Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2009)
58, *Part* 1, *pp.* 83–104

Statistical analysis of randomized experiments with non-ignorable missing binary outcomes: an application to a voting experiment

Kosuke Imai

Princeton University, USA

[Received March 2007. Final revision July 2008]

Summary. Missing data are frequently encountered in the statistical analysis of randomized experiments. I propose statistical methods that can be used to analyse randomized experiments with a non-ignorable missing binary outcome where the missing data mechanism may depend on the unobserved values of the outcome variable itself even after taking into account the information in the fully observed variables. The motivating empirical example is a German election experiment where researchers are worried that the act of voting may increase the probability of participation in the post-election survey through which the outcome variable, turnout, was measured. To address this problem, I first introduce an identification strategy for the average treatment effect under the non-ignorability assumption and compare it with the existing alternative approaches in the literature. I then derive the maximum likelihood estimator and its asymptotic distribution and discuss possible estimation methods. Furthermore, since the identification assumption proposed is not directly verifiable from the data, I show how to conduct a sensitivity analysis based on the parameterization that links the key identification assumption with the causal quantities of interest. Finally, the methodology proposed is extended to the analysis of randomized experiments with non-compliance. In addition, although the method that is introduced may not directly apply to randomized experiments with non-binary outcomes, I briefly discuss possible identification strategies in more general situations.

Keywords: Average treatment effect; Causal inference; Instrumental variables; Intention-to-treat effect; Non-compliance; Sensitivity analysis

1. Introduction and a motivating example

Missing data are frequently encountered in the statistical analysis of randomized experiments, and they raise various methodological issues that must be addressed for valid causal inference. In this paper, I propose statistical methods that can be used to analyse randomized experiments with a non-ignorable missing binary outcome where the missing data mechanism may depend on the unobserved values of the outcome variable itself even after taking into account the information in the fully observed variables. The outcome in randomized experiments is often binary. Examples include the turnout of voters in political science research, employment of workers in economic job training experiments, student graduation or dropout in education research and death or certain illness conditions in medical studies. In these studies, non-ignorable missing data may exist; even given fully observed covariates, voters who voted in an election are more willing to answer a post-election survey, and employed workers are more likely to report their employment status. Such a non-response pattern can be in part attributed to a sense of

Address for correspondence: Kosuke Imai, Department of Politics, Princeton University, Princeton, NJ 08544, USA.
E-mail: kimai@princeton.edu

© 2009 Royal Statistical Society 0035–9254/09/58083

social desirability that people have about voting and employment as well as to their tendency to participate in surveys, which they find relevant for themselves.

This paper is motivated by a randomized experiment that was conducted during the 2005 Bundestag election in Germany (see Goldstein *et al.* (2007) for more details about the experiment). The experiment used Internet-based surveys. During the week immediately before the election, a randomly selected group of participants (i.e. the control group) was asked whether they intended to vote in the forthcoming election. Another randomly selected group of participants (i.e. the treatment group) was asked whether they intended to vote, and if so whether they planned to vote in person rather than by mail. The voters in the treatment group were then asked to list the main obstacles that could affect whether they would vote in the forthcoming election (this group is different from the group that was analysed in Goldstein *et al.* (2007)). During the week after the election, the same participants were asked whether they had voted in the election.

The different questions that were posed to each group are viewed as different stimuli, defining the treatment factor. The stimulus that is given to the treatment group is designed to induce *implementation intentions* among the experimental subjects (e.g. 'since I shall be busy on the election day, I am going to vote by mail!'), whereas the stimulus that is given to the control group is designed to encourage the voters simply to form *goal intentions* (e.g. 'I am going to vote!'). In the psychological literature, there is a large amount of theoretical and empirical studies showing that implementation intentions can more effectively increase the probability of achieving one's goal by automating goal implementation through anticipatory decisions (see, for example, Gollwitzer (1999)). For example, Orbell *et al.* (1997) found that those women who had been asked to write down when and where they would perform a breast self-examination were more likely to perform it when compared with a group of women who had been only asked whether they intended to perform a breast self-examination. The German election experiment tests this implementation intentions hypothesis in the context of voting in an election.

Table 1 presents a summary of the data that are used for this paper. For both the treatment and the control groups, background characteristics, which include their gender and birth year, were also collected. The experimental subjects come from two German on-line panels, which are not necessarily a representative sample of the voting population. I focus on those subjects who have completed the pre-election survey, assuming that non-response in the pre-election survey is independent of the treatment factor (see also the appendix of Goldstein *et al.* (2007)). This assumption is reasonable given the fact that the treatment is subtle and is administered later in the pre-election survey. In fact, the data are consistent with the assumption; as expected because of the randomization of the treatment, the pretreatment covariates appear to be balanced.

Table 1. Summary of the German election data†

Group	*Pre-election data*				*Post-election data*	
	Size	*Fraction female*	*Year of birth (mean)*	*Fraction of vote intenders*	*Non-response rate*	*Reported turnout among respondents*
Treatment	548	0.546	1970.88	0.942	0.206	0.828
Control	572	0.538	1971.08	0.930	0.248	0.805

†Fraction of vote intenders refers to the sample proportion of those who said that they were planning to vote.

However, the non-response rate to the post-election survey is lower for the treatment group than for the control group. The observed difference, which has a p-value of 0.11 based on the χ^2-test of a contingency table, may have arisen because the missing data mechanism depends on the actual turnout of the participants of the experiment. There are at least two reasons why researchers worry about non-ignorable missing data in this experiment. First, those who voted in the election may have been more willing to answer the post-election survey since the act of voting itself may have increased their interest in the election and/or more generally in politics. The existing evidence suggests that those who do not vote are often not interested in politics and are less willing to participate in political surveys (e.g. Burden (2000)). Secondly, those who did not vote may be less likely to participate in the post-election survey in part because they may view abstention as a socially undesirable act. If the missing data are indeed non-ignorable, then the observed difference of reported turnout rates among respondents in Table 1 would be a biased estimate of the average treatment effect (ATE).

The rest of the paper is organized as follows. In Section 2, I first propose a method that can be used to analyse standard randomized experiments with non-ignorable missing binary outcomes. In Section 3, I then apply this method to the German election experiment data. Although the method that is introduced in this paper may not directly apply to randomized experiments with non-binary outcomes, I also discuss possible identification strategies in more general settings. In Section 4, I show how these methods can be extended to the analysis of randomized experiments with non-compliance. This is an important generalization as non-compliance is often encountered in randomized field experiments. I discuss the advantages and disadvantages of the proposed method over the existing approaches in the literature. The methods proposed are applicable to causal inference in observational studies where the ignorability of treatment assignment is assumed (Rosenbaum and Rubin, 1983a). Finally, the proposed and other related methods are available as an R package *experiment* (Imai, 2008) at the Comprehensive R Archive Network (`http://cran.r-project.org`).

My analysis begins by establishing the bounds of the ATE (Horowitz and Manski, 2000) and then considers assumptions that are sufficient for point identification. I then derive the maximum likelihood estimator of the ATE and its asymptotic distribution, and I discuss possible estimation methods. Since the identification assumption proposed is not directly verifiable from the data, I show how to conduct a sensitivity analysis based on the parameterization that links the key identification assumption with the causal quantities of interest. The sensitivity analysis proposed offers a formal assessment about how the resulting causal estimates may vary if the assumptions are violated to a specified degree. The analysis also complements the method of bounds which seeks to establish the identification region from the observed data alone without invoking any assumption.

In recent causal inference literature, several methods have been proposed to deal with missing data in randomized experiments. They include the method of bounds (Horowitz and Manski, 2000), semiparametric models (Scharfstein *et al.*, 1999) and approaches that are based on the assumptions of ignorability (Yau and Little, 2001) and latent ignorability (LI) (Frangakis and Rubin, 1999). This paper aims to contribute to this growing literature by providing alternative identification and estimation strategies. The methods proposed are closely related to those of *missingness not at random* that have been developed in the context of longitudinal data analysis (e.g. Diggle and Kenward (1994) and Hirano *et al.* (2001)). Throughout this paper, the approach proposed is compared with these alternative methods. Finally, following the previous works (see, for example, Scharfstein *et al.* (1999), Vansteelandt and Goetghebeur (2001) and Verzilli and Carpenter (2002), and references therein), the sensitivity analysis that is introduced in this paper examines the sensitivity of one's conclusion to the assumption about the missing data

mechanism. Others in the literature have focused on the assessment of sensitivity to the assumptions about unmeasured confounders in observational studies (e.g. Rosenbaum and Rubin (1983b), Lin *et al.* (1998), Robins *et al.* (1999) and Rosenbaum (2002)).

2. Methodology

In this section, I consider the statistical analysis of standard randomized experiments with non-ignorable missing outcomes. I begin by describing the potential outcomes framework of causal inference (Holland, 1986).

