UW Biostatistics Working Paper Series

3-30-2005

ITT Analysis of Randomized Encouragement Design Studies with Missing-Data

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

Richard Wang

Sierra M. Li
University of Washington, sierrali@u.washington.edu

Suggested Citation
Zhou, Xiao-Hua; Wang, Richard; and Li, Sierra M., "ITT Analysis of Randomized Encouragement Design Studies with Missing-Data" (March 2005). *UW Biostatistics Working Paper Series.* Working Paper 245.
http://biostats.bepress.com/uwbiostat/paper245

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

ITT Analysis of Randomized Encouragement Design Studies with Missing-Data

Xiao-Hua Zhou, Richard Wang and Sierra M. Li
Department of Biostatistics
University of Washington
F600, HOB, Box 357232
Seattle, WA 98195-7232

KEY WORDS: Causal Inference; Flu Shots; Non-compliance; Intention-to-Treat Analysis; and Missing-data.

Abstract

In this paper we considered a missing outcome problem in causal inferences for a randomized encouragement design study. We proposed both moment and maximum likelihood estimators for the marginal distributions of potential outcomes and the local complier average causal effect (CACE) parameter. We illustrated our methods in a real randomized encouragement design study on the effectiveness of flu shots.

1. Introduction

A traditional randomized clinical trial is the established gold standard for estimating the causal effects of treatments. However, for some treatments, it is impossible to perform such a randomized trial due to ethical and other reasons. For example, in our flu shot reminder study, it would be unethical to randomize high-risk adult patients to receive or not to receive flu shots in order to evaluate causal effects of having flu on patient outcomes. A better way to learn about the treatment causal effects is to perform a randomized encouragement design study (EDS). A randomized EDS randomly assigns subjects to receive or not to receive an encouragement for the use of a treatment. In recent years, there has been a rapid growth in utilizing encouragement designs to study causal treatment effects. Since the randomization to encouragement leads to a natural instrumental variable under some plausible assumptions, the randomized EDS provides a tool for estimating causal effects of the treatment on patient outcomes.

There are three basic elements in a randomized controlled EDS that make statistical analysis more difficult than in standard drug studies. The first is an encouragement to perform suggested

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

clinical actions is randomly given to clinicians; the second is subsequent clinical actions are recorded; and the third element is that final patient outcomes are recorded. The focus of the analysis is the causal effects of the encouraged clinical actions on subsequent patient outcomes. One major analytic issue in the EDS is non-compliance of subjects with suggested treatments, and there could be a large number of subjects with non-compliance. For example, in the randomized EDS of flu shots, patients' physicians were randomly assigned to whether to receive reminders about giving flu shots to their patients. A physician could decide to give flu shots to his/her patients, regardless of whether he/she received a reminder.

Rubin [10, 11, 12] developed a framework for the causal analysis in randomized trials with non-compliance using potential outcomes. This approach allows one to estimate the causal effect of the treatment actually received on patient outcomes. Neyman [8] first formalized the ideas of using potential outcomes in randomized trials to define the effect of the treatment actually received on outcomes [12]. Rubin's causal model [5] extended Neyman's formulation to observational studies and formulated an explicit model for the assignment mechanism exhibiting possible dependence on all potential outcomes. The resulting model allows one to use frequentist and Bayesian methods for inferring causal effects in studies of all kinds. For all-or-none compliance studies, both the likelihood approach and Bayesian approach have been proposed. For example, Cuzick et al [2] developed likelihood methods for causal analysis of studies with all-or-none compliance. Imbens and Rubin [6], and Hirano et al. [4] developed likelihood and Bayesian methods for causal analysis of randomized studies without any missing outcome values.

When some patients were missing their outcomes, Frangakis and Rubin [3] developed a moment estimator for the intention-to-treat (ITT) effect and the local complier average causal effect (CACE) of the treatment on the binary outcome in a simple two-treatment randomized trial in which only intervention patients could receive the new treatment (e.g. only control subjects could be non-compliers). When outcomes were continuous, Yau and Little [13] extended Frangakis and Rubin's method to a longitudinal study which had only compliers and never-takers, and Peng et al. [9] further extended Yau and Little's results to the case when the missing value could occur in both the outcomes and baseline covariates.

Barnard et al. [1] extended Frangakis and Rubin's method to a longitudinal study with crossover non-compliance (e.g. both intervention and control subjects could be non-compliers) when outcomes were continuous. Their method focused on parametric models (e.g. a censored normal model for

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

their outcomes).

In this paper, we focused our attention on binary outcomes and generalize the moment method proposed in Frangakis and Rubin [3] to a randomized clinical trial with crossover non-compliance and missing data, as in our flu shot reminder study. We also developed a maximum likelihood (ML) approach for this type of non-compliance problem with missing data. As in the previous papers, we focused our attention on the local complier average causal effect (CACE). This paper is organized as follows. In Section 2, we describe a randomized encouragement design study on flu shot, which has motivated this research. In Sections 3, we provide some basic notation needed for defining the Rubin's causal model, and in Section 4, we define complier types based on potential outcomes on the treatments actually received. In Section 5, we define causal effect parameters. In Section 6, we give necessary assumptions for making causal inferences when both the problems of non-compliance and missing data exist. In Sections 7, 8, and 9, we derive both the moment and maximum likelihood (ML) estimators for marginal distributions of potential outcomes as well as causal effect parameters. In Section 10 we present the results from applying the proposed method to our flu shot data set. All technical details are presented in Appendix A.

2. A Real Study

From observational studies, it has been shown that vaccination has improved outcomes in vaccinated patients. Hence, health officials in most countries recommend annual influenza vaccination for elderly persons and other people at high risk of influenza. However, no controlled randomized trials of the effects of influenza vaccination on pulmonary morbidity in high-risk adults have been published (McDonald et al, 1992). One possible reason for the lack of randomized trials is that widely accepted recommendations for vaccination raise ethical barriers against performing randomized controlled trials because it would require withholding vaccination from some subjects.

One way around this impasse would be to perform a randomized trial of an intervention that increases the use of influenza vaccine in one group of patients without changing the use of influenza vaccine in another group. McDonald and his colleagues [7] have used this idea to study influenza vaccine efficacy in reducing morbidity in high-risk adults, using a computer-generated reminder for flu shots (McDonald et al, 1992). The study was conducted over a three-year period (1978-1980) in an academic primary care practice affiliated with a large urban public teaching hospital. Physicians in the practice were randomly assigned to either an intervention or a control group at

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

the beginning of the study. Since physicians at the clinic each care for a fixed group of patients, their patients were similarly classified. During the study period, physicians in the intervention group received a computer generated reminder when a patient with a scheduled appointment was eligible for a flu shot. Our data set involved randomization of an encouragement by doctor, and doctors had multiple patients. However, the study did not keep information on the clustering of patients by doctor. Hence, for the purpose of illustrating the proposed method and avoiding an overly cumbersome analysis involving unknown cluster indicators, we assume exchangeability of patients. Each subject's randomized intervention, actual treatment receipt, and outcome of interest constitute the minimal data necessary to analyze this type of study. The example data are shown in Table 1 below. Note that when Y_i is observed $R_i = 1$ by definition, so the first two tables include those subjects with full data $(R = 1, Y, Z, D)$ and the last table includes those subjects with incomplete data $(R = 0, Z, D)$.

Table 1: Example Data

R=1	Z=0,D=0	Z=0,D=1	Z=1,D=0	Z=1,D=1
Y=0	573	143	499	256
Y=1	49	16	47	20
R=0, Y=·	492	17	497	9

The main research question in the study is to estimate the causal effect of having a flu shot on flu-related patient outcomes, such as flu-related hospitalization. Our hypothesis is that having flu shots will significantly decrease flu-related hospitalization. To assess the causal effect of flu shot on the outcome, we have to deal with the both problems of non-compliance and missing outcomes.

3. Basic Notation

Additional notation is needed to clearly state our models. In our flu shot reminder study, we are interested in assessing the effect of flu shot versus no flu shot (D) on the flu-related hospitalization (Y). We have no control over who would receive a flu shot, but we have control over who would receive a reminder for flu shot (Z). For subject i, let Z_i denote whether subject i's physician has received the reminder for flu shot; $Z_i = 1$ if subject i's physician receives a reminder, and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

$Z_i = 0$ otherwise. Let $D_i(z)$ be the binary indicator for whether subject i receives flu shot given the reminder status z of subject i, z=0,1. That is, $D_i(z) = 1$ if subject i would receive the flu shot given the reminder status z, and $D_i(z) = 0$ if subject i would not receive the flu shot given the reminder status z. Let $\underline{D}_i = (D_i(1), D_i(0))$ denote the vector of potential flu shot outcomes for subject i. Let $Y_i(z, D_i(z))$ be the binary indicator for flu-related hospitalization if subject i has the flu shot status $D_i(z)$ and the reminder z. Then, $\underline{Y}_i = (Y_i(1, D_i(1)), Y_i(0, D_i(0)))$ is a vector of potential outcomes for subject i. Note that although we used a double-argument notation for the outcome $Y_i(z, D_i(z))$, it is actually a function z only. Our observed data would consist of Z_i, $D_i = D_i(Z_i)$, and $Y_i = Y_i(Z_i, D_i(Z_i))$, $i = 1, \ldots, n$. Finally, we define the response indicator, $R_i(z)$, for the outcome Y_i of subject i. That is, $R_i(z) = 1$ if Y_i is observed given that subject i's physician receives the reminder z, and $R_i(z) = 0$ if Y_i is missing given that the subject i's physician receives the reminder status z.

