Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2003/01

Analyzing a randomized trail on breast self-examination with noncompliance and missing outcomes

Fabrizia Mealli, Guido W. Imbens,
Salvatore Ferro, Annibale Biggeri

Università degli Studi
di Firenze

Biostatistics

Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes

Fabrizia Mealli[1]
Dipartimento di Statistica "G. Parenti", Università di Firenze

Guido W. Imbens[2]
Department of Economics, University of California at Berkeley

Salvatore Ferro[3]
Istituto Oncologico Romagnolo, Faenza
and

Annibale Biggeri[4]
Dipartimento di Statistica "G. Parenti", Università di Firenze

January 2002

[1]Fabrizia Mealli is grateful to CNR (grant # 9803620.74) and to MURST-MIUR funds (2000).

[2]University of California at Berkeley, Department of Economics, 661 Evans Hall #3880, Berkeley, CA 94720-3880

[3]Currently at Ospedale S. Orsola-Malpighi, Bologna.

[4]Annibale Biggeri is grateful to MURST-MIUR funds (2000).

ABSTRACT

Recently instrumental variables methods have been used to address non-compliance in randomized experiments. Complicating such analyses is often the presence of missing data. The standard model for missing data, Missing At Random (MAR), has some unattractive features in this context. In this paper we compare MAR-based estimates of the Complier Average Causal Effect (CACE) with an estimator based on an alternative, non-ignorable model for the missing data process, developed by Frangakis and Rubin (1999). We also introduce a new missing data model that, like the Frangakis-Rubin model, is specially suited for models with instrumental variables, but makes different substantive assumptions. We analyze these issues in the context of a randomized trial of breast self-examination (BSE). In the study two methods of teaching BSE, consisting of either mail information about BSE (standard treatment) or the attendance of a course (new treatment) involving theoretical and practical sessions, were compared with the aim of assessing whether teaching programs could increase BSE practice and improve examination skills. The study was affected by the two sources of bias mentioned above: only 55% of women assigned to receive the new treatment complied with their assignment and 35% of the women did not respond to the post-test questionnaire. The results suggest that the MAR assumption is less plausible here than some of the alternatives.

KEYWORDS: *Randomized Experiments, Instrumental Variables, Non-compliance, Complier Average Causal Effect, Intention–to–treat Effect, Missing At Random, Non-ignorable Missing Data.*

1 Introduction

Estimating causal effects of interventions is often the focus of empirical studies in medicine and the social sciences. Randomized experiments are the only generally accepted tools for causal inference; yet, they may suffer from a number of complications, including non-compliance and missing outcomes, that may compromise the study and require additional assumptions.

When the outcome is observed for each subject but compliance is imperfect, the bias of the "as-treated" (where subjects are compared by treatment received) or "per protocol" (where only outcomes for subjects who comply with their assignment are analyzed) analyses is well known (Robins and Greenland, 1994; Sheiner and Rubin, 1995; Barnard *et al.* 1998). In such imperfect compliance cases researchers typically focus on either the global Intention-To-Treat (ITT) effect (by comparing all units by their assignment rather than by the treatment actually received), or on the ITT effect for the subpopulation of compliers (units who always comply with their assignment), which can be identified by exploiting appropriate instrumental variables exclusion restrictions (Bloom, 1984; Sommer and Zeger, 1991; Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996; Imbens and Rubin, 1998; Baker, 1998, 2000; Little and Yau, 1998).

When the outcome is not observed for each unit, an ITT analysis based only on complete observations would result in bias unless the missing data process is missing completely at random (MCAR, Little and Rubin, 1987). The MCAR assumption has testable implications and can often be rejected in applications. Another, potentially more plausible assumption is the Missing at Random (MAR) model proposed by Rubin (1976) and discussed in Little and Rubin (1987). An alternative missing data model that is explicitly non-ignorable and more specifically designed for this instrumental variables setting, was proposed by Frangakis and Rubin (1999). Here we propose a third missing data model also specifically designed for the class of randomized experiments with imperfect compliance.[1] The critical assumption in the Frangakis-Rubin model requires that subjects who are unwilling to take the new treatment when assigned to it, have the same response behavior irrespective of whether they are assigned to the new treatment (at which point they refuse to cooperate) or to the old treatment. Our proposed model replaces this assumption with one where instead those subjects who always comply with

[1]Note that both the Frangakis-Rubin and the model proposed in the current paper differ from the class of nonignorable missing data models discussed by Scharfstein, Rotznitzky and Robins (1999). The Frangakis-Rubin model and the current model impose some additional restrictions on the MAR while relaxing others in a way that makes them estimable, whereas in the Scharfstein-Rotnitzky-Robins setup the non-ignorable model does not impose additional restrictions and hence can only be estimated if some of the parameters are fixed a priori.

their assignment (whether it is to the new or control treatment) are not affected in their response behavior by their assignment. Neither of these assumptions is directly testable. In this paper we discuss the relative merits of the proposed missing data models, in particular comparing the new model with MAR and Frangakis-Rubin, and apply them to data from a randomized trial with substantial noncompliance and item nonresponse to investigate the sensitivity of the substantive results to these assumptions.