2.1. Framework

Let T_i be the binary treatment indicator variable, i.e. $T_i \in \{0, 1\}$. Let $Y_i(T_i)$ denote the binary *potential* outcome variable of unit i if the treatment value T_i is received, i.e. $Y_i(T_i) \in \{0, 1\}$. Thus, whereas $Y_i(1)$ represents the outcome of unit i when receiving the treatment, $Y_i(0)$ is the outcome of the same unit without the treatment. For each unit, only one of the two potential outcome variables can be observed, and this *realized* outcome variable is denoted by $Y_i = T_i\, Y_i(1) + (1 - T_i)\, Y_i(0)$. Next, let $R_i(T_i) \in \{0, 1\}$ represent the *potential* binary recording variable, which equals 1 if the realized outcome variable Y_i is observed for unit i after receiving the treatment value of T_i and is equal to 0 otherwise. Although there are two potential recording variables for each unit, only one of them is observed. The observed recording variable is denoted by $R_i = T_i\, R_i(1) + (1 - T_i)\, R_i(0)$. Finally, we use X_i to denote a vector of observed pretreatment covariates. Throughout this paper, I make the stability assumption, which states that there is neither interference between units nor different versions of the treatment (Cox, 1958; Rubin, 1990). I also assume non-zero probability of treatment assignment, i.e. $0 < \Pr(T_i = 1 | X_i = x) < 1$ for all $x \in \mathcal{X}$ where $\mathcal{X}$ is the support of X_i.

In standard randomized experiments, the treatment is randomized so that the potential outcome and potential recording variables are independent of the treatment variable. Formally, we have the following assumption.

Assumption 1 (randomization of treatment).

$$(Y_i(1), Y_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp T_i | X_i,$$

where '$\perp\!\!\!\perp$' denotes statistical independence.

Note that this conditional independence assumption is often invoked for causal inference in observational studies (Rosenbaum and Rubin, 1983a). A common quantity of interest is the ATE, which is defined as

$$\tau_{\mathrm{ATE}} \equiv E\{Y_i(1) - Y_i(0)\} = \int_{\mathcal{X}} \{E(Y_i | T_i = 1, X_i) - E(Y_i | T_i = 0, X_i)\}\, \mathrm{d}F_{X_i}, \tag{1}$$

where F_{X_i} is the distribution function of X, and the equality follows from assumption 1 and the definition of Y_i.

2.2. Identification

I first study the non-parametric identification of the ATE with non-ignorable missing outcomes. To do this, it is sufficient to consider the identification of the ATE at each point of X_i. Define $\tau_{\mathrm{ATE}}(x) = E\{Y_i(1) - Y_i(0) | X_i = x\}$ where $x \in \mathcal{X}$. Then, under assumption 1, the ATE can be rewritten as

$$\tau_{\mathrm{ATE}}(x) = \frac{p_{10}(x)\,\pi_{10}(x) + p_{11}(x)\,\pi_{11}(x)}{\pi_{10}(x) + \pi_{11}(x)} - \frac{p_{00}(x)\,\pi_{00}(x) + p_{01}(x)\,\pi_{01}(x)}{\pi_{00}(x) + \pi_{01}(x)}, \tag{2}$$

where $p_{jk}(x)$ and $\pi_{jk}(x)$ are defined as

$$p_{jk}(x) \equiv \Pr(Y_i = 1 | T_i = j, R_i = k, X_i = x), \tag{3}$$

$$\pi_{jk}(x) \equiv \Pr(T_i = j, R_i = k | X_i = x), \tag{4}$$

for $j = 0, 1$ and $k = 0, 1$ where $\Sigma_{j=0}^{1}\Sigma_{k=0}^{1}\pi_{jk}(x) = 1$ for all $x \in \mathcal{X}$. Although $p_{j1}(x)$ and $\pi_{jk}(x)$ for $j, k = 0, 1$ are identifiable, it is clear that the observed data do not impose any restriction on $p_{j0}(x)$. Hence, by setting $p_{00}(x)$ and $p_{10}(x)$ to their extreme values (0 and 1), the following sharp (i.e. best possible) bounds for $\tau_{\mathrm{ATE}}(x)$ can be obtained (see Horowitz and Manski (2000)),

$$\tau_{\mathrm{ATE}}(x) \in \left[\frac{p_{11}(x)\,\pi_{11}(x)\{\pi_{00}(x) + \pi_{01}(x)\} - \{\pi_{00}(x) + p_{01}(x)\,\pi_{01}(x)\}\{\pi_{10}(x) + \pi_{11}(x)\}}{\{\pi_{10}(x) + \pi_{11}(x)\}\{\pi_{00}(x) + \pi_{01}(x)\}}, \right.$$
$$\left. \frac{\{\pi_{10}(x) + p_{11}(x)\pi_{11}(x)\}\{\pi_{00}(x) + \pi_{01}(x)\} - p_{01}(x)\pi_{01}(x)\{\pi_{10}(x) + \pi_{11}(x)\}}{\{\pi_{10}(x) + \pi_{11}(x)\}\{\pi_{00}(x) + \pi_{01}(x)\}} \right],$$

where the width of the bounds is equal to

$$\frac{\pi_{10}(x)\{\pi_{00}(x) + \pi_{01}(x)\} + \pi_{00}(x)\{\pi_{10}(x) + \pi_{11}(x)\}}{\{\pi_{10}(x) + \pi_{11}(x)\}\{\pi_{00}(x) + \pi_{01}(x)\}},$$

which is not equal to 0 unless $\pi_{10}(x) = \pi_{00}(x) = 0$. Aggregating these bounds over the distribution of X, we obtain the sharp bounds of the ATE. Without an additional assumption, therefore, the ATE is not point identified.

A common way to point-identify the ATE is to assume that the outcome variables are missing at random and the missing data mechanism is ignorable (Little and Rubin, 1987). In the context of standard randomized experiments, using the potential outcomes notation, the missingness at random (MAR) assumption can be written as follows.

Assumption 2 (identification assumption for ignorable missing outcomes).

$$\Pr(R_i = 1 | T_i = j, Y_i = 1, X_i = x) = \Pr(R_i = 1 | T_i = j, Y_i = 0, X_i = x), \qquad \text{for } j = 0, 1, \text{ and } \forall x \in \mathcal{X}.$$

Since assumption 2 implies that $p_{10}(x) = p_{11}(x)$ and $p_{00}(x) = p_{01}(x)$, the ATE can be identified at each point of X as $\tau_{\mathrm{ATE}}(x) = p_{11}(x) - p_{01}(x)$.

Although the MAR assumption may be reasonable in many situations, researchers often worry that the missing data mechanism may further depend on the values of the outcome variable itself and hence is not ignorable. To deal with such non-ignorable missing data, I propose the following identification assumption.

Assumption 3 (identification assumption for non-ignorable missing outcomes).

$$\Pr(R_i = 1 | T_i = 1, Y_i = k, X_i = x) = \Pr(R_i = 1 | T_i = 0, Y_i = k, X_i = x), \qquad \text{for } k = 0, 1, \text{ and } \forall x \in \mathcal{X}.$$

The assumption implies that, within each of the strata that are defined by the realized (but not necessarily observed) outcome as well as covariates, the missing data mechanism does not depend on the treatment variable. Such an assumption may be reasonable if the treatment represents a relatively weak intervention in terms of future non-response. Related ideas have been discussed in the context of longitudinal data analysis (see, for example, Diggle and Kenward (1994), Scharfstein *et al.* (1999) and Hirano *et al.* (2001)). Assumption 3 does not eliminate the need to control for confounders. Indeed, even if the treatment does not directly affect the missing data mechanism the assumption can be violated. This happens if the treatment affects

the outcome and there are unmeasured factors that are associated with the outcome as well as the non-response behaviour.

In the context of the German election experiment, assumption 3 (non-ignorability (NI)) implies that those who voted in the election have the same probability of reporting their voting behaviour in the post-election survey, regardless of their treatment status after conditioning on the observed pretreatment covariates (and the same condition holds for those who did not vote). In contrast, the identification strategy that is based on assumption 2 (MAR) implies that the non-response probability is independent of the actual turnout given the treatment status and the observed pretreatment covariates. As discussed in Section 3, the NI assumption may be reasonable because the act of voting itself may increase one's interests in the election and hence make them more likely to participate in the post-election survey. It is important to note that, in the pre-election survey, voters in the treatment group were asked only two additional questions compared with those in the control group. If the burden of answering the pre-election survey would have been significantly higher for the treatment group, the treatment may directly affect the response probability even after conditioning on the voting outcome and the pretreatment covariates.

Under assumption 3, it is straightforward to show that the ATE is just identified.