To ease the notation burden, we define some statistics of the observed data Z_i, D_i, Y_i, and R_i. Let $N_{zd} = \sum_{i=1}^{N} I_{[Z_i=z, D_i=d]} R_i$, which represents the total number of subjects in the sample who have observed outcomes and are assigned to the treatment z and actually receive the treatment d. Also let $r_{zd} = \sum_{i=1}^{N} I_{[Z_i=z, D_i=d]} R_i Y_i$, which is the total number of hospitalized subjects assigned to treatment z, receiving treatment d, and having the observed outcome. Let $M_{zd} \equiv \sum_{i=1}^{N} I_{[R_i=0, Z_i=z, D_i=d]} = \sum_{i=1}^{N} I_{[Z_i=z, D_i=d]}(1 - R_i)$, which is the total number of subjects in the sample who are missing their outcomes and are assigned to the treatment z and actually receive the treatment d.

4. Complier types

One key idea in causal inferences in Rubin's model is to define the complier type, C_i, of subject i based on potential outcomes $D_i(1)$ and $D_i(0)$. Because $D_i(1)$ and $D_i(0)$ each can take two values, the complier type C_i has four different values, n for never-takers, a for always-takers, c for compliers, and d for defiers:

$$C_i \equiv \begin{cases} n \text{ if } D_i(1) = D_i(0) = 0; \\ a \text{ if } D_i(1) = D_i(0) = 1; \\ c \text{ if } D_i(0) = 0 \text{ and } D_i(1) = 1; \\ d \text{ if } D_i(1) = 0 \text{ and } D_i = 0. \end{cases}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Subjects whose treatment status is always the opposite of their treatment assignment are referred to as defiers. Also, we assume all-or-none compliance: subjects are assumed to have a fixed, though possibly unknown, complier type which does not change. If we further assume the monotonicity assumption that $D_i(1) \geq D_i(0)$ for all subjects i (i.e. no defiers), the complier status is observable when $Z_i \neq D_i$. When $(Z_i, D_i) = (1, 0)$, we know the subject is a never-taker; similarly, when $(Z_i, D_i) = (0, 1)$, we know the subject is an always-taker. For subjects with $(Z_i, D_i) = (0, 0)$, we cannot tell whether the subject is a never-taker or a complier; similarly, for subjects with $(Z_i, D_i) = (1, 1)$ we cannot tell whether the subject is an always-taker or a complier. Here arises the principal difficulty with using only marginal quantities to estimate causal parameters: some observed data arise from a mixture of the outcome distributions of different complier types. Let us define the probability of complier type given the observed treatment assignment and receipt as:

$$\psi_{tzd} \equiv P[C_i = t | Z_i = z, D_i = d]$$

. We denote the probability of assigning a subject to the treatment group as:

$$\xi \equiv P[Z_i = 1] \text{ and } \omega_t = P(C_i = t).$$

5. Causal Parameters of Interest

In order to define our causal parameters, we need to make some further assumptions. The first basic assumption is the Stable Unit treatment Value Assumption (SUTVA), which assumes no interference between study patients. Under the SUTVA, we can define intention-to-treat (ITT) causal effect of Z on Y as $ITT_Y = E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)))$. Note that this ITT causal effect parameter can be decomposed as $ITT_Y = \sum_{t=\{c,n,a\}} P(C_i = t) ITT_Y^t$, where $ITT_Y^t = E((Y_{ij}(1, D_{ij}(1)) - Y_{ij}(0, D_{ij}(0))) \mid C_i = t)$, which is the ITT effect of reminder for flu shot on the flu-related hospitalization in the subpopulation of patients with compliance types $C_i = t$. Of the three subpopulation ITT effects, neither ITT_Y^n nor ITT_Y^a address causal effects of D (the receipt of flu shot) on Y because ITT_Y^n compares the outcomes of two groups both without flu shots, and ITT_Y^a compares the outcomes of two groups both having flu shots. Only for the complier subpopulation, ITT_Y^c, compares outcomes with having flu shots to outcomes without having flu shots. ITT_Y^c is also called the local complier average causal effect of D on Y, denoted by CACE; that is,

$$CACE = E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \mid C_i = c).$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

In this article, we concentrate on estimation of the $CACE$.

6. Further Assumptions and their Implications in Missing Outcomes

With additional missing values in the outcome, to estimate causal parameters we need to make following additional assumptions.

1. Latent ignorability: potential outcomes and associated potential non-response indicators are independent within each latent complier type.

$$P[R_i(1), R_i(0)|C_i, Y_i(1, D_i(1)), Y_i(0, D_i(0))] = P[R_i(1), R_i(0)|C_i].$$

2. Compound exclusion restrictions for never-takers and always-takers: for never-takers and always-takers, Z influences Y and R only through D. The compound exclusion restriction for never-takers embodies the twin assumptions that, within the subpopulation of never-takers, the distributions of the two potential outcomes, $P[Y(1,0) = 1|Z = 1, D = 0]$ and $P[Y(0,0) = 1|Z = 0, D = 0]$, are identical, and that the distributions of the two missing-data indicators $P[R(1) = 1|Z = 1, D = 0]$ and $P[R(0) = 1|Z = 0, D = 0]$ are also identical. In other words, the treatment assignment does not affect the outcome or missing-data distributions for never-takers. The compound exclusion restriction for always-takers, comprises the twin assumptions that, within the subpopulation of always-takers, the treatment assignment does not affect the outcome or missing-data distributions. Let us define $\eta_{zt} = P(Y_i(z, D_i(z)) = 1 \mid Z_i = z, C_i = t)$ as the conditional probability of outcome given the complier type and treatment assignment, and $\gamma_{zt} \equiv P[R_i = 1|Z_i = z, C_i = t]$ as the probability of observing outcome Y for subjects with $Z_i = z$ and $C_i = t$. Then we can express the compound exclusion assumption as $\eta_{1n} = \eta_{0n} \equiv \eta_n$ and $\gamma_{1n} = \gamma_{0n} \equiv \gamma_n$.

7. Estimation

In this section, we discuss methods for estimating the parameters ψ_{zdt}, γ_{zt}, and ω_t. Under our assumptions, the complier status, C_i, is known when $Z_i \neq D_i$. For example, $C_i = n$ when $(Z_i, D_i) = (1, 0)$ and $C_i = a$ when $(Z_i, D_i) = (0, 1)$. The observations for which $Z_i = D_i$, on the other hand, are mixtures of different compliance strata, and the proportion of each compliance stratum within the observed cell is unobservable. When $(Z_i, D_i) = (0, 0)$, the subject can be either

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

a complier or a never-taker; when $(Z_i, D_i) = (1, 1)$ the subject can be either a complier or an always-taker. Hence, many of the conditional probabilities of complier types given the observed data are known. In fact, among ψ_{zdt}'s, only ψ_{n00} and ψ_{a11} are unknown parameters. To simplify notation we denote the two unknown parameters as $\psi_n = \psi_{n00}$ and $\psi_a = \psi_{a11}$. If the unconditional probabilities of complier types are of interest, we can compute them from the parameters ψ_{tzd} and the observed data. Therefore, the marginal distributions of potential outcomes and the missing data mechanism are defined by the eleven parameters, ξ, ψ_n, ψ_a, η_n, η_a, η_{0c}, η_{1c}, γ_n, γ_a, γ_{0c}, and γ_{1c}. Let θ be a vector of those parameters; that is, $\theta = (\xi, \psi_n, \psi_a, \eta_n, \eta_a, \eta_{0c}, \eta_{1c}, \gamma_n, \gamma_a, \gamma_{0c}, \gamma_{1c})$. Next, we propose two methods for estimating the vector of parameters θ.

8. Moment Estimators

We apply the method of moments to obtain estimators of the causal parameters of interest, using a principal stratification with the binary outcome. The method of moments equates sample moments to population moments and solves the resulting equations for the parameters of interest. In Appendix A, we show the following proposition regarding the moment estimator for θ.