We re-analyze data of a randomized trial of Breast-Self-Examination (BSE) conducted between January 1988 and December 1990 in Faenza; two methods of teaching BSE were compared consisting of either mailed information about BSE (standard treatment) or attendance in a course (new treatment) involving theoretical and practical sessions, the aim being to assess whether teaching programs could increase BSE practice and improve examination skills. The study (previously analyzed in Ferro *et al.*, 1996) could have been affected by the two sources of bias mentioned above: only 55% of women assigned to the new treatment complied with their assignment and 35% of the women did not respond to the questionnaire about final health status. Making different sets of assumptions on the response behavior of individuals, including MAR, the Frangakis-Rubin model, and a modification of the FR model, the goal of the present paper is to estimate the ITT effect for compliers on BSE practice, and to compare the results to assess the appropriateness of the various models for missing data. Section 2 describes the randomized trial on BSE used to illustrate the issues and introduces notation. In section 3 assumptions regarding compliance behaviour are presented together with alternative (ignorable and non ignorable) missing data mechanisms; a parametric model specification is presented in section 4 and parallel estimation results in section 5. Section 6 concludes the paper.

2 The randomized trial on breast self-examination in Faenza

Breast Self Examination remains the most controversial of commonly recommended procedures for breast cancer screening. The rationale behind extending BSE as a screening test stems from the fact that breast cancer is frequently detected by women themselves without any other symptoms. Although BSE is simple, noninvasive, and inexpensive, its effectiveness is heavily debated in spite of more than 30 years of research (Baxter, 2001; Spurgeon, 2001; Miller and Baines, 2001). Despite these controversies, many field trials have been undertaken to evaluate the effectiveness of teaching methods, particularly in developing countries. These trials usually compare a BSE class to alternative forms of health education (physician message, leaflets). Quality of execution of self-examination

and BSE practice are the two main outcomes under study. Studies differ in their sample selection and the choice of control treatment (Kalichman *et al.*, 2000; Ortega-Altamirano *et al.*, 2000; Strickland *et al.*, 1997; Mishra *et al.*, 1998; Giles *et al.*, 2001)).

We will consider one such study in which two teaching methods were compared, both of which would be feasible in practice and would be acceptable according to the cultural profile of the area. The study took place between January 1988 and December 1990 at the Oncologic Center of the Faenza Health District in Italy. A random sample of 825 women, aged 20-64 years, was drawn from the demografic files of the city of Faenza. The sample was stratified by age; women with a current breast pathology, a history of breast cancer, a mental or physical disorder, or a terminal illness were excluded from the study. 168 women declined participation; the remaining 657 women completed a self-administered pretest questionnaire aimed at evaluating their knowledge of breast pathophisiology, risk factors for breast cancer, preventive beliefs, practice of BSE and other individual characteristics. This is the population of interest for the purposes of our study. Responders were randomly assigned to either a treatment group (330) or a control group (327). All women received information about BSE in the mail in the form of a leaflet containing theoretical as well as graphical material describing how to perform BSE correctly. Each woman of the treatment group was in addition invited to the Oncologic Center to be taught BSE in an active way. The course was held by specialized medical staff and consisted of a one hour theoretical session, a group discussion and a fiteen-minute individual practical session. Only 182 of the 330 women in the treatment group complied with their assigned treatment, i.e., attended the course. One year later the learning level of the women was assessed by the same procedure used at the start of the study, i.e., by a self-administered questionnaire sent to all 657 women. Only 429 women of the total population provided information on post-treatment BSE practice by completing this questionnaire. In this study the question of interest is the effect of teaching on BSE. However, the study suffers from noncompliance: the receipt of the treatment was not random, as some women assigned to receive the instruction on BSE practice did not receive any instruction. In addition, there was a substantial amount of missing outcome data, partly due to the fact that the outcome data were collected at a later date than the covariate and assignment data.

2.1 Notation

In order to address the noncompliance and missing data problem we first introduce some notation. The study presented above is a two-arm randomized experiment that compares a new versus a standard treatment, with access to the new treatment only in the new

treatment arm, and all-or-nothing compliance. So, each individual i who participates in the study can either be assigned to the new treatment, $Z_i = 1$, or the standard one, $Z_i = 0$. For each individual, let $D_i(z)$ be an indicator for the treatment received (1 for new or active, 0 for standard or control) given assignement z and let $D_i = D_i(Z_i)$ be the actual treatment received; $D_i(0) = 0$ by definition in our study as those assigned to the standard treatment have no access to the new treatment. Also define $Y_i(z)$ and and $R_i(z)$, the potential outcomes and potential indicators for response (1 if a subject responds to the post-test questionnaire, 0 for non-response) if an individual is assigned to treatment z, and let $Y_i = Y_i(Z_i)$ be the actual outcome and $R_i = R_i(Z_i)$ the actual response indicator. Also, a vector of pre-treatment variables is observed for each subject, $\mathbf{X}_i$. Hence, the observed data are