Proposition 1 (identification of the ATE with non-ignorable missing outcomes). Under assumptions 1 and 3,

$$\tau_{\mathrm{ATE}}(x) = \frac{\{a(x)-b(x)\}\{\pi_{00}(x)\,\pi_{11}(x)-\pi_{01}(x)\,\pi_{10}(x)-a(x)+b(x)\}}{\pi_{11}(x)\,\pi_{01}(x)\{p_{01}(x)-p_{11}(x)\}\{\pi_{00}(x)+\pi_{01}(x)\}\{\pi_{10}(x)+\pi_{11}(x)\}},$$

where $a(x) = p_{11}(x)\,\pi_{11}(x)\{\pi_{00}(x)+\pi_{01}(x)\}$ and $b(x) = p_{01}(x)\,\pi_{01}(x)\{\pi_{10}(x)+\pi_{11}(x)\}$. In addition, given assumption 1, assumption 3 is a minimal assumption which allows $\tau_{\mathrm{ATE}}(x)$ to be identified.

A proof is given in appendix A.1. Equations (16) and (17) in the proof also imply observable implications of assumption 3. In particular, if the value of either $p_{00}(x)$ or $p_{10}(x)$ lies outside the unit interval, then assumption 3 is violated (the converse, however, is not true).

2.3. Generalization to multivalued treatment and outcome variables

Now, I briefly consider the generalization of the above identification strategy for the non-ignorable missing outcomes beyond the setting of the binary outcome and treatment variables. Suppose that the treatment is a J-valued variable and the outcome is a K-valued variable, i.e. $T_i \in \{0, 1, \ldots, J-1\}$ and $Y_i(j) \in \{0, 1, \ldots, K-1\}$ where $J \geqslant 2$ and $K \geqslant 2$. Then, the ATE is defined at each of the treatment levels (and at each value of $X_i = x$ where $x \in \mathcal{X}$), $\tau^{(j)}_{\mathrm{ATE}}(x) = E\{Y_i(j) - Y_i(j-1)|X_i = x\}$ for $j = 1, \ldots, J-1$. Once $\tau^{(j)}_{\mathrm{ATE}}(x)$ has been identified for each $j = 1, \ldots, J-1$, then other non-contiguous ATEs are also identified, e.g. $E\{Y_i(3) - Y_i(1)|X_i = x\} = \tau^{(3)}_{\mathrm{ATE}}(x) + \tau^{(2)}_{\mathrm{ATE}}(x)$.

An argument that is similar to that made for the case of the binary outcome variable implies that there are $J \times (K-1)$ unknown probabilities, i.e. $\Pr(Y_i = k|T_i = j, R_i = 0, X_i = x)$ for each (j, k), and that the identification of these probabilities is required to identify the ATE. Similarly to assumption 3, I assume that

$$\Pr(R_i = 1|T_i = j, Y_i = k, X_i = x) = \Pr(R_i = 1|T_i = j', Y_i = k, X_i = x),$$

where $j \neq j'$. Since there are $J(J-1)K/2$ possible such constraints, one can choose $J(K-1)$ constraints of them to identify all the ATEs, i.e. $\tau^{(j)}_{\mathrm{ATE}}(x)$ for all $j = 1, \ldots, J-1$, so long as $J \geqslant 3 - 2/K$. (The resulting estimator depends on the selection of these constraints since different sets of

constraints imply different assumptions about the missing data mechanism.) Thus, $J=2$ and $K \geqslant 3$ represent the only case where the identification of the ATE is not achieved without an additional restriction. This is a limitation of the proposed identification strategy under the NI assumption. In contrast, the MAR assumption always provides $J(K-1)$ linearly independent constraints. To point-identify the ATE in such situations under the NI assumption, therefore, a general strategy is to make a parametric assumption about the conditional independence between $R_i(T_i)$ and T_i given $Y_i(T_i)$, e.g.

$$\Pr\{R_i(j)=1|T_i=j, Y_i(j)=y, X_i=x\}=\frac{\exp(\alpha+\beta y+\gamma x)}{1+\exp(\alpha+\beta y+\gamma x)}$$

where α, β and γ are unknown parameters.

2.4. Inference

I now discuss the estimation of the ATE under assumption 3. First, we consider the estimation of the ATE within each stratum defined by the pretreatment covariate. The inference can be based on the following (observed data) likelihood function:

$$\prod_{i\in\{i:X_i=x\}} [p_{11}(x)^{Y_i}\{1-p_{11}(x)\}^{1-Y_i}\pi_{11}(x)]^{R_iT_i}[p_{01}(x)^{Y_i}\{1-p_{01}(x)\}^{1-Y_i}\pi_{01}(x)]^{R_i(1-T_i)} \times \{\pi_{10}(x)^{T_i}\pi_{00}(x)^{1-T_i}\}^{1-R_i}. \quad (5)$$

Given this likelihood function, the maximum likelihood (ML) estimator of the ATE and its asymptotic properties can be derived.

Proposition 2 (asymptotic distribution of the ML estimator under assumption 3).

$$n^{1/2}\{\hat{\tau}_{\mathrm{ATE}}(x)-\tau_{\mathrm{ATE}}(x)\} \xrightarrow{\mathrm{d}} N\{0, \delta(x)^{\mathrm{T}}\Sigma(x)\,\delta(x)\},$$

where $\hat{\tau}_{\mathrm{ATE}}(x)$ is the ML estimator of $\tau_{\mathrm{ATE}}(x)$, and $\Sigma(x)$ and $\delta(x)$ are given in Appendix A.2.

A proof is given in Appendix A.2. Given this ML estimate of $\tau_{\mathrm{ATE}}(x)$, we can consistently estimate the ATE by estimating the distribution of the pretreatment covariates. Alternatively, we may be interested in estimating the quantity which averages $\tau_{\mathrm{ATE}}(x)$ over the empirical distribution of X, which is sometimes called the conditional average treatment effect (CATE) (Imbens, 2004),

$$\tau_{\mathrm{CATE}} \equiv \frac{1}{n}\sum_{i=1}^{n}\tau_{\mathrm{ATE}}(X_i). \quad (6)$$

This avoids the modelling of the covariate distribution, which can be high dimensional, and results in smaller uncertainty of estimation.

In practice, the number of observations within strata defined by the values of the pretreatment covariates is so small that the asymptotic approximation of proposition 2 within each stratum may be quite poor. In such situations, we may use regression-based methods with the observed pretreatment covariates by modelling $p_{j1}(x)$ and $\pi_{jk}(x)$ for $j,k=0,1$. Then, the inference will still be based on the likelihood function of equation (5).

However, there is an alternative modelling strategy. In particular, applied researchers may prefer to model the outcome variables directly, i.e. $q_j(x) \equiv \Pr(Y_i=1|T_i=j, X_i=x)$, and the response indicator, i.e. $r_{jk}(x) \equiv \Pr(R_i=1|T_i=j, Y_i=k, X_i=x)$, for $j=0,1$ and $k=0,1$, since they are more likely to have substantive knowledge about these probabilities, than $\pi_{jk}(x)$ and $p_{j1}(x)$. Under assumption 3, $r_{\cdot k}(x)=r_{1k}(x)=r_{0k}(x)$ for $k=0,1$. Once $q_j(x)$ has been estimated,

the CATE can be estimated as

$$\tau_{\mathrm{CATE}} = \frac{1}{n}\sum_{i=1}^{n}\{q_1(X_i) - q_0(X_i)\}.$$

Although $r_{jk}(x)$ is a nuisance parameter in this set-up, it is non-parametrically identified under assumption 3. For example, it can be shown that $r_{\cdot 1}(x) = p_{11}(x)\,\pi_{11}(x)/\{p_{10}(x)\,\pi_{10}(x) + p_{11}(x)\,\pi_{11}(x)\}$, implying that $r_{\cdot 1}(x)$ is identified because $p_{10}(x)$ can be also written as a function of identifiable parameters (see equation (17)). Similarly, one can show that $r_{\cdot 0}(x)$ is non-parametrically identified.

Finally, to obtain the ML estimates of model parameters, we maximize the observed data likelihood function. However, in the empirical example that is given in Section 3 and other simulation settings, standard numerical optimization algorithms are found to be somewhat unstable. Thus, I use the expectation and maximization (EM) algorithm to obtain the ML estimates (Dempster *et al.*, 1977) (see Appendix B and also Stubbendick and Ibrahim (2003)) by integrating out the missing data in the following complete-data likelihood function:

$$\prod_{i=1}^{n}[r_{\cdot 1}(X_i)^{R_i}\{1-r_{\cdot 1}(X_i)\}^{1-R_i}]^{Y_i}[r_{\cdot 0}(X_i)^{R_i}\{1-r_{\cdot 0}(X_i)\}^{1-R_i}]^{1-Y_i}[q_1(X_i)^{Y_i}\{1-q_1(X_i)\}^{1-Y_i}]^{T_i} \times [q_0(X_i)^{Y_i}\{1-q_0(X_i)\}^{1-Y_i}]^{1-T_i}. \tag{7}$$

The asymptotic variance can be calculated in the usual way (see, for example, Louis (1982)). It is also possible to conduct a Bayesian analysis based on the likelihood function of equation (7). Once prior distributions have been specified, a Markov chain Monte Carlo algorithm can be constructed to sample model parameters from their posterior distribution.