Proposition 1 *Under the assumptions made in Sections 5 and 6, the moment estimators for the eleven components in θ are given as follows:*

1. *For parameters in non-complier subpopulations, the moment estimators are*

$$\widehat{\xi} = \frac{N_{10} + M_{10} + N_{11} + M_{11}}{N_{00} + M_{00} + N_{10} + M_{10} + N_{01} + M_{01} + N_{11} + M_{11}}$$

$$\widehat{\gamma_n} = \frac{N_{10}}{N_{10} + M_{10}}, \qquad \widehat{\gamma_a} = \frac{N_{01}}{N_{01} + M_{01}},$$
$$\widehat{\eta_n} = \frac{r_{10}}{N_{10}}, \qquad \widehat{\eta_a} = \frac{r_{01}}{N_{01}},$$
$$\widehat{\psi_n} = \frac{N_{10} + M_{10}}{N_{00} + M_{00}}, \qquad \widehat{\psi_a} = \frac{N_{01} + M_{01}}{N_{11} + M_{11}}.$$

2. *For parameters in complier subpopulations, the moment estimators are*

$$\begin{aligned}
\widehat{\gamma_{0c}} &= \frac{1}{1 - \widehat{\psi_n}} \left[\frac{N_{00}}{N_{00} + M_{00}} - \widehat{\gamma_n}\widehat{\psi_n} \right], \\
\widehat{\gamma_{1c}} &= \frac{1}{1 - \widehat{\psi_a}} \left[\frac{N_{11}}{N_{11} + M_{11}} - \widehat{\gamma_a}\widehat{\psi_a} \right], \\
\widehat{\eta_{0c}} &= \frac{1}{1 - \widehat{\gamma_{0c}}\widehat{\psi_n}} \left\{ \left[\widehat{\gamma_n}\widehat{\psi_n} + \gamma_{0c}\widehat{(1 - \psi_n)} \right] \cdot \frac{r_{00}}{N_{00}} - \widehat{\eta_n}\widehat{\gamma_n}\widehat{\psi_n} \right\}, \\
\widehat{\eta_{0c}} &= \frac{1}{1 - \widehat{\gamma_{1c}}\widehat{\psi_a}} \left\{ \left[\widehat{\gamma_a}\widehat{\psi_a} + \gamma_{1c}\widehat{(1 - \psi_a)} \right] \cdot \frac{r_{11}}{N_{11}} - \widehat{\eta_a}\widehat{\gamma_a}\widehat{\psi_a} \right\}.
\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

Using the results in Proposition 1, we can also estimate the proportions of the three complier types, $\omega_t = P(C_i = t)$. Since

$$\omega_n = P(C_i = n) = P(C_i = n \mid Z_i = 1, D_i = 0)P(Z_i = 1, D_i = 0) +$$

$$P(C_i = n \mid Z_i = 0, D_i = 0)P(Z_i = 0, D_i = 0),$$

$$\widehat{\omega_n} = \frac{N_{10} + M_{10}}{N} + \widehat{\psi}_n \frac{N_{00} + M_{00}}{N}.$$

Hence,

$$\widehat{\omega_n} = 2 \cdot \frac{N_{10} + M_{10}}{N}.$$

Similarly, we can show that

$$\widehat{\omega_a} = 2 \cdot \frac{N_{01} + M_{01}}{N}$$

where

$$N = \text{ total number of subjects} = N_{00} + M_{00} + N_{10} + M_{10} + N_{01} + M_{01} + N_{11} + M_{11}.$$

Under the assumption of monotonicity, $\omega_c = 1 - \omega_n - \omega_a$ and the estimator of ω_c is

$$\widehat{\omega_c} = 1 - \widehat{\omega_n} - \widehat{\omega_a}.$$

Since we can write $CACE = \eta_{1c} - \eta_{0c}$, we obtain the estimator for the local complier average causal effect CACE as follows:

$$\widehat{CACE} = \widehat{\eta_{1c}} - \widehat{\eta_{0c}}.$$

We apply the bootstrap method to compute the standard errors for the moment estimators.

9. Maximum Likelihood Estimation

One potential problem with the moment estimators in Section 8 is that they are not constrained to the range of the parameters they estimate. In this section, we also develop maximum-likelihood estimators for estimating causal parameters.

We first derive the observed-data likelihood function for the subjects with observed outcomes $(R_i = 1, Y_i, Z_i, D_i)$, and denote it $L_o(\theta)$. We then derive the observed-data likelihood function for the subjects with missing outcomes $(R_i = 0, Z_i, D_i)$, and denote it $L_m(\theta)$. The overall observed-data likelihood function, $L(\theta)$, is then equal to the product of the two likelihood functions, $L_o(\theta) \cdot L_m(\theta)$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

For subjects with observed outcomes $R_i = 1$, we have

$$\begin{aligned} P(Y_i, R_i = 1, Z_i, D_i) &= \sum_{C_i=n,a,c} P(Y_i, R_i = 1, Z_i, D_i, C_i) \\ &= \sum_{C_i=n,a,c} P(Y_i|Z_i, C_i) \cdot P(R_i = 1|Z_i, C_i) \cdot P(D_i|Z_i, C_i) \cdot P(Z_i, C_i) \\ &= \sum_{C_i=n,a,c} P(Y_i|Z_i, C_i) \cdot P(R_i = 1|Z_i, C_i) \cdot P(D_i|Z_i, C_i) \cdot P(C_i) \cdot P(Z_i). \end{aligned}$$

Since $P(Z_i = 1) = \xi$ and $P(Z_i = 0) = 1 - \xi$ and

$$P(D_i|Z_i, C_i) = 1 \text{ or } 0,$$

we obtain that

$$\begin{aligned} L_o(\theta) = [\eta_n\gamma_n\omega_n + \eta_{0c}\gamma_{0c}(1 - \omega_n - \omega_a)]^{r_{00}} \cdot [(1 - \eta_n)\gamma_n\omega_n + (1 - \eta_{0c})\gamma_{0c}(1 - \omega_n - \omega_a)]^{N_{00}-r_{00}} \\ \cdot [\eta_n\gamma_n\omega_n]^{r_{10}} \cdot [(1 - \eta_n)\gamma_n\omega_n]^{N_{10}-r_{10}} \cdot [\eta_a\gamma_a\omega_a]^{r_{01}} \cdot [(1 - \eta_a)\gamma_a\omega_a]^{N_{01}-r_{01}} \\ \cdot [\eta_a\gamma_a\omega_a + \eta_{1c}\gamma_{1c}(1 - \omega_n - \omega_a)]^{r_{11}} \cdot [(1 - \eta_a)\gamma_a\omega_a + (1 - \eta_{1c})\gamma_{1c}(1 - \omega_n - \omega_a)]^{N_{11}-r_{11}} \\ \cdot(\xi)^{N_{10}+N_{11}} \cdot (1 - \xi)^{N_{00}+N_{01}}. \end{aligned}$$

For subjects with missing outcomes $R_i = 0$, we have

$$\begin{aligned} P(R_i = 0, Z_i, D_i) &= \sum_{C_i=n,a,c} P(R_i = 0|Z_i, C_i) \cdot P(D_i|Z_i, C_i) \cdot P(Z_i, C_i) \\ &= \sum_{C_i=n,a,c} P(R_i = 0|Z_i, C_i) \cdot P(D_i|Z_i, C_i) \cdot P(C_i) \cdot P(Z_i). \end{aligned}$$

Hence we obtain that

$$\begin{aligned} L_m(\theta) = [(1 - \gamma_n)\omega_n + (1 - \gamma_{0c})(1 - \omega_n - \omega_a)]^{M_{00}} \cdot [(1 - \gamma_n)\omega_n]^{M_{10}} \\ \cdot [(1 - \gamma_a)\omega_a]^{M_{01}} \cdot [(1 - \gamma_a)\omega_a + (1 - \gamma_{1c})(1 - \omega_n - \omega_a)]^{M_{11}} \\ \cdot(\xi)^{M_{10}+M_{11}} \cdot (1 - \xi)^{M_{00}+M_{01}}. \end{aligned}$$

Therefore, the observed-data log-likelihood $l(\theta)$ is

$$\begin{aligned} & l(\theta) \equiv logL(\theta) \qquad (1) \\ = \; & r_{00} \cdot log[\eta_n\gamma_n\omega_n + \eta_{0c}\gamma_{0c}(1 - \omega_n - \omega_a)] \\ & +(N_{00} - r_{00}) \cdot log[(1 - \eta_n)\gamma_n\omega_n + (1 - \eta_{0c})\gamma_{0c}(1 - \omega_n - \omega_a)] \\ & +r_{10} \cdot log(\eta_n\gamma_n\omega_n) + (N_{10} - r_{10}) \cdot log[(1 - \eta_n)\gamma_n\omega_n] + r_{01} \cdot log(\eta_a\gamma_a\omega_a) \end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

$$
\begin{aligned}
&+(N_{01}-r_{01})\cdot log[(1-\eta_a)\gamma_a\omega_a]+r_{11}\cdot log[\eta_a\gamma_a\omega_a+\eta_{1c}\gamma_{1c}(1-\omega_n-\omega_a)]\\
&+(N_{11}-r_{11})\cdot log[(1-\eta_a)\gamma_a\omega_a+(1-\eta_{1c})\gamma_{1c}(1-\omega_n-\omega_a)]+M_{00}\cdot log[(1-\gamma_n)\omega_n\\
&+(1-\gamma_{0c})(1-\omega_n-\omega_a)]+M_{10}\cdot log[(1-\gamma_n)\omega_n]+M_{01}\cdot log[(1-\gamma_a)\omega_a]\\
&+M_{11}\cdot log[(1-\gamma_a)\omega_a+(1-\gamma_{1c})(1-\omega_n-\omega_a)]\\
&+(N_{10}+M_{10}+N_{11}+M_{11})\cdot log(\xi)+(N_{00}+M_{00}+N_{01}+M_{01})\cdot log(1-\xi). \qquad (2)
\end{aligned}
$$