$$\{Z_i, D_i, R_i, \mathbf{X}_i, (Y_i : R_i = 1), i = 1, ..., N\}.$$

Following Frangakis and Rubin (1999), let $U_i := D_i(1)$. If $U_i = 1$ then person i is a complier (someone who would always comply with their assignment, and thus with $D_i(z) = z$ for $z = 0, 1$), and if $U_i = 0$ the person is a never-taker (would never take the new treatment irrespective of the assigned treatment that is, $D_i(z) = 0$, for $z = 0, 1$). This compliance status U_i can be viewed as a covariate (Angrist *et al.*, 1996), which is observed only for persons with $Z_i = 1$ and is not observed if $Z_i = 0$, although, by randomization, it is guaranteed to have the same distribution in both treatment arms. This latent compliance covariate plays a crucial role. Only for compliers can we hope to learn anything about the effect of the treatment, as never-takers are never observed exposed to the new treatment. However, even for compliers inferring causal effects of the treatment is controversial (Angrist *et al.*, 1996).

2.2 Summary statistics

Table 1 presents some summary statistics[2] for the sample of 657 women classified by assignment Z_i and treatment received D_i. The outcome Y_i is a binary indicator whether BSE was practiced. Later we also use a second indicator whether the quality of BSE was above a cutoff level. The observed baseline covariates are X_{i1}, an indicator for previous BSE practice, X_{i2}, an indicator of good knowledge of breast pathophysiology and X_{i3}, age in years. As can be seen in columns (2) and (3), due to randomization, pretreatment variables are well balanced in the two subsamples defined by assigment; however randomization does not imply that the pretreatment variables are balanced in

[2]For a complete descriptive data analysis, including a full description of the variables obtained from the questionnaire see Ferro *et al.* (1996).

the subsamples defined by the actual treatment received (columns (5) and (6)); note that $D = 1$ implies $Z_i = 1$, so that the column with $(Z_i = 1, D_i = 1)$ contains all observations with $D_i = 1$). Knowledge of breast pathophysiology, for example is significantly different between women who attend the course and women who did not attend it. This imbalance suggests that simply comparing outcomes by treatment received, as one would do with a naive per-protocol or as-treated analysis, is invalid (e.g., Sheiner and Rubin, 1995).

Concerning the response behavior, we observe that within the group assigned to receive the treatment, response rates differ significantly between women who comply and do not comply with their assigned treatment (columns (4) and (5)), suggesting that the compliance covariate U_i may be related to the willingness to respond of the subjects. As Barnard *et al.* (1998) point out "... reasons for missing outcomes can be different for compliers and never takers, and also, can be affected by treatment assignment, creating even more disparity between the types of people being compared." Frangakis and Rubin (1999) have shown that in such cases the complete-case (or respondent-based) ITT estimator (with an estimate of -0.021 (0.040) in the present study) is generally biased for the ITT effect.

Table 1: Faenza BSE study - Summary statistics

	Means					
	(1)	(2)	(3)	(4)	(5)	(6)
	All	$Z_i = 0$	$Z_i = 1$	$Z_i = 1$ $D_i = 0$	$Z_i = 1$ $D_i = 1$	$D_i = 0$
N	657	327	330	148	182	475
Assignment (Z_i)	0.502	0	1	1	1	0.312
Course attendance (D_i)	0.277	0	0.551	0	1	0
Response (R_i)	0.653	0.688	0.618	0.399	0.797	0.598
BSE practice (Y_i)*	0.785	0.796	0.774	0.475	0.897	0.729
Pretreatment variables:						
BSE practice (X_{i1}) **	0.585	0.591	0.579	0.551	0.601	0.579
Knowledge of breast pathophysiology (X_{i2})	0.554	0.560	0.548	0.439	0.637	0.522
Age (X_{i3})	41.4	41.5	41.3	41.7	41.0	41.6
ITT		-0.021 (0.040)				

(*) Computed on respondents only. (**) Available for 615 women.

3 Assumptions regarding compliance and response behavior

3.1 Noncompliance

The summary statistics of the previous section suggest that an ITT analysis and other naive alternatives, such as the as-treated or per-protocol analysis, are potentially very misleading, as the treatment received is correlated with important pre-treatment variables. The as-treated analysis would compare individuals by the actual treatment received (columns (5) and (6)), while the per-protocol analysis would compare women who did receive the treatment (column (5)) with the control group (column (2)): in both cases pre-treatment covariates are not well balanced between the two groups, as the treated women practice BSE more and have a better knowledge of breast pathophysiology.

In order to address the noncompliance problem we consider the following assumption:

Assumption 1 (Exclusion Restriction for Never-takers)

$$Y_i(Z_i) \perp Z_i | \mathbf{X}_i, U_i = 0.$$

This assumption implies that $\Pr(Y_i(1)|\mathbf{X}_i, U_i = 0) = \Pr(Y_i(0)|\mathbf{X}_i, U_i = 0)$, so that for subpopulations of never-takers with the same value of the covariates, the distributions of the two potential outcomes are the same. This is a type of instrumental variables assumption because it rules out a direct effect of the assignment on the outcome for a specific subpopulation.