2.5. Sensitivity analysis

Next, I propose a sensitivity analysis for the NI assumption. In particular, we assess how the deviation from the assumption alters the resulting conclusions. First, I consider a sensitivity analysis based on the ratio of the two conditional probabilities,

$$\theta_k^{\mathrm{NI}}(x) \equiv \frac{\Pr(R_i = 1 | T_i = 1, Y_i = k, X_i = x)}{\Pr(R_i = 1 | T_i = 0, Y_i = k, X_i = x)}, \tag{8}$$

for $k = 0, 1$ and all $x \in \mathcal{X}$. In the German election experiment, $\theta_1^{\mathrm{NI}}(x)$ and $\theta_0^{\mathrm{NI}}(x)$ represent the ratio of response probabilities in the strata that are defined by $X_i = x$ between the treated and control individuals who voted or did not vote in the election respectively. Assumption 3 holds when $\theta_k^{\mathrm{NI}}(x) = 1$ for $k = 0, 1$ and all x. The ATE can be written as a function of the sensitivity parameters and identifiable quantities,

$$\tau^{\mathrm{ATE}}(x) = \frac{[a(x)\,p_{01}(x)\{1-p_{11}(x)\} - \theta_0^{\mathrm{NI}}(x)\,b(x)\,p_{11}(x)\{1-p_{01}(x)\}]\{a(x) - \theta_1^{\mathrm{NI}}(x)\,b(x)\}}{a(x)\,b(x)[\theta_1^{\mathrm{NI}}(x)\{1-p_{11}(x)\}\,p_{01}(x) - \theta_0^{\mathrm{NI}}(x)\{1-p_{01}(x)\}\,p_{11}(x)]},$$

where $a(x)$ and $b(x)$ are defined in proposition 1. Given this expression, sensitivity analysis can be conducted by computing the ATE based on various values of $\theta_0^{\mathrm{NI}}(x)$ and $\theta_1^{\mathrm{NI}}(x)$ within their range, i.e.

$$\frac{\{1-p_{11}(x)\}\,\pi_{11}(x)}{\{1-p_{11}(x)\}\,\pi_{11}(x) + \pi_{10}(x)} \leqslant \theta_0^{\mathrm{NI}} \leqslant \frac{\{1-p_{01}(x)\}\,\pi_{01}(x) + \pi_{00}(x)}{\{1-p_{01}(x)\}\,\pi_{01}(x)},$$

$$\frac{p_{11}(x)\,\pi_{11}(x)}{p_{11}(x)\,\pi_{11}(x) + \pi_{10}(x)} \leqslant \theta_1^{\mathrm{NI}}(x) \leqslant \frac{p_{01}(x)\,\pi_{01}(x) + \pi_{00}(x)}{p_{01}(x)\,\pi_{01}(x)}.$$

Similarly, for the MAR assumption, we can base the sensitivity analysis on the following ratio of these conditional probabilities:

$$\theta_j^{\mathrm{MAR}}(x) \equiv \frac{\Pr(R_i = 1 | T_i = j, Y_i = 1, X_i = x)}{\Pr(R_i = 1 | T_i = j, Y_i = 0, X_i = x)}, \tag{9}$$

for $j = 0, 1$. Clearly, if $\theta_j^{\mathrm{MAR}}(x) = 1$ for $j = 0, 1$, then assumption 2 is satisfied. Using the identifiable parameters, the possible range of values for $\theta_j^{\mathrm{MAR}}(x)$ is given by

$$\frac{p_{j1}(x)\,\pi_{j1}(x)}{\pi_{j0}(x) + p_{j1}(x)\,\pi_{j1}(x)} \leqslant \theta_j^{\mathrm{MAR}}(x) \leqslant \frac{\{1 - p_{j1}(x)\}\pi_{j1}(x) + \pi_{j0}(x)}{\{1 - p_{j1}(x)\}\,\pi_{j1}(x)}, \qquad \text{for } j = 0, 1.$$

A straightforward calculation shows that the ATE can be estimated once the value of θ_j^{MAR} has been specified within this range for $j = 0, 1$:

$$\tau_{\mathrm{ATE}}(x) = \frac{p_{11}(x)}{\theta_1^{\mathrm{MAR}}(x)\{1 - p_{11}(x)\} + p_{11}(x)} - \frac{p_{01}(x)}{\theta_0^{\mathrm{MAR}}(x)\{1 - p_{01}(x)\} + p_{01}(x)}.$$

Thus, sensitivity analysis can be conducted by changing the values of $\theta_j^{\mathrm{MAR}}(x)$ and observing how the ATE changes with $\theta_j^{\mathrm{MAR}}(x)$.

Although this sensitivity analysis is formulated in terms of the risk ratio, we can also use the risk difference. One disadvantage of these approaches (the risk ratio and risk difference), however, is that, as we saw above, the sensitivity parameters are bounded by unknown (though identifiable) quantities. Also, researchers may wish to conduct a sensitivity analysis within their parametric analysis (see Scharfstein *et al.* (1999) and references therein for a semiparametric or non-parametric approach in more general settings). In this case, a sensitivity analysis that is based on the odds ratio may be reasonable though its main disadvantage is the difficulty of interpretation (e.g. King and Zeng (2002), pages 1411–1412). Following the literature (e.g. Rosenbaum (1987, 2002)), the sensitivity analysis for the MAR and NI assumptions can be conducted by using the following logistic regression model for the response probabilities:

$$r_{jk}(x; \eta_{jk}) = \frac{\exp(\alpha_{jk} + \beta x)}{1 + \exp(\alpha_{jk} + \beta x)},$$

where $\eta_{jk} = (\alpha_{jk}, \beta)$ is a vector of unknown parameters. An alternative sensitivity analysis is to consider all possible enumeration of missing data patterns (thereby avoiding the parametric model for the non-response mechanism) and to assess the sensitivity of parameter estimates in the outcome model (Vansteelandt and Goetghebeur (2001) and Verzilli and Carpenter (2002), and references therein).

Under this model, the sensitivity parameters proposed define the range of odds ratio for the conditional probabilities of missingness, to assess the sensitivity of the causal estimates to the violation of the assumptions: for MAR,

$$\Gamma_j^{\mathrm{MAR}} = \frac{r_{j1}(x; \eta_{j1})/\{1 - r_{j1}(x; \eta_{j1})\}}{r_{j0}(x; \eta_{j0})/\{1 - r_{j0}(x; \eta_{j0})\}} = \exp(\alpha_{j1} - \alpha_{j0}); \tag{10a}$$

for NI

$$\Gamma_k^{\mathrm{NI}} = \frac{r_{1k}(x; \eta_{1k})/\{1 - r_{1k}(x; \eta_{1k})\}}{r_{0k}(x; \eta_{0k})/\{1 - r_{0k}(x; \eta_{0k})\}} = \exp(\alpha_{1k} - \alpha_{0k}). \tag{10b}$$

Here $\Gamma_j^{\mathrm{MAR}} \geqslant 0$ and $\Gamma_k^{\mathrm{NI}} \geqslant 0$ for $j = 0, 1$ and $k = 0, 1$. Thus, sensitivity analysis may proceed by first specifying the values of Γ_j^{MAR} or Γ_k^{NI}, which are equivalent to setting $\alpha_{j1} - \alpha_{j0} = \log(\Gamma_j^{\mathrm{MAR}})$ and

$\alpha_{1k} - \alpha_{0k} = \log(\Gamma_k^{\mathrm{NI}})$ respectively. Under this parametric approach, the sensitivity parameters do not depend on the value of the covariates.

Furthermore, as shown in Appendix B, if we postulate parametric models for $q_j(x;\phi_j) \equiv \Pr(Y_i = 1 | T_i = j, X_i = x; \phi_j)$, e.g. the logistic regression models, then the EM algorithm can be used to obtain the ML estimate of the CATE. Alternatively, a Bayesian analysis can be conducted by specifying the prior distribution on model parameters, while respecting the constraints that are implied by the values of the sensitivity parameters.