It is also easy to show that

$$
\begin{aligned}
\psi_n &= P(C_i=n|Z_i=0,D_i=0)\\
&= \frac{P(C_i=n,Z_i=0,D_i=0)}{P(Z_i=0,D_i=0)}\\
&= \frac{P(C_i=n,Z_i=0,D_i=0)}{P(Z_i=0,D_i=0,C_i=n)+P(Z_i=0,D_i=0,C_i=c)}\\
&= \frac{P(D_i=0|C_i=n,Z_i=0)\cdot P(C_i=n,Z_i=0)}{P(D_i=0|C_i=n,Z_i=0)\cdot P(C_i=n,Z_i=0)+P(D_i=0|C_i=c,Z_i=0)\cdot P(C_i=c,Z_i=0)}.
\end{aligned}
$$

Since

$$P(D_i=0|C_i=n,Z_i=0)=P(D_i=0|C_i=c,Z_i=0)=1,$$

we obtain

$$
\begin{aligned}
\psi_n &= \frac{P(C_i=n,Z_i=0)}{P(C_i=n,Z_i=0)+P(C_i=c,Z_i=0)}\\
&= \frac{P(C_i=n)\cdot P(Zi=0)}{P(C_i=n)\cdot P(Z_i=0)+P(C_i=c)\cdot P(Z_i=0)}\\
&= \frac{\omega_n}{\omega_n+\omega_c}=\frac{\omega_n}{1-\omega_a}.
\end{aligned}
$$

Similarly, we can show that

$$\psi_a=\frac{\omega_a}{1-\omega_n}.$$

To obtain the maximum likelihood estimator $\hat{\theta}$ for θ, we can directly solve the score function of θ, which is the partial derivative of the log likelihood with respect to θ, $S(\theta)=\partial l(\theta)/\partial\theta$. Using the Fisher information matrix, we obtain standard error estimates for these parameters.

Since the observed-data likelihood given in Equation 2 has a complicated form involving a mixture structure over a large amount of missing-data, directly maximizing the observed-data likelihood may present a computational challenge. To avoid directly maximizing the observed-data likelihood, we propose an EM algorithm to find the ML estimator $\hat{\theta}$ of θ by treating C_i as missing-data because the complete-data likelihood has a simple form if C_i were known for all patients.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

In the E-step of the EM algorithm, we compute the conditional expectation of the complete-data likelihood function given the previous parameter estimate, denoted by $\theta^{(k)}$, and the observed data. In the M-step, we maximize this expectation with respect to θ, typically by differentiation, to obtain the new parameter estimate $\theta^{(k+1)}$. We repeat the E- and M-steps until the process converges at step K, where $\left|\theta^{(K+1)} - \theta^{(K)}\right| < \epsilon$. Here ϵ is a very small constant. The one major advantage of using the EM algorithm here is that the EM algorithm yields an explicit solution for $\theta^{(k+1)}$ in terms of $\theta^{(k)}$ and the observed data, which simplifies computation.

The complete-data likelihood function is given by

$$L_c(\theta) = \prod_{i=1}^{N} P[R, Y, C, Z, D] = \prod_{i=1}^{N} \underbrace{P[R|C, Z, D]}_{(a)} \underbrace{P[Y|C, Z, D]}_{(b)} \underbrace{P[D|C, Z]}_{c} \underbrace{P[C]}_{(d)} \underbrace{P[Z]}_{(e)}.$$

Using this expression, in Appendix B we derive the E- and M- steps in the EM algorithm. Before we state our EM algorithm results, we need additional notation. Let us define

$$x_{yrzt} = \text{ number of subjects with } Y = y, R = r, Z = t \text{ and } C = t$$

where y, r and z take values of 0 or 1 and $t = a, n$ or c.

Proposition 2 *Let $\theta^{(k)}$ be the estimate for θ after the kth iteration in the EM algorithm. Then the next iteration estimate $\theta^{(k+1)}$ for θ in the EM algorithm is given as follows:*

E-step: In the E step, we take the expectation of the complete-data log-likelihood, given the observed data and the previous parameter estimate $\theta = \theta^{(k)}$. Our observed data, denoted by obsn, consist of $(R_i = 1, Y_i, D_i, Z_i)$ and $(R_i = 0, D_i, Z_i)$. Let $x_{yrzt}^{(k+1)} = E(x_{yrzt} \mid obsn, \theta = \theta^{(k)})$, which are given in Appendix B. Then,

$$\begin{aligned}
E[l_c(\theta)|obsn, \theta = \theta^{(k)})] &= x_{110n}^{(k+1)} \cdot log(\eta_n \gamma_n \omega_n) + x_{110c}^{(k+1)} \cdot log(\eta_{0c} \gamma_{0c}(1 - \omega_n - \omega_n)) \\
&+ x_{010n}^{(k+1)} \cdot log[(1 - \eta_n)\gamma_n \omega_n] + x_{010c}^{(k+1)} \cdot log[(1 - \eta_{0c})\gamma_{0c}(1 - \omega_n - \omega_a)] \\
&+ x_{111n}^{(k+1)} \cdot log(\eta_n \gamma_n \omega_n) + x_{011n}^{(k+1)} \log((1 - \eta_n)\gamma_n \omega_n) + x_{110a}^{(k+1)} \cdot log(\eta_a \gamma_a \omega_a) \\
&+ x_{010a}^{(k+1)} \cdot log[(1 - \eta_a)\gamma_a \omega_a] + x_{111a}^{(k+1)} \cdot log(\eta_a \gamma_a \omega_a) + x_{111c} \cdot log[\eta_{1c} \gamma_{1c}(1 - \omega_n - \omega_a)] \\
&+ x_{011a}^{(k+1)} \cdot log[(1 - \eta_a)\gamma_a \omega_a] + x_{011c}^{(k+1)} \cdot log[(1 - \eta_{1c})\gamma_{1c}(1 - \omega_n - \omega_a)] \\
&+ x_{100n}^{(k+1)} \cdot log[\eta_n(1 - \gamma_n)\omega_n] + x_{100c}^{(k+1)} \cdot log[\eta_{0c}(1 - \gamma_{0c})(1 - \omega_n - \omega_a)] \\
&+ x_{000n}^{(k+1)} \cdot log[(1 - \eta_n)(1 - \gamma_n)\omega_n] + x_{000c}^{(k+1)} \cdot log[(1 - \eta_{0c})(1 - \gamma_{0c})(1 - \omega_n - \omega_a)]
\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

$$+x_{101n}^{(k+1)} \cdot log[\eta_n(1-\gamma_n)\omega_n] + x_{001n}^{(k+1)} \cdot log[(1-\eta_n)(1-\gamma_n)\omega_n] + x_{100a}^{(k+1)} \cdot log[\eta_a(1-\gamma_a)\omega_a]$$

$$+x_{000a}^{(k+1)} \cdot log[(1-\eta_a)(1-\gamma_a)\omega_a] + x_{101a}^{(k+1)} \cdot log[\eta_a(1-\gamma_a)\omega_a]$$

$$+x_{101c}^{(k+1)} \cdot log[\eta_{1c}(1-\gamma_{1c})(1-\omega_n-\omega_a)] + x_{001a}^{(k+1)} \cdot log[(1-\eta_a)(1-\gamma_a)\omega_a]$$

$$+x_{001c}^{(k+1)} \cdot log[(1-\eta_{1c})(1-\gamma_{1c})(1-\omega_n-\omega_a)].$$

M-step: we find the next iteration estimate $\theta^{(k+1)}$ *by maximizing* $E[l_c(\theta)|obsn, \theta = \theta^{(k)})]$, *and* $\theta^{(k+1)}$ *has the following explicit expressions.*

$$\begin{aligned}
\omega_n^{(k+1)} &= \frac{b_1^{(k)}(a_2^{(k)} - b_2^{(k)})}{a_1^{(k)}a_2^{(k)} - b_1^{(k)}b_2^{(k)}} \\
\omega_a^{(k+1)} &= \frac{b_2^{(k)}(a_1^{(k)} - b_1^{(k)})}{a_1^{(k)}a_2^{(k)} - b_1^{(k)}b_2^{(k)}} \\
\psi_n^{(k+1)} &= \frac{b_1^{(k)}}{a_1^{(k)}} \\
\psi_a^{(k+1)} &= \frac{b_2^{(k)}}{a_a^{(k)}} \\
\gamma_n^{(k+1)} &= \frac{x_{110n}^{(k)} + x_{010n}^{(k)} + x_{111n}^{(k)} + x_{011n}^{(k)}}{x_{110n}^{(k)} + x_{010n}^{(k)} + x_{111n}^{(k)} + x_{011n}^{(k)} + x_{100n}^{(k)} + x_{000n}^{(k)} + x_{101n}^{(k)} + x_{001n}^{(k)}} \\
\gamma_a^{(k+1)} &= \frac{x_{110a}^{(k)} + x_{010a}^{(k)} + x_{111a}^{(k)} + x_{011a}^{(k)}}{x_{110a}^{(k)} + x_{010a}^{(k)} + x_{111a}^{(k)} + x_{011a}^{(k)} + x_{100a}^{(k)} + x_{000a}^{(k)} + x_{101a}^{(k)} + x_{001a}^{(k)}} \\
\gamma_{0c}^{(k+1)} &= \frac{x_{110c}^{(k)} + x_{010c}^{(k)}}{x_{110c}^{(k)} + x_{010c}^{(k)} + x_{100c}^{(k)} + x_{000c}^{(k)}} \\
\gamma_{1c}^{(k+1)} &= \frac{x_{111c}^{(k)} + x_{011c}^{(k)}}{x_{111c}^{(k)} + x_{011c}^{(k)} + x_{101c}^{(k)} + x_{001c}^{(k)}} \\
\eta_n^{(k+1)} &= \frac{x_{110n}^{(k)} + x_{111n}^{(k)} + x_{100n}^{(k)} + x_{101n}^{(k)}}{x_{110n}^{(k)} + x_{010n}^{(k)} + x_{111n}^{(k)} + x_{011n}^{(k)} + x_{100n}^{(k)} + x_{000n}^{(k)} + x_{101n}^{(k)} + x_{001n}^{(k)}} \\
\eta_a^{(k+1)} &= \frac{x_{110a}^{(k)} + x_{111a}^{(k)} + x_{100a}^{(k)} + x_{101a}^{(k)}}{x_{110a}^{(k)} + x_{010a}^{(k)} + x_{111a}^{(k)} + x_{011a}^{(k)} + x_{100a}^{(k)} + x_{000a}^{(k)} + x_{101a}^{(k)} + x_{001a}^{(k)}} \\
\eta_{0c}^{(k+1)} &= \frac{x_{110c}^{(k)} + x_{100c}^{(k)}}{x_{110c}^{(k)} + x_{010c}^{(k)} + x_{100c}^{(k)} + x_{000c}^{(k)}} \\
\eta_{1c}^{(k+1)} &= \frac{x_{111c}^{(k)} + x_{101c}^{(k)}}{x_{111c}^{(k)} + x_{011c}^{(k)} + x_{101c}^{(k)} + x_{001c}^{(k)}},
\end{aligned}$$