In the absence of non-response this exclusion restriction would allow to identify the intention-to-treat effect for the subpopulation defined by the compliance status covariate, namely the ITT effect for compliers (Angrist, Imbens and Rubin, 1996; Imbens, Rubin, 1997b) defined as

$$ITT_C = E(Y(1) - Y(0)|U = 1)$$

without any further assumption. Assumption 1 has some testable restrictions (Imbens and Rubin, 1997a; Balke and Pearl, 1997) in the form of inequalities, but in order to estimate models that relax it, it is typically useful to make additional assumptions, such as imposing some parametric form of the likelihood function or using informative prior distributions within a Bayesian approach (Hirano *et al.*, 2000); we will return to this issue later. The ITT_C effect is the only intention to treatment effect in this case that potentially addresses the causal effect of the receipt of the new treatment, because it compares outcomes under the new treatment with outcomes under the standard one. At

least in the present study, and especially under the exclusion restriction that requires that for never-takers there is no direct effect of the assignment, it seems plausible to attribute the effect of assignment for the compliers to the effect of the receipt of the treatment. This ITT effect is sometimes referred to as the Local Average Treatment Effect (Imbens and Angrist, 1994) or the Complier Average Causal Effect, CACE, as we will do in the sequel.

3.2 Non response

As far as the treatment of the missing data problem is concerned we now review the two principal models that have been proposed in the literature to address it. The first model assumes that Y is Missing At Random (MAR; Rubin, 1976): the probability of observing Y is the same for all the subjects with the same value of the observed covariates, treatment assigned and treatment received. In our case the i.i.d MAR assumption can thus be stated as follows:

Assumption 2 (Missing at Random, MAR)

$$Y_i \perp R_i | Z_i, \mathbf{X}_i, D_i.$$

This assumption implies that $\Pr(R_i|Y_i, Z_i, \mathbf{X}_i, D_i) = \Pr(R_i|Z_i, \mathbf{X}_i, D_i)$. If the parameters of the missing data mechanism are distinct from those of the data distribution, the missing data mechanism is said to be ignorable (Rubin, 1976; Little and Rubin, 1987). In terms of the response behavior of never-takers and compliers, this assumption implies that never-takers and compliers may have different response behavior in the new treatment arm, but the same behavior in the standard one. Under this assumption it is the treatment received, which is a deterministic function of Z_i and U_i, rather than the true compliance covariate U_i, that determines the response behavior. This model has no testable implications. We can estimate the model under Assumptions 1 and 2 without additional restrictions.

A special case of MAR arises when

$$R_i \perp Z_i, \mathbf{X}_i, D_i$$

and the data are said to be missing completely at random (MCAR). Limiting the analysis to the complete data would result in no bias in this case[3]. This model does have testable

[3]As the analysis is always conditional on Z, the complete data analysis would not be biased also under the weaker assumption: $R_i(0) \perp \mathbf{X}_i, D_i(0)$ and $R_i(1) \perp \mathbf{X}_i, D_i(1)$, which implies that $\Pr(R_i|Y_i, Z_i, \mathbf{X}_i, D_i) = \Pr(R_i|Z_i)$.

restrictions, as it implies that the distribution of covariates is the same in the complete data and incomplete data subsamples.

Before discussing the second model, it is useful to show what can be learned from the data about the response behavior of the two groups of individuals, namely the never-takers and the compliers, using a method-of-moment estimation reasoning. Avoiding here for simplicity the conditioning on the pre-treatment covariates $\mathbf{X}_i$, denote $\Pr(R_i = 1|Z_i = z, U_i = u)$ by π_{zu}, which are the four response probabilities for never-takers and compliers in the standard and in the new treatment arm. From the data, only π_{11} and π_{10} can be estimated directly as $\hat{\pi}_{11} = \sum R_i Z_i D_i / \sum Z_i D_i$ and $\hat{\pi}_{10} = \sum R_i Z_i (1 - D_i) / \sum Z_i (1 - D_i)$ using individuals assigned to the treatment, whereas in the control group only the mixture $\Pr(U_i = 1)\pi_{01} + (1 - \Pr(U_i = 1))\pi_{00}$ can be estimated as $\sum R_i (1 - Z_i) / \sum (1 - Z_i)$, because we cannot identify the 'true' compliance status for subjects in the control group. If one knew (or could estimate) either π_{01} or π_{00} a priori, they could be used to estimate both π_{01} and π_{00} separately. But as long as the data give information only on the mixture, one has to rely on assumptions, that are not testable without auxiliary information, in order to disentangle the mixture and get separate estimate of π_{01} and π_{00}. Some such assumptions are implicit or explicit in the following models proposed in the literature.

The second model, introduced by Frangakis and Rubin (1999), makes use of the compliance covariate:

Assumption 3 (Latent ignorability)

$$Y_i \perp R_i | Z_i, \mathbf{X}_i, U_i.$$

Under this assumption potential outcomes and potential nonresponse indicators are independent within each level of the compliance covariate with the same covariates'value. This assumption implies that:

$$\Pr(Y_i, R_i | Z_{i,} \mathbf{X}_i, U_i) = \Pr(Y_i | Z_{i,} \mathbf{X}_i, U_i) \Pr(R_i | Z_{i,} \mathbf{X}_i, U_i)$$

so that, if U_i were observed for all subjects (and the parameters of the missing data process are distinct from those of the outcome distribution) the missing data process would be ignorable. But because the true compliance covariate is missing, the missing data process is in fact non-ignorable.