Finally, a referee suggested that, although the sensitivity analyses proposed are formulated in terms of the missingness probabilities, one may wish to connect the proposed sensitivity parameters to the outcome success probabilities of those whose outcome variables are unobserved within each treatment arm. This can be done in a straightforward manner. For example, $p_{00}(x)$ can be written as a function of the sensitivity parameters $\theta_k(x)^{\mathrm{NI}}$ and identifiable quantities:

$$p_{00}(x) = \frac{c_1(x)[\pi_{01}(x)\{1 - p_{01}(x)\} + \pi_{00}(x)] - c_0(x)[\theta_1^{\mathrm{NI}}(x)\{\pi_{11}(x) + \pi_{10}(x)\} + \pi_{11}(x)\, p_{11}(x)]}{c_1(x)\,\pi_{00}(x) - c_0(x)\,\pi_{11}(x)\, p_{11}(x)/p_{01}(x)},$$

where $c_j(x) = \{1 - p_{j1}(x)\}\pi_{j1}(x)$. An alternative strategy, which was suggested by another referee, is to express the magnitude of bias under the NI assumption as a function of the sensitivity parameter, i.e. a degree to which the NI assumption is violated.

3. Empirical analysis of the German election experiment

In this section, I analyse the German election experiment that was described in Section 1 by applying the methods that were introduced above.

3.1. Estimation of the average treatment effects

I first estimate the ATE in equation (1) under the NI assumption by using the results in propositions 1 and 2 and compare them with the estimate under the MAR assumption. I then estimate the CATE in equation (6) by modelling the turnout and recording indicator variables parametrically under the NI assumption and base our inference on the likelihood function in equation (5). Specifically, I use the following parametric models: $q_j(X_i) = \Pr(Y_i = 1 | T_i = j, X_i = x) = \exp(\alpha_j + x^{\mathrm{T}}\beta)/\{1 + \exp(\alpha_j + x^{\mathrm{T}}\beta)\}$ and $r_{.k}(X_i) = \Pr(R_i = 1 | Y_i = k, X_i = x) = \exp(\gamma_k + x^{\mathrm{T}}\delta)/\{1 + \exp(\gamma_k + x^{\mathrm{T}}\delta)\}$, where Y_i is the voting indicator variable which is equal to 1 if respondent i voted, X_i includes the gender (1 if female, and 0 otherwise), the last two digits of the year of birth, voting intention (1 if intended to vote, and 0 otherwise), and indicator variables representing the region of residence of each respondent. The EM algorithm in Appendix B is used to obtain the ML estimate of the CATE. For the comparison, I also obtain the ML estimate of the CATE under the MAR assumption by modelling

$$p_j(x) = \Pr(Y_i = 1 | R_i = 1, T_i = j, X_i = x) = \exp(\eta_j + x^{\mathrm{T}}\xi)/\{1 + \exp(\eta_j + x^{\mathrm{T}}\xi)\}.$$

Table 2 presents the ML estimates of the ATE and the CATE as well as their standard errors and 95% confidence intervals. The standard errors and confidence intervals for the ATE are based on its asymptotic variance, whereas those for the CATE are based on 1000 (standard non-parametric) bootstrap replications. The results show that the estimated effect size is larger under the NI assumption than under the MAR assumption. In particular, when controlling for the observed pretreatment covariates, the point estimate is approximately 4.6 percentage points with the 95% confidence interval just overlapping zero. In contrast, under the MAR assumption, there is little evidence that the treatment increases the probability of turnout. Moreover,

Table 2. Estimated ATEs (increase in turnout) for the German election experiment†

Assumption	*Point estimate*	*Standard error*	*95% confidence interval*	
			Lower	*Upper*
MAR				
No covariate	0.021	0.026	−0.030	0.073
With covariates	0.014	0.025	−0.035	0.063
NI				
No covariate	0.035	0.051	−0.049	0.119
With covariates	0.046	0.036	−0.011	0.129

†Under the assumptions of MAR and NI, the ML estimate of the ATE and CATE are presented without and with pretreatment covariates respectively. The standard errors and confidence intervals for the ATE are based on its asymptotic variance, whereas those for the CATE are based on 1000 bootstrap replications.

the inclusion of pretreatment covariates significantly reduces the uncertainty estimates under the NI assumption, whereas this does not appear to be so under the MAR assumption.

3.2. Sensitivity analysis

Next, following the strategies that were described in Section 2.5, I conduct a sensitivity analysis under the two assumptions. Fig. 1 presents the results without pretreatment covariates. Here, I calculate the estimated ATE based on each set of the values of the sensitivity parameters, $(\theta_0^{\mathrm{MAR}}, \theta_1^{\mathrm{MAR}})$ for the MAR assumption (Fig. 1(a)) and $(\theta_0^{\mathrm{NI}}, \theta_1^{\mathrm{NI}})$ for the NI assumption (Fig. 1(c)), where I chose $[0.8, 1.2]$ as the range of the values although the bounds on these sensitivity parameters are wider (e.g. [0.76, 1.41] for θ_1^{NI} and [0.76, 2.51] for θ_1^{MAR}). For example, this means that under the NI assumption I investigate the sensitivity of the conclusion that is based on the estimates that are given in Table 2 to the key identifying assumption by letting the response probability of the treated units who actually voted be different by a maximum 20% from the response probability of the control units who voted. Figs 1(a) and 1(c) plot the estimated ATE under each of various combinations of the values of the sensitivity parameters. Similar graphs are created by using t-statistics where the standard errors are based on the standard non-parametric bootstrap. They are presented in of Figs 1(b) and 1(d).

Fig. 1 shows that under the NI assumption the point estimate of the ATE can vary from 0.15 to −0.075, suggesting that the empirical evidence supporting a positive effect of the treatment in this experiment is quite sensitive to the key identifying assumption. Although the sensitivity of the point estimate under the NI assumption appears greater than that under the MAR assumption, the degree of the variation in t-statistics is similar under the two assumptions, implying that the standard error under the NI assumption is larger.

Finally, another sensitivity analysis is conducted by using the observed pretreatment covariates. The analysis is based on the two sensitivity parameters, Γ_0^{NI} and Γ_1^{NI} that were defined in equation (10), which represent the odds ratios of the (conditional) recording probabilities between treated and control units given the actual voting behaviour and the observed pre-

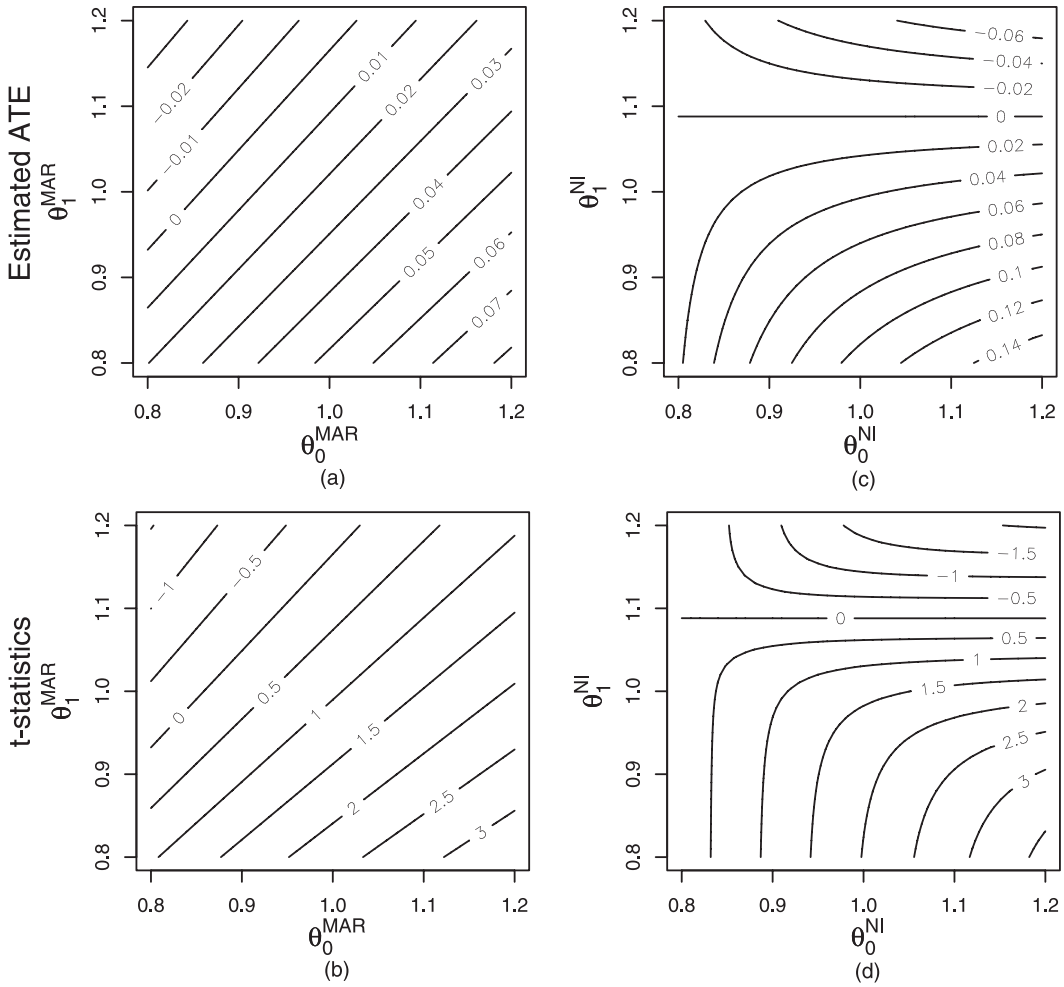

Fig. 1. Sensitivity analysis for the German election experiment without pretreatment covariates, (a), (b) under the MAR assumption and (c), (d) under the NI assumption: in each graph, the values of two sensitivity parameters (representing the horizontal and vertical axes), $(\theta_0^{\text{MAR}}, \theta_1^{\text{MAR}})$ for the MAR assumption and $(\theta_0^{\text{NI}}, \theta_1^{\text{NI}})$ for the NI assumption, range from 0.8 to 1.2, and the estimated ATE ((a) and (c)) and *t*-statistics ((b) and (d)) are calculated and plotted for each set of the values; *t*-statistics are based on the standard non-parametric bootstrap

