00}}, 0, \frac{\nu_0-\nu_{10}}{(\pi_{10}-\pi_{00})}, 0, \frac{\nu_0-\nu_{10}}{(\pi_{10}-\pi_{00})})$.

where $V_1 = \delta' \Sigma \delta$ and $V_0 = \beta' \Sigma \beta$ are defined in section 4. Then noting that $\sqrt{n}(\hat{\eta}_1c - \eta_1c)$ and $\sqrt{n}(\hat{\eta}_0c - \eta_0c)$ are asymptotically independent since $\delta' \Sigma \beta = 0$, we can use Slutsky’s theorem to derive the asymptotic distribution for $CACE^L$ as in Theorem 5.1.

1.2 Proof of Theorem 7.1

Using the multivariate delta method, the asymptotic distributions for $\hat{\eta}_1c$ and $\hat{\eta}_0c$ are $\sqrt{n}(\hat{\eta}_1c - \eta_1c) \rightarrow_d N(0, \delta' \Sigma \delta)$, where $\delta' = (0, \delta_1, 0, \delta_2, 0, \delta_3, 0, \delta_4)$ for $\delta$ defined in Theorem 7.1, and $\sqrt{n}(\hat{\eta}_0c - \eta_0c) \rightarrow_d N(0, \beta' \Sigma \beta)$ where $\beta' = (0, \beta_1, 0, \beta_2, 0, \beta_3, 0, \beta_4)$ for $\beta$ defined in Theorem 7.1. Again noting that $\sqrt{n}(\hat{\eta}_1c - \eta_1c)$ and $\sqrt{n}(\hat{\eta}_0c - \eta_0c)$ are asymptotically independent, we can use Slutsky’s theorem to derive the asymptotic distribution for $CACE^T$ as in Theorem 7.1.
2. Proof of Result 7.1

For \( \psi_n = P(C = n | Z = 0, D = 0) \) note that under randomization \( P(C = n, Z = 0) = P(C = n, Z = 1) \) and since never-takers by definition have \( D = 0 \), then we expect \( P(C = n, Z = 0, D = 0) = P(C = n, Z = 1, D = 0) \). Similarly for always-takers, \( P(C = a, Z = 1, D = 1) = P(C = a, Z = 0, D = 1) \).

Therefore:

\[
\psi_n = \frac{P(C = n, Z = 0, D = 0)}{P(Z = 0, D = 0)} = \frac{P(C = n, Z = 1, D = 0)}{P(Z = 0, D = 0)} = \frac{P(Z = 1, D = 0)}{P(Z = 0, D = 0)}
\]

\[
\psi_a = \frac{P(C = a, Z = 1, D = 1)}{P(Z = 1, D = 1)} = \frac{P(C = a, Z = 0, D = 1)}{P(Z = 1, D = 1)} = \frac{P(Z = 0, D = 1)}{P(Z = 1, D = 1)}
\]

Thus we can obtain the following estimators for \( \psi_n \) and \( \psi_a \):

\[
\hat{\psi}_n = \frac{\hat{\xi}_{10}}{\hat{\xi}_{00}}; \quad \hat{\psi}_a = \frac{\hat{\xi}_{01}}{\hat{\xi}_{11}}
\]

Then note the following relationship:

\[
v_{zd} = P(Y = 1, R = 1, Z = z, D = d)
\]

\[
= \sum_{t \in \{n,c,a\}} P(R = 1, Z = z, D = d, C = t, Y = 1)
\]

\[
= \sum_{t \in \{n,c,a\}} P(R = 1 | Z = z, D = d, C = t, Y = 1) P(Y = 1 | Z = z, D = d, C = t)
\]

\[
P(C = t | Z = z, D = d) P(Z = z, D = d)
\]

\[
\Leftrightarrow v_{zd} = \sum_{t \in \{n,c,a\}} P(R = 1 | Z = z, D = d, C = t, Y = 1)
\]

\[
P(Y = 1 | Z = z, D = d, C = t) P(C = t | Z = z, D = d)
\]

\[
= \sum_{t \in \{n,c,a\}} \phi_{zt1} \eta_{zt} \psi_{zd}
\]