On its own, this assumption is not sufficient to identify the ITT effect for compliers. To address the complications due to the fact the U_i is only partially observed, different assumptions can be exploited. Frangakis and Rubin propose the following assumption:

Assumption 4 (Response Exclusion Restriction for Never-takers)

$$R_i(Z_i) \perp Z_i | \mathbf{X}_i, U_i = 0.$$

This assumption implies that $\Pr(R_i(1)|\mathbf{X}_i, U_i = 0) = \Pr(R_i(0)|\mathbf{X}_i, U_i = 0)$, i.e., $\pi_{00} = \pi_{10}$. Assumptions 1 and 4 combined are the stochastic version of the compound exclusion restriction of Frangakis and Rubin (1999) and the combination of assumptions 1, 3 and 4 will be referred to as the FR model. As explained in Frangakis and Rubin (1999), under this set of assumptions, all the quantities on which the ITT effect for compliers depends have a sample counterpart, and the estimator that can be derived using the sample analogues is in fact the Frangakis-Rubin estimator.

Assumption 4 implies that never-takers have the same response behavior irrespective of the treatment arm they are assigned to. The missing data process is not ignorable in this case. While the exclusion restriction on the outcome variable seems plausible in many circumstances, the response exclusion restriction for never-takers appears to be more questionable. Especially when there is no comparable set-up in the control arm, i.e. no blind placebo-like treatment, never-takers who were assigned to the treatment and declined participation might in fact easily lower their subsequent response probability. Plausibly, their explicit refusal to comply with their assigned treatment might induce them to deny response at the posttest questionnaire as well.

An alternative to the Frangakis-Rubin response exclusion restriction for the never-takers is to assume that compliers do not change their response behavior with assignment. This rationale is behind the following assumption:

Assumption 5 (Response Exclusion Restriction for Compliers)

$$R_i(Z_i) \perp Z_i | \mathbf{X}_i, U_i = 1.$$

This assumption implies that compliers have the same response behavior irrespective of the treatment arm they are assigned to. As compliers are willing to follow the protocol in their treatment assignment, it seems more plausible that they would not be affected in their response behavior by their assignment either. The missing data process is again not ignorable in this case. The set of assumptions 1, 3 and 5 will be considered as an alternative to the MAR and FR models, referred to as the Modified Frangakis-Rubin (MFR) model. It can be easily shown that, similarly to the Frangakis-Rubin model, under this set of assumption all the quantities on which the ITT effect for compliers depends have a sample counterpart, and using the sample analogues a simple estimator could be derived.

The three sets of Assumptions, 1 and 2, (MAR), 1, 3 and 4, (FR) and 1, 3 and 5 (MFR) have no testable implication beyond inequality restrictions of the type discussed in Imbens and Rubin (1997a) and Balke and Pearl (1997). Unless one of these assumptions is made, the model would not have unique maximum likelihood estimates, the same problem encountered when relaxing assumption 1 in randomized experiment with non compliance, although the presence of observed pretreatment variables might help investigating violations of these various restrictions (Imbens and Rubin, 1997a).

4 Model specification

Method-of-moment estimators are useful to understand where information comes from the observed data and what assumptions help us identifying estimands of interest. In the presence of covariates, method-of-moment estimators are not easily implemented. Here we prefer to use likelihood based estimators, that have been proved to improve upon conventional IV estimators (Imbens and Rubin, 1997a and 1997b; see also Hirano *et al.*, 2000). In particular, we model the conditional distribution of U given the pre-treatment variables, the conditional distribution of potential outcomes given pre-treatment variables and compliance covariate U, and specify also the conditional distribution of potential response indicators given pre-treatment variables and compliance covariate U. As all the variables of interest are dichotomous, we assume that their distributions have the form of logistic regressions:

$$\Pr(U_i = 1|\mathbf{X}_i = \mathbf{x}; \alpha) = \frac{\exp(\alpha_0 + \alpha_1'\mathbf{x})}{1 + \exp(\alpha_0 + \alpha_1'\mathbf{x})}$$

$$\Pr(R_i = 1|\mathbf{X}_i = \mathbf{x}, Z_i = z, U_i = u; \beta) = \frac{\exp(\beta_{zu0} + \beta_{zu1}'\mathbf{x})}{1 + \exp(\beta_{zu0} + \beta_{zu1}'\mathbf{x})}$$

$$\Pr(Y_i = 1|\mathbf{X}_i = \mathbf{x}, Z_i = z, U_i = u; \gamma) = \frac{\exp(\gamma_{zu0} + \gamma_{zu1}'\mathbf{x})}{1 + \exp(\gamma_{zu0} + \gamma_{zu1}'\mathbf{x})}$$