treatment covariates. Table 3 represents the results of this analysis for a range of values of the sensitivity parameters. We observe that, when $(\Gamma_0^{\text{NI}}, \Gamma_1^{\text{NI}}) = (\frac{1}{3}, 3)$ (the upper right-hand cell), the estimates are quite different from the results under the NI assumption (reproduced in the middle cell with $(\Gamma_0^{\text{NI}}, \Gamma_1^{\text{NI}}) = (1, 1)$). However, this scenario may be unlikely if we expect the treatment to affect directly voters' response probability in the same direction regardless of whether they actually voted in the election. For the same reason, the scenario corresponding to the lower left-hand cell, i.e. $(\Gamma_0^{\text{NI}}, \Gamma_1^{\text{NI}}) = (\frac{1}{3}, 3)$, may not be plausible although it yields the estimate which is much larger than that obtained under the NI assumption. If these scenarios are ignored, Table 3 appears to suggest that the treatment either increases the turnout by 5 percentage points on average or has a negligible effect for the range of values of the sensitivity parameters that is considered here.

Table 3. Sensitivity analysis for the NI assumption with the pretreatment covariates†

	$\Gamma_1^{\mathrm{NI}} = \frac{1}{3}$	$\Gamma_1^{\mathrm{NI}} = 1$	$\Gamma_1^{\mathrm{NI}} = 3$
$\Gamma_0^{\mathrm{NI}} = \frac{1}{3}$	0.046 (0.027) [−0.006, 0.100]	0.003 (0.020) [−0.032, 0.046]	−0.075 (0.027) [−0.128, −0.024]
$\Gamma_0^{\mathrm{NI}} = 1$	0.045 (0.029) [−0.015, 0.097]	0.046 (0.036) [−0.011, 0.129]	0.004 (0.039) [−0.073, 0.080]
$\Gamma_0^{\mathrm{NI}} = 3$	0.134 (0.029) [0.080, 0.192]	0.047 (0.033) [−0.020, 0.111]	0.046 (0.028) [−0.009, 0.101]

†Γ_k^{NI} for $k = 0, 1$, which is defined in equation (10), is the sensitivity parameter which equals the odds ratio of the (conditional) recording probabilities between the treated and control units given the actual voting behaviour and the observed pretreatment covariates. The middle cell with $(\Gamma_0^{\mathrm{NI}}, \Gamma_1^{\mathrm{NI}}) = (1, 1)$ shows the results when the NI assumption holds. Within each cell, the first row represents the ML estimate of the ATE, whereas the second and third rows give the estimated standard errors and 95% confidence intervals, based on 1000 bootstrap replications.

4. Extension to randomized experiments with non-compliance

In this section, I extend the identification and estimation strategies that were described in Section 2 to the analysis of randomized experiments with non-compliance and non-ignorable binary missing outcomes. In voting experiments that are similar to the German experiment that was analysed above, non-compliance is a common problem. For example, to study the influence of a Web-based political campaign on voting behaviour, Horiuchi *et al.* (2007) used an encouragement design where randomly selected voters were encouraged to view political parties' Web sites. However, some voters who were encouraged did not view the designated Web sites. Another example is a standard 'get out the vote' experiment where researchers are interested in estimating the causal effects of different mobilization techniques on turnout (e.g. Gerber and Green (2000) and Imai (2005)). In these experiments, some voters in the treatment group cannot be contacted by canvassers whereas others in the control group may be exposed to this type of campaign. These examples suggest that the extension of the proposed methods to randomized experiments with non-compliance is of practical importance.

4.1. Framework

Let Z_i be the binary (randomized) encouragement variable, which is equal to 1 if unit i is encouraged to receive the binary treatment and is equal to 0 otherwise. The observed binary (non-randomized) treatment variable is denoted by $T_i = Z_i\, T_i(1) + (1 - Z_i)\, T_i(0)$ where $T_i(1)$ and $T_i(0)$ represent the potential binary treatment variables with and without the encouragement. Finally, the binary potential outcomes are denoted as $Y_i(Z_i)$, whereas the potential recording variables are represented by $R_i(Z_i)$. As before, I denote the realized outcome variable as $Y_i = Z_i\, Y_i(1) + (1 - Z_i)\, Y_i(0)$ and the observed recording variable as $R_i = Z_i\, R_i(1) + (1 - Z_i)\, R_i(0)$. Although I ignore the pretreatment covariates throughout this section for notational simplicity,

a generalization of the results that are presented here can be easily accomplished by following the approaches that were given in the previous section.

In this section, the encouragement variable Z_i is assumed to be randomized.

Assumption 4 (randomization of encouragement).

$$(Y_i(1), Y_i(0), T_i(1), T_i(0), R_i(1), R_i(0)) \perp\!\!\!\perp Z_i,$$

for each $i = 1, 2, \ldots, n$.

In the absence of missing data, assumption 4 alone is sufficient to identify the intention-to-treat (ITT) effect or the average causal effect of encouragement:

$$\tau_{\mathrm{ITT}} \equiv E\{Y_i(1) - Y_i(0)\} = E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0). \tag{11}$$

In addition to the ITT effect, another causal quantity of interest can be defined by following Angrist *et al.* (1996) and defining the latent compliance type as follows: unit i is a complier if $T_i(Z_i) = Z_i$ (which is denoted by $C_i = c$), a never-taker if $T_i(1) = T_i(0) = 0$ (which is denoted by $C_i = n$), an always-taker if $T_i(1) = T_i(0) = 1$ (which is denoted by $C_i = a$) and a defier if $T_i(Z_i) = 1 - Z_i$ (which is denoted by $C_i = d$), where C_i represents the compliance covariate for unit i. Angrist *et al.* (1996) showed that under the assumption of no defier (or the monotonicity assumption), i.e. $T_i(1) \geqslant T_i(0)$, and the exclusion restriction for never-takers and always-takers, i.e. $Y_i(1) = Y_i(0)$ for units with $C_i = a$ and $C_i = n$, we can identify the ATE for compliers, which Angrist *et al.* (1996) called the complier average causal effect (CACE),

$$\tau_{\mathrm{CACE}} \equiv E\{Y_i(1) - Y_i(0)|C_i = c\} = \frac{E\{Y_i(1) - Y_i(0)\}}{E\{T_i(1) - T_i(0)\}}.$$

For compliers the ATE equals the ITT effect. Since the treatment variable is assumed to be without missing values, $E\{T_i(1) - T_i(0)\}$ can be identified. Thus, the identification of τ_{ITT} implies that of τ_{CACE} under the monotonicity assumption and exclusion restriction.

4.2. Identification

To study the identification problem in this setting, a useful decomposition of the ITT effect under assumption 4, which is similar to that in equation (2), is

$$\tau_{\mathrm{ITT}} = \frac{\sum_{j=0}^{1}\sum_{k=0}^{1} p_{jk1}\pi_{jk1}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1}\sum_{k=0}^{1} p_{jk0}\pi_{jk0}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk0}}, \tag{12}$$

where p_{jkl} and π_{jkl} are defined as $p_{jkl} \equiv \Pr(Y_i = 1|T_i = j, R_i = k, Z_i = l)$ and $\pi_{jkl} \equiv \Pr(T_i = j, R_i = k, Z_i = l)$. Whereas π_{jkl} and p_{j1l} are identifiable, p_{j0l} is not. Since p_{j0l} can vary within the unit interval, the following bounds for the ITT effect can be derived:

$$\tau_{\mathrm{ITT}} \in \left[\frac{\sum_{j=0}^{1} p_{j11}\pi_{j11}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1} (p_{j10}\pi_{j10} + \pi_{j00})}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk0}}, \frac{\sum_{j=0}^{1} p_{j11}\pi_{j11}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1} p_{j10}\pi_{j10}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk0}} \right],$$

where the width of the bounds equals

$$\frac{(\pi_{101}+\pi_{001})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk0}+(\pi_{100}+\pi_{000})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk1}}{\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk1}\right)\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk0}}.$$

The bounds for the CACE can be obtained by dividing the upper and lower bounds of the ITT by $\pi_{111}+\pi_{101}-(\pi_{110}+\pi_{100})$.