where

$$a_1^{(k)} = N_{00} + M_{00} + N_{10} + M_{10} + x_{111c}^{(k)} + x_{011c}^{(k)} + x_{101c}^{(k)} + x_{001c}^{(k)},$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$$b_1^{(k)} = N_{10} + M_{10} + x_{110n}^{(k)} + x_{010n}^{(k)} + x_{100n}^{(k)} + x_{000n}^{(k)},$$

$$a_2^{(k)} = N_{11} + M_{11} + N_{01} + M_{01} + x_{110c}^{(k)} + x_{010c}^{(k)} + x_{100c}^{(k)} + x_{000c}^{(k)},$$

and

$$b_2^{(k)} = N_{01} + M_{01} + x_{111a}^{(k)} + x_{011a}^{(k)} + x_{101a}^{(k)} + x_{001a}^{(k)}.$$

For a proof, see Appendix B.

10. Simulations

In this section, we conducted simulation studies to assess the finite sample performance of proposed moment and ML estimators. Simulations are run with the total number of subjects of $N = 500$; 500 replications of the data were generated with for each fixed N. Subjects were randomly assigned to a treatment or placebo group with $\xi = P(Z = 1) = 0.5$. The proportions of never-takers, always-takers, and compliers are $\omega_n = 0.2$, $\omega_a = 0.3$ and $\omega_c = 0.5$ respectively. We set the true values of ψ_n and ψ_a as $\psi_n = 0.286$ and $\psi_a = 0.375$. The response rates for never-takers and always-takers are $\gamma_n = 0.5$ and $\gamma_a = 0.6$. The response rates for compliers in placebo and treatment groups are $\gamma_{0c} = 0.7$ and $\gamma_{1c} = 0.8$. The probabilities of $Y_i = 1$ for never-takers and always takers are $\eta_n = 0.2$ and $\eta_a = 0.3$, respectively; they are $\eta_{0c} = 0.4$ and $\eta_{1c} = 0.5$ for compliers in placebo and treatment groups. Data was generated under the assumptions of latent ignorability and compound exclusion criteria. Parameters were estimated for each of the 500 simulated data sets and the mean. We report bias and mean squared error of the parameter estimates in Table 2 below.

Our simulation results show that both the moment and ML estimators are almost unbiased and have small mean squared errors. The ML estimators have slightly smaller MSE than the Moment estimators.

11. Results

In this section, we illustrate the application of the proposed methods in the flu shot data set discussed in Section 2.

We summarize both the moment and ML estimates in Table 1 along with their associated standard errors.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

Table 2: Simulation results: N=500 with 500 replication of the data

True Parameter value	Moment estimators			ML estimators		
	Mean($\tilde{\theta}$)	Bias($\tilde{\theta}$)	MSE($\tilde{\theta}$)	Mean($\hat{\theta}$)	Bias($\hat{\theta}$)	MSE($\hat{\theta}$)
ξ=0.5	0.4998	-0.0002	0.0005	0.4998	-0.0002	0.0005
ω_n=0.2	0.2006	0.0006	0.0007	0.2006	0.0006	0.0006
ω_a=0.3	0.2989	-0.0011	0.0009	0.2988	-0.0012	0.0008
ω_c=0.5	0.5005	0.0005	0.0014	0.5006	0.0006	0.0014
ψ_n=0.286	0.2876	0.0016	0.0023	0.2865	0.0005	0.0015
ψ_a=0.375	0.3757	0.0007	0.0023	0.3741	-0.0009	0.0014
γ_n=0.5	0.4996	-0.0004	0.0054	0.4996	-0.0004	0.0054
γ_a=0.6	0.5965	-0.0035	0.0033	0.5966	-0.0034	0.0033
γ_{0c}=0.7	0.6992	-0.0008	0.0037	0.6989	-0.0011	0.0037
γ_{1c}=0.8	0.8017	0.0017	0.0042	0.8004	0.0004	0.0038
η_n=0.2	0.2015	0.0015	0.0062	0.2020	0.0020	0.0062
η_a=0.3	0.3067	0.0067	0.0049	0.3066	0.0066	0.0049
η_{0c}=0.4	0.4042	0.0042	0.0043	0.4035	0.0035	0.0042
η_{1c}=0.5	0.4949	-0.0051	0.0048	0.4944	-0.0056	0.0047

The EM algorithm finds a maximum for the log-likelihood of -5057.885 at ($\omega_n = 0.783, \omega_a = .134, \gamma_n = .523, \gamma_a = .926, \gamma_{0c} = .885, \gamma_{1c} = 1.000, \eta_n = .086, \eta_a = .101, \eta_{0c} = .038, \eta_{1c} = .031$). Convergence is rapid and stable with respect to various different starting points, indicating that this solution is a global maximum; the log-likelihood has no local maxima. These maximum-likelihood estimates also solve the score equations, as required. Unfortunately, there is a boundary value problem with respect to γ_{1c}. The likelihood method yielded an estimate of 1.0 ($\widehat{\gamma}_{1c} = 1.000$), and the moment method yielded an estimate of greater than 1.0 ($\widetilde{\gamma}_{1c} = 1.073$).

The moment and ML methods yielded identical estimates for η_n, η_a, and γ_n. The agreement is very close for complier distributions ω_n, ω_a, and ω_c. The conditional complier distributions, ψ_n

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Table 3: Results on the flu shot data set

Parameter	$\widetilde{\theta}$	SE (bootstrap)	$\widehat{\theta}$	SE (MLE)
ξ	0.507	0.009	0.507	0.010
ω_n	0.797	0.020	0.783	0.011
ω_a	0.134	0.009	0.134	0.009
ω_c	0.069	0.020	0.083	0.015
ψ_n	0.936	0.041	0.904	0.016
ψ_a	0.618	0.053	0.615	0.059
γ_n	0.523	0.016	0.523	0.015
γ_a	0.903	0.025	0.926	0.020
γ_{0c}	1.070	2.627	0.885	0.218
γ_{1c}	1.073	0.045	1.000	0.046
η_n	0.086	0.012	0.086	0.012
η_a	0.101	0.023	0.101	0.023
η_{0c}	0.026	0.269	0.038	0.097
η_{1c}	0.034	0.053	0.031	0.053

and ψ_n, are very close too. The parameters for missing data, γ_{0c} and γ_{1c}, are not very close due to the boundary solution.

From the Fisher information matrix, we can see that the asymptotic correlation between estimators for the parameters of primary interest, η_{0c} and η_{1c}, is zero. Hence the estimate for the local complier average causal effect $CACE$ is $0.031 - 0.038 = -0.009$ with a standard error of 0.112 using the ML method. Using the moment method, we obtain the estimate of $CACE$ as $0.034 - 0.026 = 0.009$ with a standard error of 0.479; the corresponding 95% confidence intervals for $CACE$ using the ML and moment methods are $(-0.211, 0.229)$, and $(-0.948, 0.930)$, respectively. Although the moment and ML estimates for $CACE$ are different, both the moment and ML methods give the same conclusion that influenza vaccination is not associated with reduced risk of hospitalization for respiratory illness.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

There are several limitations to the results of this application; some of which are due to a limitation in the data collection. First, our data set involved randomization of an encouragement to physicians, and physicians had multiple patients. Due to lacking of information on the clustering of patients by physicians, for the purpose of illustrating the proposed method, we ignored a possible clustering effect.