Under assumptions 1 and latent ignorability the actual (observed) likelihood function is:

$$L(\theta|\mathbf{Z}, \mathbf{X}, \mathbf{D}, \mathbf{R}, \mathbf{Y}) =$$

$$\prod_{i:Z_i=1,D_i=1,R_i=1} \pi^U \pi_{11}^R f_{11}(Y_i) \prod_{i:Z_i=1,D_i=1,R_i=0} \pi^U (1 - \pi_{11}^R) \tag{1}$$

$$\prod_{i:Z_i=1,D_i=0,R_i=1} (1 - \pi^U) \pi_{10}^R f_{10}(Y_i) \prod_{i:Z_i=1,D_i=0,R_i=0} (1 - \pi^U)(1 - \pi_{10}^R)$$

$$\prod_{i:Z_i=0,D_i=0,R_i=1} \left(\pi^U \pi_{01}^R f_{01}(Y_i) + (1 - \pi^U) \pi_{00}^R f_{10}(Y_i)\right)$$

$$\prod_{i:Z_i=0,D_i=0,R_i=0} \left(\pi^U(1-\pi^R_{01}) + (1-\pi^U)(1-\pi^R_{00})\right)$$

where $\pi^U = \Pr(U_i = 1|\mathbf{X}_i;\alpha)$, $\pi^R_{Z_i,U_i} = \Pr(R_i = 1|\mathbf{X}_i, Z_i, U_i;\beta)$, $f_{Z_i,U_i}(Y_i) = \Pr(Y_i|\mathbf{X}_i, Z_i, U_i;\gamma)$, $\theta = \{\alpha, \beta, \gamma\}$ and $f_{10}(Y_i) = f_{00}(Y_i)$ for assumption 1. The assumptions MAR, FR and MFR can be imposed using the following restrictions respectively: $\pi^R_{00} = \pi^R_{01}$, $\pi^R_{00} = \pi^R_{10}$, $\pi^R_{01} = \pi^R_{11}$.

In the application in this paper we impose prior equality of the slope coefficients in the logit models for the outcome of compliers: $\gamma'_{011} = \gamma'_{111}$. Maximum likelihood estimates can be obtained using the EM algorithm (Dempster *et al.*, 1977) or standard maximization routines. In the application the Newton-Raphson algorithm was implemented and standard errors computed using the Delta method.

In the application, another variable can be used as an outcome: that is the quality of BSE practice measured by adding different indicators and resulting in a variable that can take on integer values between 0 and 21 (see Ferro *et al.*, 1996 for details on this variable). The estimation of causal effects for such an outcome is more problematic for the fact that the quality can be observed only on women who do practice BSE and is not only unobserved but also undefined when Y is equal zero. The solution to such a problem is often that of assuming the quality as missing or censored or assigning it a value of zero; although often done these approaches do not lead to properly defined causal estimands (see Rubin (2000) and Frangakis and Rubin (2002) for more discussion on this). In principle, a causal estimand of interest would be the effect of the treatment on the quality for those women who would practice BSE under both assignments; in a randomized experiment with non compliance, such a causal estimands would be the effect of the treatment on the quality for compliers who would practice BSE under both treatments. The estimation of this causal effect would involve additional assumptions that are discussed in the sequel. Here we consider, as suggested in other works (Ferro *et al.*, 1996; Miller and Baines, 2001), an alternative binary outcome Q that assumes value 1 if the quality indicator is greater than its overall median value (17) and 0 otherwise. We specify a conditional logit model for this secondary outcome, given pre-treatment variables and compliance covariate, and conditional on practicing BSE ($Y = 1$):

$$\Pr(Q_i = 1|\mathbf{X}_i = \mathbf{x}, Z_i = z, U_i = u, Y_i = 1) = \frac{\exp(\delta_{zu0} + \delta'_{zu1}\mathbf{x})}{1 + \exp(\delta_{zu0} + \delta'_{zu1}\mathbf{x})},$$

under the additional exclusion restriction for never-takers that:

Assumption 7

$$Q_i(Z_i) \perp Z_i|\mathbf{X}_i, U_i = 0, Y_i = 1.$$

This will allow to obtain an estimate of:

$$E(Q(1) - Q(0)|U = 1, Y = 1)$$

which, as explained above, cannot be interpreted as a causal effect as it is conditional on the value of the outcome Y (i.e. of a post-treatment variable). A set of assumptions that would allow to interpret this as a causal effect are the following assumptions 8 and 9:

Assumption 8 - Monotonicity of the outcome for compliers

$$Y_i(1) \geq Y_i(0)|U_i = 1$$

Assumption 9 (Quality independence of the primary outcome under control for compliers)

$$Q_i(1) \perp Y_i(0)|U_i = 1, Y_i(1) = 1$$

Assumption 8 basically says that those practicing BSE under control would have done so also under treatment, while assumption 9 says that the quality of BSE of those who practice BSE under treatment is the same irrespective of them practicing BSE or not under control. Assumption 8 cannot be verified directly at the individual level but we can have some indirect evidence of it from the estimate of CACE; assumption 9 cannot be tested, but if it would not hold the expected sign of the bias should be towards an underestimation of the real effect on BSE quality for compliers practicing BSE under both treatment arms, because the compliers who practice BSE under treatment but not under control should plausibly practice BSE with a lower average quality than those practicing BSE also under control.

In the application we impose $\delta_{zu1} = \mathbf{0}$ in order to reduce the number of parameters[4] in each model to 22.