A standard approach to point-identify the ITT is to assume that the data are missing at random (Yau and Little, 2001). In the context of causal inference, the assumption can be written as follows.

Assumption 5 (identification assumption for randomized experiments with non-compliance and ignorable missing data).

$$\Pr(R_i=1|Y_i=1,T_i=j,Z_i=l)=\Pr(R_i=1|Y_i=0,T_i=j,Z_i=l),$$

for $j=0,1$ and $l=0,1$, which implies that, within the strata that are defined by the observed treatment and encouragement variables, the outcome variable is missing completely at random. The assumption can be generalized by further conditioning on the observed pretreatment covariates X_i. Assumption 5 implies that $p_{j0l}=p_{j1l}$ for $j=0,1$ and $l=0,1$. Hence, the ITT effect is identified.

By extending the strategy that was developed in Section 2, I propose the following identifying assumption for the non-ignorable missing data mechanism.

Assumption 6 (identification assumption for randomized experiments with non-compliance and non-ignorable missing data).

$$\Pr(R_i=1|T_i=j,Y_i=k,Z_i=1)=\Pr(R_i=1|T_i=j,Y_i=k,Z_i=0),$$

for $j=0,1$, and $k=0,1$, which implies that the missing data mechanism does not depend on the randomized encouragement within the strata defined by the observed treatment and realized outcome variables. Assumption 6 can be generalized by further conditioning on the observed pretreatment variables X_i.

The next proposition establishes the identification of the ITT effect under assumption 6.

Proposition 3 (identification of the ITT effect). Under assumptions 4 and 6,

$$\tau_{\mathrm{ITT}}=\frac{\sum_{j=0}^{1}p_{j11}\{(p_{j10}-p_{j11})\pi_{j11}+(1-p_{j11})\pi_{j00}\pi_{j11}/\pi_{j10}-(1-p_{j10})\pi_{j01}\}}{(p_{j10}-p_{j11})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk1}}$$
$$-\frac{\sum_{j=0}^{1}p_{j10}\{(p_{j10}-p_{j11})\pi_{j10}+(1-p_{j11})\pi_{j00}-(1-p_{j10})\pi_{j10}\pi_{j01}/\pi_{j11}\}}{(p_{j10}-p_{j11})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk0}}.$$

In addition, given assumption 4, assumption 6 is a minimal assumption which allows τ_{ITT} to be identified.

A proof is given in Appendix A.3. Although assumption 6 is not directly verifiable from the observed data, it has testable implications. In particular, if the value of either p_{j00} or p_{j01}, whose expressions are given in equations (18) and (19), lie outside the unit interval, then assumption 6 is violated (the converse, however, is not true). If we further assume monotonicity and the exclusion restriction that was discussed in Section 4.1, then the CACE can be identified as

$$\tau_{\mathrm{CACE}} = \frac{\tau_{\mathrm{ITT}}\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk1}\right)\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk0}}{(\pi_{101}+\pi_{111})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk0} - (\pi_{100}+\pi_{110})\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jk1}},$$

where τ_{ITT} is given in proposition 3.

4.3. Inference

The estimation strategy of the ITT effect under assumption 6 parallels the methodology that was described in Section 2.4. When there is no pretreatment covariate, we may use sample averages to estimate π_{jkl} and p_{j1l} for each j, k and l, and then use proposition 3 to obtain the ML estimates of the ITT effect. The asymptotic distributions of the ML estimators of the ITT effect as well as the CACE can be derived by applying the delta method as done in proposition 2.

When some pretreatment covariates are available, however, we may directly model the outcome, treatment and recording variables, i.e. $p_{jl}(x) \equiv \Pr(Y_i = 1 | T_i = j, Z_i = l, X_i = x)$, $q_l(x) \equiv \Pr(T_i = 1 | Z_i = l, X_i = x)$ and $r_{jk}(x) \equiv \Pr(R_i = 1 | T_i = j, Y_i = k, X_i = x)$, for each j, k and l. Then, the ITT effect conditional on $X_i = x$ is given by $\tau_{\mathrm{ITT}}(x) = p_{11}(x)\,q_1(x) + p_{01}(x)\{1 - q_1(x)\} - [p_{10}(x)\,q_0(x) + p_{00}(x)\{1 - q_0(x)\}]$, whereas the conditional CACE is given by $\tau_{\mathrm{CACE}}(x) = \tau_{\mathrm{ITT}}(x)/\{q_1(x) - q_0(x)\}$. As before, we may summarize these quantities by averaging over the sample values of X_i.

4.4. Comparison with the latent ignorability approach

In the literature, a popular approach to the estimation of the ITT effect and CACE with non-ignorable missing data is the assumption of LI, which was proposed by Frangakis and Rubin (1999). The LI approach assumes the ignorability of the missing data mechanism only after conditioning on the latent compliance covariate (e.g. Barnard *et al.* (2003) and Mealli *et al.* (2004)). In the context of randomized experiments with binary outcomes, the assumption can be written as follows.

Assumption 7 (LI assumption).

$$\Pr\{R_i(l) = 1 | Y_i(l) = 1, C_i = t, Z_i = l\} = \Pr\{R_i(l) = 1 | Y_i(l) = 0, C_i = t, Z_i = l\}, \qquad \text{for } l = 0, 1.$$

Under this approach, in addition to assumptions 4 and 7, the exclusion restrictions with respect to both the outcome and the recording indicator variables are assumed for non-compliers, i.e. $Y_i(1) = Y_i(0)$ and $R_i(1) = R_i(0)$ for units with $C_i = n$ and $C_i = a$. Furthermore, the monotonicity assumption is assumed so that there is no defier. All together, it can be shown that these assumptions identify the ITT effect and the CACE (see Frangakis and Rubin (1999) and O'Malley and Normand (2005)).

The advantage of the NI approach proposed, i.e. assumption 6 in Section 4.2, over the LI approach that is reviewed here is that the former does not require either the monotonicity assumption or the exclusion restrictions for the identification of the ITT effect (as shown in proposition 3). In contrast, the LI approach requires such assumptions even when the quantity of interest is the ITT effect rather than the CACE. For the identification of the CACE under the NI approach proposed, however, both the monotonicity assumption and the exclusion restriction for non-compliers (with respect to the potential outcomes but not to the potential recording variables) are necessary. In contrast, the main advantage of the LI approach is that it can be easily extended to non-binary outcomes, whereas the NI approach requires additional assumptions for reasons that are similar to those discussed in Section 2.2.

4.5. Sensitivity analysis

Like in randomized experiments, assumptions 5–7 are not directly verifiable from the data. Thus, sensitivity analysis is an important tool to assess the degree to which one's conclusion is sensitive to the key identifying assumption. I first consider sensitivity analysis for the MAR assumption based on the sensitivity parameter

$$\psi_{jl}^{\mathrm{MAR}} \equiv \frac{\Pr(R_i = 1 | T_i = j, Y_i = 1, Z_i = l)}{\Pr(R_i = 1 | T_i = j, Y_i = 0, Z_i = l)},$$

for $j = 0, 1$ and $l = 0, 1$. Clearly, if $\psi_{jl}^{\mathrm{MAR}} = 1$ for $j = 0, 1$ and $l = 0, 1$, the MAR assumption is satisfied. The bounds for this parameter are given by

$$\frac{p_{j1l}\pi_{j1l}}{p_{j1l}\pi_{j1l} + \pi_{j0l}} \leqslant \psi_{jl}^{\mathrm{MAR}} \leqslant \frac{(1 - p_{j1l})\pi_{j1l} + \pi_{j0l}}{(1 - p_{j1l})\pi_{j1l}}.$$

A calculation similar to that described in Section 2.5 allows us to write the ITT effect in terms of the sensitivity parameters and identifiable quantities:

$$\tau_{\mathrm{ITT}} = \frac{\sum_{j=0}^{1} \{p_{j11}(\pi_{j11} + \pi_{j01})\} / \{\psi_{j1}^{\mathrm{MAR}}(1 - p_{j11}) + p_{j11}\}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk1}} - \frac{\sum_{j=0}^{1} \{p_{j10}(\pi_{j10} + \pi_{j00})\} / \{\psi_{j0}^{\mathrm{MAR}}(1 - p_{j10}) + p_{j10}\}}{\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk0}}.$$