(2)

Also note the following relationship:

\[
\frac{\pi_{zd}}{\xi_{zd}} = P(R = 1 | Z = z, D = d)
\]

\[
= \sum_{t \in \{n,c,a\}} P(R = 1 | Z = z, D = d, C = t) P(C = t | Z = z, D = d)
\]
\[
\begin{align*}
\sum_{t \in (n,c,a)} \left( \sum_{y \in (0,1)} P(R = 1|Z = z, D = d, C = t, Y = y)P(Y = y|Z = z, D = d, C = t) \right) \frac{1}{P(C = t|Z = z, D = d)} \\
= \sum_{t \in (n,c,a)} P(R = 1|Z = z, D = d, C = t, Y = 1)P(Y = 1|Z = z, D = d, C = t) \\
& \quad P(C = t|Z = z, D = d) + P(R = 1|Z = z, D = d, C = t, Y = 0) \\
& \quad P(Y = 0|Z = z, D = d, C = t)P(C = t|Z = z, D = d) \\
= \sum_{t \in (n,c,a)} \phi_{zt1}\eta_{zt}\psi_{zd} + \phi_{zt0}(1 - \eta_{zt})\psi_{zd} \\
= \sum_{t \in (n,c,a)} \frac{\phi_{zt1}\eta_{zt}\psi_{zd} + f_{zt}\phi_{zt1}(1 - \eta_{zt})\psi_{zd}}{\pi_{zt1}} \\
= \sum_{t \in (n,c,a)} \phi_{zt1}\psi_{zd}(\eta_{zt} + f_{zt}(1 - \eta_{zt})) \\
\end{align*}
\]

(3)

Letting \((z, d)\) equal \((0, 1)\) and \((1, 0)\) in expression (3) above and solving for \(\phi_{0a1}\) and \(\phi_{1n1}\) we get:

\[
\phi_{0a1} = \frac{v_{01}}{\xi_{01}\eta_{a}}; \quad \phi_{1n1} = \frac{v_{10}}{\xi_{10}\eta_{n}}
\]

(4)

Letting \((z, d)\) equal \((0, 1)\) and \((1, 0)\) in expression (4) above and substituting the expressions from (5):

\[
\frac{\pi_{01}}{\xi_{01}} = \frac{v_{01}}{\xi_{01}\eta_{a}}(\eta_{a} + f_{0a}(1 - \eta_{a})); \quad \frac{\pi_{10}}{\xi_{10}} = \frac{v_{10}}{\xi_{10}\eta_{n}}(\eta_{n} + f_{1n}(1 - \eta_{n}))
\]

Then solving for \(\eta_{a}\) and \(\eta_{n}\) above, we obtain the following estimators:

\[
\hat{\eta}_{a} = \frac{f_{0a}\hat{v}_{01}}{\hat{\pi}_{01} + (f_{0a} - 1)v_{01}}; \quad \hat{\eta}_{n} = \frac{f_{1n}\hat{v}_{10}}{\hat{\pi}_{10} + (f_{1n} - 1)v_{10}}
\]

We can then obtain estimators for \(\phi_{0a1}\) and \(\phi_{1n1}\):

\[
\hat{\phi}_{0a1} = \frac{\hat{v}_{01}}{\hat{\xi}_{01}\hat{\eta}_{a}}; \quad \hat{\phi}_{1n1} = \frac{\hat{v}_{10}}{\hat{\xi}_{10}\hat{\eta}_{n}}
\]

Using the exclusion criteria for never-takers and always-takers and (2), note that:

\[
\phi_{0n1}(\eta_{n} + f_{0n}(1 - \eta_{n})) = \phi_{1n1}(\eta_{n} + f_{1n}(1 - \eta_{n}))
\]
\[ \phi_{1a}(\eta_a + f_{1a}(1 - \eta_a)) = \phi_{0a1}(\eta_a + f_{0a}(1 - \eta_a)) \]