5 Results

We first estimated the model using no pre-treatment variables. Table 2 shows the estimates of the effect on BSE practice and quality for compliers and the estimates of the four response probabilities under different assumptions.[5] The estimates under the MAR

[4]Given the relatively small sample size, relaxing this restriction, as well as the one on the slope coefficients in the logit for the outcome of compliers, would increase the computational burden and lead to imprecise estimates; in the final models age has been excluded.

[5]Note that some of the estimated probabilities are identical under the different models, due to the structure of the likelihood function. For example, the proportion of compliers is the same under both the FR and MFR models, but not under the MAR model.

assumption show an surprising negative effect of the course on BSE practice. Although it is conceivable that the course has little or no effect on BSE practice, it is more difficult to understand how, as a causal effect of attending the course for this population of volunteers, BSE practice goes down significantly. The alternative models show more plausible small and non-significant effect of the course on BSE practice and a positive effect on BSE quality. Assumptions 1, 3 and 4 (FR model) give figures for the response probabilities for compliers that are not very plausible: compliers have a lower response rate if assigned to treatment than if assigned control. Assumptions 1, 3 and 5 combined (last column, MFR model) give more plausible figures for the response probabilities: never-takers have lower response rate than compliers. In addition, never-takers have a lower response rate if assigned to the treatment arm than if assigned to the control group. This would agree with the hypothesis that once never-takers show that they are unwilling to follow the assignment protocol, they are less inclined to respond to the survey.

Table 2: Effects for compliers under various missing data assumptions, without pretreatment variables. Exclusion restriction for never-takers is always maintained. (standard errors in parentheses)

	Complete Data	MAR	FR	MFR
CACE				
on BSE practice	-0.030	-0.103	-0.012	-0.081
	(0.056)	(0.025)	(0.054)	(0.067)
Compliers' Effect				
on BSE quality	0.235	0.206	0.239	0.225
conditional on BSE practice				
	(0.063)	(0.077)	(0.062)	(0.067)
π_{01}^R	0.69	0.69	0.92	0.80
	(0.026)	(0.026)	(0.062)	(0.030)
π_{11}^R	0.62	0.80	0.80	0.80
	(0.030)	(0.030)	(0.030)	(0.030)
π_{00}^R	0.69	0.69	0.40	0.55
	(0.026)	(0.026)	(0.040)	(0.066)
π_{10}^R	0.40	0.40	0.40	0.40
	(0.040)	(0.040)	(0.040)	(0.040)
π^U	0.71	0.56	0.55	0.55
	(0.032)	(0.026)	(0.027)	(0.027)

In Table 3 we present results where we condition on the pretreatment variables. The estimates are slightly more precise, but pretreatment variables seem to have little effect on the estimated treatment effects: adding pretreatment variables in the analysis changes

the sign under latent ignorability (last two columns) to positive, although the CACE effect on BSE practice remains not significant.

Table 3: Effects for compliers under various missing data assumptions, using pretreatment variables. The exclusion restriction for never-takers is always maintained. (standard errors in parentheses)

	Complete Data	MAR	FR	MFR
CACE				
on BSE practice	-0.001	-0.111	0.011	0.024
	(0.053)	(0.026)	(0.049)	(0.054)
Compliers' Effect				
on BSE quality	0.239	0.226	0.244	0.227
conditional on BSE practice				
	(0.066)	(0.076)	(0.063)	(0.073)

6 Concluding remarks

In this paper we compare MAR-based estimates of the Complier Average Causal Effect (CACE) with estimators based on alternative models for the missing data process, including one developed by Frangakis and Rubin (1999) specifically for this instrumental variables context and a modification of the Frangakis-Rubin model. We illustrate these methods by re-analyzing data of a randomized trial of breast self-examination (BSE). In the study two methods of teaching BSE, consisting of either mail information about BSE (standard treatment) or the attendance of a course (new treatment) involving theoretical and practical sessions, were compared with the aim of assessing whether teaching programs could increase BSE practice and improve examination skills.

Since the three models are all untestable, we judge the plausability of the models by the results they produce. The MAR assumptions lead to significant negative effects of the course on BSE practice. Although one can easily imagine a positive or zero effect of the treatment, it is difficult to understand why a program designed to encourage BSE would have a negative effect. The two alternative models lead to more plausible, small, insignificant effects. One of the alternative models in particular also leads to a a priori plausible pattern of response rates: those that are unwilling to comply with their assignment are also less likely to respond to the survey, and they are less willing to respond if they have declined to participate in the treatment program. The model appears to be particularly appropriate in the context of randomized encouragement designs where one-sided non-compliance is an issue, and double blinding is not feasible.

References

[1] Angrist J.D., Imbens G.W. and Rubin D. (1996) Identification of Causal Effects using Instrumental Variables, *Journal of the American Statistical Association*, 91, 444-472.

[2] Baker S.G. (1998) Analysis of Survival Data From a Randomized Trial with All-or-None Compliance: Estimating the Cost-Effectiveness of a Cancer Screening Program, *Journal of the American Statistical Association*, 93, 929-934.