Second, although the exclusion restriction for never-takers is reasonable, the exclusion restriction for always–takers may be questionable. Never-takers are these patients who would not receive the flu shot in any case. If these patients and their physicians did not regard the risk of flu as high enough to warrant flu shot, it might be reasonable to assume that these patients were completely unaffected by their physicians' receipt of the letter, implying that the exclusion restriction would be satisfied for the never–takers. However, for the always-takers, the situation may be quite different. The always–takers are patients who would receive the flu shot irrespective of the receipt of the encouragement letter by their physician. Such patients are most likely at higher risk for getting the flu. If these patients and their physicians regard the risk of getting flu as high enough to warrant flu shot, they might be also subject to other medical actions. Therefore, it is possible that the flu reminder may prompt the physician to take other measures beyond giving the flu shot. If these other measures affect health outcomes, the exclusion restriction would be violated. Given these limitations we advise caution with the clinical interpretation of the results.

12. Conclusion

In this paper we have proposed both moment and ML methods for estimating the causal marginal outcome distributions, as well as the local complier average causal treatment effects when both the problems of missing-data and noncompliance exist. Our methods are extensions of the method proposed by Frangakis and Rubin [3]. The four crucial assumptions in our methods are SUTVA, monotonicity assumption, compound exclusion assumption and, the latent ignorability. We illustrated the proposed methods in the randomized encouragement design flu shot study.

One future research topic would be to extend the proposed moment method to a longitudinal setting considered by both Yau and Little [13] and Barnard et al. [1]. Using such the moment method, we can then perform a goodness-of-fit test for the parametric models used in these papers.

ACKNOWLEDGEMENT

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

This work was supported in part by NIH/NHLBI grant R01HL62567. We like to thank two reviewers for their helpful comments.

References

[1] J. Barnard, C. E. Frangakis, J. L. Hill, and D. B. Rubin. Principal stratification approach to broken randomized experiments: A case study of school choice voucher in new york city. *JASA*, 98:299–311, 2003.

[2] J Cuzick, R Edward, and N Segnan. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine*, 16:1017–1029, 1997.

[3] C.E. Frangakis and D. B. Rubin. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86:365–379, 1999.

[4] K. Hirano, G. W. Imbens, D. B. Rubin, and X. H. Zhou. Causal inference in encouragement designs with covariates. *Biostatistics*, pages 1–20, 2000.

[5] P. Holland. Causal inference, path analysis, and recursive structural equations models. In *Sociological Methodology*, pages 305–327. American Sociological Association, 1988.

[6] G. W. Imbens and D. B. Rubin. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25:305–327, 1997.

[7] C. J. McDonald, S. L. Hui, and W. M. Tierney. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *M. D. Computing*, 9:304–312, 1992.

[8] J. Neyman. On the application of probability theory to agricultural experiments. essay on principles. *Annals of Agricultural Science*, pages 1–51, 1923.

[9] Y. Peng, R. Little, and T. E. Raghuanthan. An extended general location model for causal inference from data subject to noncompliance and missing valus. *The University of Michigan Department of Biostatistics Working Paper Series*, 7:1232–1244, 2003.

[10] D. B. Rubin. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

[11] D. B. Rubin. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6:34–58, 1978.

[12] D. B. Rubin. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5:472–480, 1990.

[13] L. H. Y. Yau and R. J. Little. Inference for complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *JASA*, 96:1232–1244, 2001.

Appendix A. A proof for Proportion 1

For notational convenience, we define the following three parameters that are directly estimable from the observed data:

1. $\xi_{zd} \equiv P[Z = z, D = d]$ denotes the joint probability of observing treatment assignment $Z = z$ and treatment receipt $D = d$.

2. $\pi_{zd} \equiv P[R = 1, Z = z, D = d]$ denotes the joint probability of observing $R_i = 1$ (a subject's outcome is observed), treatment assignment $Z_i = z$, and treatment receipt $D_i = d$.

3. $\nu_{zd} \equiv P[Y = 1, R = 1, Z = z, D = d]$ denotes the joint probability of observing outcome $Y_i = 1$, treatment assignment $Z_i = z$ and treatment receipt $D_i = d$.

We first derive moment estimators for ψ_n and ψ_a. Under randomization we expect that $P[Z = 0|C = n] = P[Z = 1|C = n]$. Because never-takers by definition have $D_i = 0$, we expect that $P[Z = 0, D = 0|C = n] = P[Z = 1, D = 0|C = n]$. Hence the expected number of never-takers with $(Z_i = 0, D_i = 0)$ is the same as the number of never-takers with $(Z_i = 1, D_i = 0)$. Similarly, the expected number of always-takers with $(Z_i = 1, D_i = 1)$ is the same as the number of always-takers with $(Z_i = 0, D_i = 1)$. Therefore, we obtain that

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

and

$$\psi_a = \frac{P(C = c, Z = 1, D = 1)}{P(Z = 1, D = 1)} = \frac{P(C = c, Z = 0, D = 1)}{P(Z = 1, D = 1)} = \frac{P(Z = 0, D = 1)}{P(Z = 1, D = 1)}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Therefore we obtain the moment estimators of ψ_n and ψ_a as follows:

$$\widehat{\psi_n} = \frac{\sum_{i=1}^N Z_i(1-D_i)}{\sum_{i=1}^N (1-Z_i)(1-D_i)} = \frac{N_{10}+M_{10}}{N_{00}+M_{00}}$$

and

$$\widehat{\psi_a} = \frac{\sum_{i=1}^N (1-Z_i)D_i}{\sum_{i=1}^N Z_i D_i} = \frac{N_{01}+M_{01}}{N_{11}+M_{11}}.$$

We next derive the moment estimators for η_n, η_a, γ_n, and γ_a. By the latent ignorability, we have $P[Y=1|Z=1,C=n] = P[Y=1|Z=1,C=n,R=1]$ and $P[Y=1|Z=0,C=a] = P[Y=1|Z=0,C=a,R=1]$. Hence, we have

$$\begin{aligned}
\eta_n &= \frac{P(Y=1,Z=1,C=n,R=1)}{P(Z=1,C=n,R=1)} = \frac{P[Y=1,Z=1,D=0,R=1]}{P[Z=1,D=0,R=1]}, \\
\eta_a &= \frac{P(Y=1,Z=0,C=a,R=1)}{P(Z=0,C=n,R=1)} = \frac{P[Y=1,Z=0,D=1,R=1]}{P[Z=0,D=1,R=1]}.
\end{aligned}$$

Therefore, the moment estimators for η_n and η_a are

$$\begin{aligned}
\widehat{\eta_n} &= \frac{\sum_{i=1}^N Y_i R_i Z_i (1-D_i)}{\sum_{i=1}^N R_i Z_i (1-D_i)} = \frac{r_{10}}{N_{10}}, \\
\widehat{\eta_a} &= \frac{\sum_{i=1}^N Y_i R_i (1-Z_i) D_i}{\sum_{i=1}^N R_i (1-Z_i) D_i} = \frac{r_{01}}{N_{01}}.
\end{aligned}$$

Using the compound excludsion asumption, we have as follows:

$$\widehat{\gamma_n} = P(R=1|Z=1,C=n) = \frac{P(R=1,Z=1,D=0)}{Z=1,D=0}$$

and

$$\widehat{\gamma_a} = P(R=1|Z=0,C=a) = \frac{P(R=1,Z=0,D=1)}{Z=0,D=1}.$$

Hence the moment estimators for γ_n and γ_a are

$$\begin{aligned}
\widehat{\gamma_n} &= \frac{\sum_{i=1}^N R_i Z_i (1-D_i)}{\sum_{i=1}^N Z_i (1-D_i)} = \frac{N_{10}}{N_{10}+M_{10}}, \\
\widehat{\eta_a} &= \frac{\sum_{i=1}^N R_i (1-Z_i) D_i}{\sum_{i=1}^N (1-Z_i) D_i} = \frac{N_{01}}{N_{01}+M_{01}}.
\end{aligned}$$

Next we derive moment estimators for complier-type parameters, γ_{zc} and η_{zc}. Similar to the estimators for the ψ_n and ψ_a parameters, these estimators are not true moment estimators, but functions of other estimators and the observed data. Unlike the ψ_t 'moment' estimators, which are functions of a single estimator and the observed data, these estimators are functions of *two* other estimators and the observed data. They each contain a $\frac{1}{1-\psi_t}$ factor and also a product term, both

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

of which contribute to the variability of the estimator. We first observe that we can express the directly estimable quantity $P(R=1 \mid Z=z, D=d)$ as mixtures of the unobserved γ_{zt} and ψ_{zdt}, as well as the ratio of $P(R=1, Z=z, D=d)$ and $P(Z=z, D=d)$. That is,

$$\frac{\pi_{zd}}{\xi_{zd}} = P[R=1|Z=z, D=d] =$$
$$\sum_{t=n,a,c} P[R=1|Z=z, D=d, C=t]P[C=t|Z=z, D=d] = \sum_{t=n,a,c} \gamma_{zt}\psi_{zdt}.$$

By taking $z=0$ and 1 in the above expression, we obtain that

$$\frac{\pi_{00}}{\xi_{00}} = \gamma_n\psi_n + \gamma_{0c}(1-\psi_n),\ \frac{\pi_{11}}{\xi_{11}} = \gamma_a\psi_a + \gamma_{1c}(1-\psi_a). \tag{3}$$