Then solving for \( \phi_{0n1} \) and \( \phi_{1a1} \) above we obtain the following estimators:

\[
\hat{\phi}_{1a1} = \frac{\hat{\phi}_{0a1}(\hat{\eta}_a + f_{0a}(1 - \hat{\eta}_a))}{\hat{\eta}_a + f_{1a}(1 - \hat{\eta}_a)}; \quad \hat{\phi}_{0n1} = \frac{\hat{\phi}_{1n1}(\hat{\eta}_a + f_{1n}(1 - \hat{\eta}_n))}{\hat{\eta}_n + f_{0n}(1 - \hat{\eta}_n)}
\]

Letting \((z, d) = (0, 0)\) in expressions (3) and (4) and solving for \( \eta_{0c} \) and \( \phi_{0c1} \) we obtain:

\[
\phi_{0c1} = \frac{\phi_{0n1}\psi_n\xi_{00}(f_{0c}\eta_n + f_{0n}(1 - \eta_n)) + (1 - f_{0c})v_{00} - \pi_{00}}{\xi_{00}f_{0c}(\psi_n - 1)}
\]

\[
\eta_{0c} = \frac{f_{0c}(\phi_{0n1}\psi_n\xi_{00}\eta_n - v_{00})}{\phi_{0n1}\psi_n\xi_{00}(f_{0c}\eta_n + f_{0n}(1 - \eta_n)) + (1 - f_{0c})v_{00} - \pi_{00}}
\]

Letting \((z, d) = (1, 1)\) in expression (3) and (4) and solving for \( \eta_{1c} \) and \( \phi_{1c1} \) we obtain:

\[
\phi_{1c1} = \frac{\phi_{1a1}\psi_a\xi_{11}(f_{1a}\eta_a + f_{1a}(1 - \eta_a)) + (1 - f_{1c})v_{11} - \pi_{11}}{\xi_{11}f_{1c}(\psi_a - 1)}
\]

\[
\eta_{1c} = \frac{f_{1c}(\phi_{1a1}\psi_a\xi_{11}\eta_a - v_{11})}{\phi_{1a1}\psi_a\xi_{11}(f_{1a}\eta_a + f_{1a}(1 - \eta_a)) + (1 - f_{1c})v_{11} - \pi_{11}}
\]

Substituting in the expressions for \( \phi_{0n1}, \psi_n, \eta_n, \phi_{1a1}, \psi_a, \) and \( \eta_a \) we obtain the following expressions for \( \hat{\eta}_{0c} \) and \( \hat{\eta}_{1c} \) (which reduce to the expressions found in Result 7.1):

\[
\hat{\eta}_{0c} = \frac{f_{0c}[f_{1n}\hat{\psi}_{10}(\hat{v}_{00} - \hat{\pi}_{10}) + f_{0n}\hat{\psi}_{00}(\hat{\pi}_{10} - \hat{v}_{10})]}{(f_{0c} - 1)[f_{1n}\hat{\psi}_{10}(\hat{v}_{00} - \hat{\pi}_{10}) + f_{0n}\hat{\psi}_{00}(\hat{\pi}_{10} - \hat{v}_{10})] + f_{1n}\hat{\psi}_{10}(\hat{\pi}_{00} - \pi_{10}) + f_{0n}(\hat{\pi}_{10}\hat{\pi}_{00} - \hat{\pi}_{10}^2 - \hat{v}_{10}\hat{\pi}_{00} + \hat{v}_{10}\hat{\pi}_{10})}
\]

\[
\hat{\eta}_{1c} = \frac{f_{1c}[f_{1a}\hat{\psi}_{10}(\hat{v}_{11} - \hat{\pi}_{01}) + f_{0a}\hat{\psi}_{11}(\hat{\pi}_{01} - \hat{v}_{01})]}{(f_{1c} - 1)[f_{1a}\hat{\psi}_{10}(\hat{v}_{11} - \hat{\pi}_{01}) + f_{0a}\hat{\psi}_{11}(\hat{\pi}_{01} - \hat{v}_{01})] + f_{1a}\hat{\psi}_{10}(\hat{\pi}_{11} - \pi_{01}) + f_{0a}(\hat{\pi}_{01}\hat{\pi}_{11} - \hat{\pi}_{01}^2 - \hat{v}_{01}\hat{\pi}_{11} + \hat{v}_{01}\hat{\pi}_{01})}
\]
References