[3] Baker S.G. (2000) Analyzing a Randomized Cancer Prevention Trial with a Missing binary Outcome, an Auxiliary Variable, and All-or-None Compliance, *Journal of the American Statistical Association*, 95, 449, 43-50.

[4] Balke A. and Pearl J. (1997) Bounds on Treatment Effects from Studies with Imperfect Compliance, *Journal of the American Statistical Association*, 92, 439, 1171-1176.

[5] Barnard J, Du J., Hill J.L. and Rubin D.B. (1998) A Broader Template for Analyzing Broken Randomized Experiments, *Sociological Methods and Research*, 27, 2, 285-317.

[6] Baxter N. (2001) Preventive health care, 2001 update: should women be routinely taught breast self-examination to screen for breast cancer? *Canadian Medical Association Journal*, 164, 1837-1846.

[7] Bloom, H. (1984), "Accounting for No–shows in Experimental Evaluation Designs," *Evaluation Review*, 8(2) 225–246.

[8] Dempster A.P., Laird N. and Rubin D.B. (1977) Maximun Likelihood Estimation from incomplete data using the EM algorithm (with discussion), *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

[9] Ferro S., Caroli A., Nanni O., Biggeri A. and Gambi A. (1996) A Randomized Trail on Breast Self-Examination in Faenza (Northern-Italy), *Tumori*, 82, 4, 329-334.

[10] Frankgakis C.E. and Rubin D.B. (1999) 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.

[11] Frankgakis C.E. and Rubin D.B. (2002) "Principal Stratification in Causal Inference," *Biometrics*, 58, 21-29.

[12] Giles J.T., Kennedy D.T., Dunn E.C., Wallace W.L., Meadows S.L. and Cafiero A.C. (2001) Results of a community pharmacy-based breast cancer risk-assessment and education program, *Pharmacotherapy*, 21(2), 243-253.

[13] Holland P. (1986) Statistics and Causal Inference (with discussion), *Journal of the American Statistical Association*, 81, 396, 945-970.

[14] Hirano K., Imbens G.W., Rubin D.B. and Zhou X. (2000) Assessing the effect of an influenza vaccine in an encouragement design, *Biostatistics*, 1, 69-88.

[15] Imbens, G., and D. Rubin (1997a) "Estimating Outcome Distributions for Compliers in Instrumental Variable *Review of Economic Studies*, October.

[16] Imbens G. W. and Rubin D.B. (1997b) Bayesian Inference for causal effects in randomized experiments with noncompliance, *Annals of Statistics*, 25, 305-327.

[17] Imbens G. W. and Angrist J.D. (1994) Identification and estimation of local average treatment effects, *Econometrica*, 62, 467-476.

[18] Kalichman S.C., Williams E. and Nachimson D. (2000) Randomized community trial of a brest self-examination skills-building intervention for inner-city African-American women, *Journal of the American Medical Women Association*, 55(1), 47-50.

[19] Little R. and Rubin D. (1987) *Statistical analysis with missing data*, Wiley, New York.

[20] Little, R., and Yau, L., (1998), "Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No–Shows Using Rubin's Causal Model," *Psychological Methods*, Vol. 3, No. 2, 147-159.

[21] Miller A.B. and Baines C. (2001) Self examination contributes to reduction in mortality, *British Medical Journal* (letter), 322, 792.

[22] Mishra S.I., Chavez L.R., Magana J.R., Nava P., Burciaga Valdez R. and Hubbel F.A. (1998) Improving breast cancer control among Latinas: evaluation of a thery-based educational program,*Health Education Behaviour* 25(5), 653-670.

[23] Ortega-Altamirano D., Lopez-Carrillo L. and Lopez-Cervantes M. (2000) Strategies for teaching self-examination of the breast to women in reproductive age, *Salud Publica Mexicana*, 42(1), 17-25.

[24] Robins J.M. and Greenland S. (1994) Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial, *Journal of the American Statistical Association*, 89, 737-749.

[25] Rubin D.B. (1976) Inference and Missing Data, *Biometrika*, 63, 581-592.

[26] Rubin D.B. (2000) Comment on "Causal Inference without Conterfactuals" by P. Dawid, *Journal of the American Statistical Association*, 95, 407-448.

[27] Scharfstein, D., A. Rotnitzky, and J. Robins (1999) Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models, *Journal of the American Statistical Association*, 94, 1096-1120.

[28] Sheiner L.B. and Rubin D.B. (1995) Intention-to-Treat Analysis and the Goals of Clinical Trials, *Clinical Pharmacology and Therapy*, 57, 6-10.

[29] Sommer, A., and S. Zeger, (1991), "On Estimating Efficacy from Clinical Trials", *Statistics in Medicine*, Vol. 10, 45–52.

[30] Spurgeon D. (2001) Breast self examination may do more harm than good, *British Medical Journal* (letter), 323, 11.

[31] Strickland C.J., Feigl P., Upchurch C., King D.K., Pierce H.I., Grevstad P.K., Bearden J.D., Dawson M., Loewen W.C. and Meyskens F.L.Jr (1997) Improving breast self-examination compliance: a Southwest Oncology Group randomized trial of three interventions, *Preventive Medicine*, 26(3), 320-332.