Solving γ_{0c} and γ_{1c} in (3), we obtain that

$$\begin{aligned}
\gamma_{0c} &= \frac{1}{1-\psi_n}\left(\frac{\pi_{00}}{\xi_{00}} - \gamma_n\psi_n\right), \\
\gamma_{1c} &= \frac{1}{1-\widehat{\psi_a}}\left(\frac{\pi_{11}}{\xi_{11}} - \gamma_a\psi_a\right).
\end{aligned}$$

Since

$$\sum_{i=1}^{N} R_i I_{[Z_i=d, D_i=d]}/N \text{ and } \sum_{i=1}^{N} I_{[Z_i=d, D_i=d]}/N$$

are unbiased estimators of π_{zd} and ξ_{zd}, we obtain the moment estimators of γ_{0c} and γ_{1c} as follows:

$$\begin{aligned}
\widehat{\gamma_{0c}} &= \frac{1}{1-\widehat{\psi_n}}\left(\frac{\widehat{\pi_{00}}}{\widehat{\xi_{00}}} - \widehat{\gamma_n}\widehat{\psi_n}\right), \\
\widehat{\gamma_{1c}} &= \frac{1}{1-\widehat{\psi_a}}\left(\frac{\widehat{\pi_{11}}}{\widehat{\xi_{11}}} - \widehat{\gamma_a}\widehat{\psi_a}\right),
\end{aligned}$$

where

$$\widehat{\pi_{zd}} = \sum_{i=1}^{N} R_i I_{[Z_i=d, D_i=d]}/N = N_{zd}/N,$$

$$\widehat{\xi_{zd}} = \sum_{i=1}^{N} I_{[Z_i=d, D_i=d]}/N = (N_{zd}+M_{zd})/N.$$

We can show that

$$\begin{aligned}
\frac{\nu_{00}}{\pi_{00}} &= P(Y=1|Z=0, D=0, R=1) \\
&= P(Y=1|Z=0, D=0, R=1, C=n)\cdot P(C=n|Z=0, D=0, R=1) \\
&\quad + P(Y=1|Z=0, D=0, R=1, C=c)\cdot P(C=c|Z=0, D=0, R=1).
\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

We can also show that

$$
\begin{aligned}
& P(C=n|Z=0,D=0,R=1) \\
= & \frac{P(R=1|C=n,Z=0)P(C=n|Z=0,D=0)}{P(R=1,C=n,Z=0)P(C=n|Z=0,D=0)+P(R=1|C=c,Z=0)P(C=c|Z=0,D=0)} \\
= & \frac{\gamma_n\psi_n}{\gamma_n\psi_n+\gamma_{0c}(1-\psi_n)}.
\end{aligned}
$$

Similarly, we can show that

$$P(C=c|Z=0,D=0,R=1)=\frac{\gamma_{0c}(1-\psi_n)}{\gamma_n\psi_n+\gamma_{0c}(1-\psi_n)},$$

Therefore

$$\frac{\nu_{00}}{\pi_{00}}=\eta_n\cdot\frac{\gamma_n\psi_n}{\gamma_n\psi_n+\gamma_{0c}(1-\psi_n)}+\eta_{0c}\cdot\frac{\gamma_{0c}(1-\psi_n)}{\gamma_n\psi_n+\gamma_{0c}(1-\psi_n)}.$$

Similar to the expression for $\frac{\nu_{00}}{\pi_{00}}$, we obtain

$$\frac{\nu_{11}}{\pi_{11}}=\eta_a\cdot\frac{\gamma_a\psi_a}{\gamma_a\psi_a+\gamma_{1c}(1-\psi_a)}+\eta_{1c}\cdot\frac{\gamma_{1c}(1-\psi_a)}{\gamma_a\psi_a+\gamma_{1c}(1-\psi_a)}.$$

By solving for η_{0c} and η_{1c} in the above equations, we obtain that

$$\eta_{0c}=\frac{1}{\gamma_{0c}(1-\psi_n)}\left(\frac{\nu_{00}}{\pi_{00}}\cdot[\gamma_n\psi_n+\gamma_{0c}(1-\psi_n)]-\eta_n\gamma_n\psi_n\right)$$

and

$$\eta_{1c}=\frac{1}{\gamma_{1c}(1-\psi_a)}\left(\frac{\nu_{11}}{\pi_{11}}\cdot[\gamma_a\psi_a+\gamma_{1c}(1-\psi_a)]-\eta_a\gamma_a\psi_a\right).$$

Since $\sum_{i=1}^N Y_iR_iI_{[Z_i=d,D_i=d]}/N=r_{zd}/N$ is an unbiased estimator of ν_{zd}, we obtain the moment estimators of η_{0c} and η_{1c} as

$$
\begin{aligned}
\widehat{\eta_{0c}} &= \frac{1}{\widehat{\gamma_{0c}}(1-\widehat{\psi_n})}\left(\frac{r_{00}}{N_{00}}\cdot[\widehat{\gamma_n}\widehat{\psi_n}+\widehat{\gamma_{0c}}(1-\widehat{\psi_n})]-\widehat{\eta_n}\widehat{\gamma_n}\widehat{\psi_n}\right) \\
\widehat{\eta_{1c}} &= \frac{1}{\widehat{\gamma_{1c}}(1-\widehat{\psi_a})}\left(\frac{r_{11}}{N_{11}}\cdot[\widehat{\gamma_a}\widehat{\psi_a}+\widehat{\gamma_{1c}}(1-\widehat{\psi_a})]-\widehat{\eta_a}\widehat{\gamma_a}\widehat{\psi_a}\right).
\end{aligned}
$$

Appendix B. A Proof for Proposition 2

The complete-data likelihood function is given by

$$
\begin{aligned}
L_c(\theta) &= \prod_{i=1}^N P[Y,R,Z,D,C] \\
&= \prod_{i=1}^N P(Y|Z,C)P(R|Z,C)P(D|Z,C)P(Z)P(C)
\end{aligned}
$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

where $P(Z) = \xi$ can be dropped during the EM steps.

Hence, by define

$$x_{yrzt} = \text{ number of subjects with} Y = y, R = r, Z = t \text{ and } C = t,$$

we obtain the complete-data log-likelihood as

$$\begin{aligned}
l_c(\theta) \;=\; & x_{110n} \cdot log(\eta_n \gamma_n \omega_n) + x_{110c} \cdot log(\eta_{0c}\gamma_{0c}(1-\omega_n-\omega_n)) \\
& + x_{010n} \cdot log[(1-\eta_n)\gamma_n\omega_n] + x_{010c} \cdot log[(1-\eta_{0c})\gamma_{0c}(1-\omega_n-\omega_a)] \\
& + x_{111n} \cdot log(\eta_n\gamma_n\omega_n) + x_{011n} \log((1-\eta_n)\gamma_n\omega_n) + x_{110a} \cdot log(\eta_a\gamma_a\omega_a) \\
& + x_{010a} \cdot log[(1-\eta_a)\gamma_a\omega_a] + x_{111a} \cdot log(\eta_a\gamma_a\omega_a) + x_{111c} \cdot log[\eta_{1c}\gamma_{1c}(1-\omega_n-\omega_a)] \\
& + x_{011a} \cdot log[(1-\eta_a)\gamma_a\omega_a] + x_{011c} \cdot log[(1-\eta_{1c})\gamma_{1c}(1-\omega_n-\omega_a)] \\
& + x_{100n} \cdot log[\eta_n(1-\gamma_n)\omega_n] + x_{100c} \cdot log[\eta_{0c}(1-\gamma_{0c})(1-\omega_n-\omega_a)] \\
& + x_{000n} \cdot log[(1-\eta_n)(1-\gamma_n)\omega_n] + x_{000c} \cdot log[(1-\eta_{0c})(1-\gamma_{0c})(1-\omega_n-\omega_a)] \\
& + x_{101n} \cdot log[\eta_n(1-\gamma_n)\omega_n] + x_{001n} \cdot log[(1-\eta_n)(1-\gamma_n)\omega_n] + x_{100a} \cdot log[\eta_a(1-\gamma_a)\omega_a] \\
& + x_{000a} \cdot log[(1-\eta_a)(1-\gamma_a)\omega_a] + x_{101a} \cdot log[\eta_a(1-\gamma_a)\omega_a] \\
& + x_{101c} \cdot log[\eta_{1c}(1-\gamma_{1c})(1-\omega_n-\omega_a)] + x_{001a} \cdot log[(1-\eta_a)(1-\gamma_a)\omega_a] \\
& + x_{001c} \cdot log[(1-\eta_{1c})(1-\gamma_{1c})(1-\omega_n-\omega_a)].
\end{aligned}$$

In the E step, we take the expectation of the complete-data log-likelihood, given the observed data and the previous parameter estimate $\theta = \theta^{(k)}$ and obtain the following result:

$$\begin{aligned}
x_{110n}^{(k+1)} &= \frac{\eta_n^{(k)}\gamma_n^{(k)}\omega_n^{(k)}}{\eta_n^{(k)}\gamma_n^{(k)}\omega_n^{(k)} + \eta_{0c}^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}-\omega_a^{(k)})} \cdot r_{00} \\
x_{110c}^{(k+1)} &= \frac{\eta_{0c}^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}\omega_n^{(k)}}{\eta_n^{(k)}\gamma_n^{(k)}\omega_n^{(k)} + \eta_{0c}^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}-\omega_a^{(k)})} \cdot r_{00} \\
x_{010n}^{(k+1)} &= \frac{(1-\eta_n)^{(k)}\gamma_n^{(k)}\omega_n^{(k)}}{(1-\eta_n)^{(k)}\gamma_n^{(k)}\omega_n^{(k)} + (1-\eta_{0c})^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}-\omega_a^{(k)})} \cdot (N_{00}-r_{00}) \\
x_{010c}^{(k+1)} &= \frac{(1-\eta_{0c})^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}-\omega_a^{(k)})}{(1-\eta_n)^{(k)}\gamma_n^{(k)}\omega_n^{(k)} + (1-\eta_{0c})^{(k)}\gamma_{0c}^{(k)}(1-\omega_n^{(k)}-\omega_a^{(k)})} \cdot (N_{00}-r_{00}) \\
x_{111n}^{(k+1)} &= r_{10} \\
x_{011n}^{(k+1)} &= N_{10} - r_{10} \\
x_{110a}^{(k+1)} &= r_{01}
\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$$
\begin{aligned}
x_{010a}^{(k+1)} &= N_{01} - r_{01} \\
x_{111a}^{(k+1)} &= \frac{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)}}{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)} + \eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)}) - \omega_a^{(k)}} \cdot r_{11} \\
x_{111c}^{(k+1)} &= \frac{\eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)}) - \omega_a^{(k)}}{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)} + \eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)}) - \omega_a^{(k)}} \cdot r_{11} \\
x_{011a}^{(k+1)} &= \frac{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)}}{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)} + \eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot (N_{11} - r_{11}) \\
x_{011c}^{(k+1)} &= \frac{\eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)} - \omega_a^{(k)})}{\eta_a^{(k)}\gamma_a^{(k)}\omega_a^{(k)} + \eta_{1c}^{(k)}\gamma_{1c}^{(k)}(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot (N_{11} - r_{11}) \\
x_{100n}^{(k+1)} &= \frac{\eta_n^{(k)}(1-\gamma_n^{(k)})\omega_n^{(k)}}{(1-\gamma_n^{(k)})\omega_n^{(k)} + (1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{00} \\
x_{100c}^{(k+1)} &= \frac{\eta_{0c}^{(k)}(1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})}{(1-\gamma_n^{(k)})\omega_n^{(k)} + (1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{00} \\
x_{000n}^{(k+1)} &= \frac{(1-\eta_n^{(k)})(1-\gamma_n^{(k)})\omega_n^{(k)}}{(1-\gamma_n^{(k)})\omega_n^{(k)} + (1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{00} \\
x_{000c}^{(k+1)} &= \frac{(1-\eta_{0c}^{(k)})(1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})}{(1-\gamma_n^{(k)})\omega_n^{(k)} + (1-\gamma_{0c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{00} \\
x_{101n}^{(k+1)} &= \eta_n^{(k)} \cdot M_{10} \\
x_{001n}^{(k+1)} &= (1-\eta_n^{(k)}) \cdot M_{10} \\
x_{100a}^{(k+1)} &= \eta_a^{(k)} \cdot M_{01} \\
x_{000a}^{(k+1)} &= (1-\eta_a^{(k)}) \cdot M_{01} \\
x_{101a}^{(k+1)} &= \frac{\eta_a^{(k)}(1-\gamma_a^{(k)})\omega_a^{(k)}}{(1-\gamma_a^{(k)})\omega_a^{(k)} + (1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{11} \\
x_{101c}^{(k+1)} &= \frac{\eta_{1c}^{(k)}(1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})}{(1-\gamma_a^{(k)})\omega_a^{(k)} + (1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{11} \\
x_{001a}^{(k+1)} &= \frac{(1-\eta_a^{(k)})(1-\gamma_a^{(k)})\omega_a^{(k)}}{(1-\gamma_a^{(k)})\omega_a^{(k)} + (1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_1 \\
x_{001c}^{(k+1)} &= \frac{(1-\eta_{1c}^{(k)})(1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})}{(1-\gamma_a^{(k)})\omega_a^{(k)} + (1-\gamma_{1c}^{(k)})(1-\omega_n^{(k)} - \omega_a^{(k)})} \cdot M_{11}.
\end{aligned}
$$

In the M step, we find the next iteration estimate $\theta^{(k+1)}$ by finding the roots of the partial derivatives of $E[l_c(\theta)|Y, D, Z, \theta = \theta^{(k)})]$ with respect to θ,

$$
S_c(\theta) = \frac{\partial E[l_c(\theta)|\underline{x}, \theta = \theta^{(k)})]}{\partial \theta},
$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper245

where $Y = (Y_1, \ldots, Y_n)$, $D = (D_1, \ldots, D_n)$, and $Z = (Z_1, \ldots, Z_n)$ are observed data. We can show $S_c(\theta)$ has the following elements:

$$
\begin{aligned}
\dot{l}_c(\omega_n) &= \frac{x^{(k)}_{110n} + x^{(k)}_{010n} + x^{(k)}_{111n} + x^{(k)}_{011n} + x^{(k)}_{100n} + x^{(k)}_{000n} + x^{(k)}_{101n} + x^{(k)}_{001n}}{\omega_n} \\
&\quad - \frac{x^{(k)}_{110c} + x^{(k)}_{010c} + x^{(k)}_{111c} + x^{(k)}_{011c} + x^{(k)}_{100c} + x^{(k)}_{000c} + x^{(k)}_{101c} + x^{(k)}_{001c}}{1 - \omega_n - \omega_a} \\
\dot{l}_c(\omega_a) &= \frac{x^{(k)}_{110a} + x^{(k)}_{010a} + x^{(k)}_{111a} + x^{(k)}_{011a} + x^{(k)}_{100a} + x^{(k)}_{000a} + x^{(k)}_{101a} + x^{(k)}_{001a}}{\omega_a} \\
&\quad - \frac{x^{(k)}_{110c} + x^{(k)}_{010c} + x^{(k)}_{111c} + x^{(k)}_{011c} + x^{(k)}_{100c} + x^{(k)}_{000c} + x^{(k)}_{101c} + x^{(k)}_{001c}}{1 - \omega_n - \omega_a} \\
\dot{l}_c(\gamma_n) &= \frac{x^{(k)}_{110n} + x^{(k)}_{010n} + x^{(k)}_{111n} + x^{(k)}_{011n}}{\gamma_n} - \frac{x^{(k)}_{100n} + x^{(k)}_{000n} + x^{(k)}_{101n} + x^{(k)}_{001n}}{1 - \gamma_n} \\
\dot{l}_c(\gamma_a) &= \frac{x^{(k)}_{110a} + x^{(k)}_{010a} + x^{(k)}_{111a} + x^{(k)}_{011a}}{\gamma_a} - \frac{x^{(k)}_{100a} + x^{(k)}_{000a} + x^{(k)}_{101a} + x^{(k)}_{001a}}{1 - \gamma_a} \\
\dot{l}_c(\gamma_{0c}) &= \frac{x_{110c} + x_{010c}}{\gamma_{0c}} - \frac{x_{100c} + x_{000c}}{1 - \gamma_{0c}} \\
\dot{l}_c(\gamma_{1c}) &= \frac{x_{111c} + x_{011c}}{\gamma_{1c}} - \frac{x_{101c} + x_{001c}}{1 - \gamma_{1c}} \\
\dot{l}_c(\eta_n) &= \frac{x^{(k)}_{110n} + x^{(k)}_{111n} + x^{(k)}_{100n} + x^{(k)}_{101n}}{\eta_n} - \frac{x^{(k)}_{010n} + x^{(k)}_{011n} + x^{(k)}_{000n} + x^{(k)}_{001n}}{1 - \eta_n} \\
\dot{l}_c(\eta_a) &= \frac{x^{(k)}_{110a} + x^{(k)}_{111a} + x^{(k)}_{100a} + x^{(k)}_{101a}}{\eta_a} - \frac{x^{(k)}_{010a} + x^{(k)}_{011a} + x^{(k)}_{000a} + x^{(k)}_{001a}}{1 - \eta_a} \\
\dot{l}_c(\eta_{0c}) &= \frac{x_{110c} + x_{100c}}{\eta_{0c}} - \frac{x_{010c} + x_{000c}}{1 - \eta_{0c}} \\
\dot{l}_c(\eta_{1c}) &= \frac{x_{111c} + x_{101c}}{\eta_{1c}} - \frac{x_{011c} + x_{001c}}{1 - \eta_{1c}}.
\end{aligned}
$$

Solving above functions would give us the estimates

$$(\omega_n^{(k+1)}, \omega_a^{(k+1)}, \gamma_n^{(k+1)}, \gamma_a^{(k+1)}, \gamma_{0c}^{(k+1)}, \gamma_{1c}^{(k+1)}, \eta_n^{(k+1)}, \eta_a^{(k+1)}, \eta_{0c}^{(k+1)}, \eta_{1c}^{(k+1)}).$$

Since

$$\psi_n = \frac{\omega_n}{1 - \omega_a} \qquad \psi_a = \frac{\omega_a}{1 - \omega_n},$$

we can also obtain $(\psi_n^{(k+1)}, \psi_a^{(k+1)})$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press