Sensitivity analysis can be conducted for the NI assumption in a similar manner by using the following sensitivity parameters:

$$\psi_{jk}^{\mathrm{NI}} \equiv \frac{\Pr(R_i = 1 | T_i = j, Y_i = k, Z_i = 1)}{\Pr(R_i = 1 | T_i = j, Y_i = k, Z_i = 0)},$$

for $j = 0, 1$ and $k = 0, 1$. When $\psi_{jk} = 1$ for all $j = 0, 1$ and $k = 0, 1$, then the NI assumption is satisfied. The possible ranges for these parameters are given by

$$\frac{p_{j11}\pi_{j11}}{p_{j11}\pi_{j11} + \pi_{j01}} \leqslant \psi_{j1}^{\mathrm{NI}} \leqslant \frac{p_{j10}\pi_{j10} + \pi_{j00}}{p_{j10}\pi_{j10}}$$

and

$$\frac{(1-p_{j11})\pi_{j11}}{(1-p_{j11})\pi_{j11}+\pi_{j01}} \leqslant \psi_{j0}^{\mathrm{NI}} \leqslant \frac{(1-p_{j10})\pi_{j10}+\pi_{j00}}{(1-p_{j10})\pi_{j10}}.$$

The ITT effect can be written as a function of sensitivity parameters and identifiable quantities:

$$\tau_{\mathrm{ITT}} = \sum_{j=0}^{1}(\{(1-p_{j11})(\pi_{j10}+\pi_{j00})\pi_{j11} - \psi_{j0}^{\mathrm{NI}}(1-p_{j10})(\pi_{j11}+\pi_{j01})\pi_{j10}\}(p_{j11}\pi_{j11}\pi_{\cdot\cdot 0} - \psi_{j1}^{\mathrm{NI}}p_{j10}\pi_{j10}\pi_{\cdot\cdot 1})[\pi_{\cdot\cdot 0}\pi_{\cdot\cdot 1}\pi_{j10}\pi_{j11}\{\psi_{j1}^{\mathrm{NI}}(1-p_{j11})p_{j10} - \psi_{j0}^{\mathrm{NI}}(1-p_{j10})p_{j11}\}]^{-1}),$$

where $\pi_{\cdot\cdot l} = \Sigma_{j=0}^{1}\Sigma_{k=0}^{1}\pi_{jkl}$ for $l = 0, 1$.

Finally, the following sensitivity parameters can be used for the LI assumption:

$$\psi_{tl}^{\mathrm{LI}} = \frac{\Pr\{R_i(l)=1|C_i=t, Y_i(l)=1, Z_i=l\}}{\Pr\{R_i(l)=1|C_i=t, Y_i(l)=0, Z_i=l\}}, \tag{13}$$

where $t \in \{c, a, n\}$ and $l = 0, 1$, and the LI assumption is satisfied if $\psi_{tl}^{\mathrm{LI}} = 1$ for all $t \in \{c, n, a\}$ and $l = 0, 1$. The sensitivity parameters can be partitioned into groups that are defined by types of complier. Thus, for example, we may make the LI assumption for compliers and always-takers and assess the sensitivity to the assumption about never-takers.

To obtain the expression of the ITT effect as a function of identifiable quantities and sensitivity parameters, I start with the decomposition

$$\Pr\{Y_i(l)=1|C_i=t, Z_i=l\} = p_{t1l}r_{tl} + p_{t0l}(1-r_{tl}), \tag{14}$$

where $p_{tkl} \equiv \Pr\{Y_i(l)=1|C_i=t, R_i(l)=k, Z_i=l\}$ and $r_{tl} \equiv \Pr\{R_i(l)=1|C_i=t, Z_i=l\}$, for $t \in \{c, n, a\}$ and $l = 0, 1$. Then, the ITT effect can be written as the weighted average of the ITT effect for each type of complier:

$$\tau_{\mathrm{ITT}} = \sum_{t\in\{c,a,n\}} \Pr(C_i=t)\{p_{t11}r_{t1} + p_{t01}(1-r_{t1}) - p_{t10}r_{t0} - p_{t00}(1-r_{t0})\}. \tag{15}$$

Since the encouragement is randomized, under the monotonicity assumption the population proportion of each type of complier, i.e. $\Pr(C_i = t)$, can be identified and consistently estimated as $\Sigma_{i=1}^{n}(1-T_i)Z_i/\Sigma_{i=1}^{n}Z_i$ for never-takers and $\Sigma_{i=1}^{n}T_i(1-Z_i)/\Sigma_{i=1}^{n}(1-Z_i)$ for always-takers, whereas the population proportion of compliers is given by $\Pr(C_i = c) = 1 - \Pr(C_i = a) - \Pr(C_i = n)$. Appendix C of Imai (2007) proves that, given the values of ψ_{tl}^{LI}, the parameters p_{tkl} and r_{tl} are identified for $k = 0, 1$ and $l = 0, 1$. Therefore, the sensitivity analysis proposed can be conducted by examining how the value of τ_{ITT} changes as a function of ψ_{tl}^{LI} by using equation (15).

5. Concluding remarks and remaining methodological issues

Missing data are frequently encountered even in randomized experiments, and applied researchers often worry that the missing data mechanism may depend on their unobserved values. In this paper, I offer a set of identification, estimation and sensitivity analysis strategies that can be used to adjust for non-ignorable missing binary outcomes in randomized experiments both with and without non-compliance. The methods proposed are motivated by and applied to the German election experiment where it is hypothesized that the act of voting itself may increase the response probability of the post-election survey. The empirical analysis shows that the esti-

mated ATEs under the NI assumption are larger than under the usual MAR assumption, but the results are quite sensitive to the key identifying assumptions. Whether the same methods are applicable beyond voting experiments is an issue that needs to be considered in the future.

Although the discussion is given in the context of randomized experiments, the methods proposed are in principle applicable to causal inference in observational studies where the ignorability of treatment assignment is assumed to hold conditioning on observed pretreatment covariates. Furthermore, there are some possible extensions that could address the remaining methodological issues. First, although a non-parametric identification analysis has been conducted in this paper, the estimation with covariates is based on parametric modelling assumptions. The development and application of robust non-parametric estimation techniques are of interest. Second, one could consider how the methods proposed can be extended to repeated measures. Another important extension is to generalize the methods that are described in Section 4 to the situations where the encouragement, treatment and/or outcome variables are multivalued. Finally, the present paper considers missing outcome data alone, but the coexistence of missing data in covariates and outcome variables is also an important topic to be studied.

Acknowledgements

The proposed and other related methods are available as an R package *experiment* at the Comprehensive R Archive Network (`http://cran.r-project.org`). An earlier and longer version of the paper is available as Imai (2007). I acknowledge helpful comments from Larry Bartels, Jennifer Hill, Gary King, Jas Sekhon, Dylan Small, Liz Stuart and the seminar participants at Columbia University (the Mid-Atlantic Causal Inference Workshop), Kobe University (Department of Economics), at the University of Tokyo (Faculty of Economics) and at Queen's University (Department of Economics and School of Public Policy). Financial support from the National Science Foundation (grants SES-0550873 and SES-0752050) and the Princeton University Committee on Research in the Humanities and Social Sciences is acknowledged. Suggestions from two referees, the Associate Editor and the Joint Editor greatly improved the presentation of the materials.

Appendix A: Proofs of propositions

A.1. Proof of proposition 1

Applying the Bayes rule, it is easy to show that these equality constraints are sufficient to identify $p_{00}(x)$ and $p_{10}(x)$, i.e.

$$p_{00}(x) = \frac{p_{01}(x)[\{1-p_{11}(x)\}\pi_{00}(x)\,\pi_{11}(x) - \{1-p_{01}(x)\}\,\pi_{01}(x)\,\pi_{10}(x)]}{\pi_{00}(x)\,\pi_{11}(x)\{p_{01}(x)-p_{11}(x)\}}, \tag{16}$$

$$p_{10}(x) = \frac{p_{11}(x)[\{1-p_{11}(x)\}\,\pi_{00}(x)\,\pi_{11}(x) - \{1-p_{01}(x)\}\,\pi_{01}(x)\,\pi_{10}(x)]}{\pi_{01}(x)\,\pi_{10}(x)\{p_{01}(x)-p_{11}(x)\}}. \tag{17}$$

Substituting these into equation (2) gives the desired expression of the ATE as a function of identifiable quantities. To show that assumption 3 is a minimal assumption for the identification of $\tau_{\mathrm{ATE}}(x)$, recall that without the assumption about the missing data mechanism the observed data impose no restriction on $p_{j0}(x)$ for $j=0,1$. Thus, to identify $\tau_{\mathrm{ATE}}(x)$, at least two linearly independent restrictions are required. Assumption 3 provides such restrictions.

A.2. Proof of proposition 2

For notational simplicity, we omit the conditioning on $X_i = x$ from this proof (i.e. (x) will be omitted). The ML estimator of the ATE, $\hat{\tau}_{\mathrm{ATE}}$, is obtained by replacing π_{jk} and p_{j1} in proposition 1 with their

corresponding sample estimates, i.e. $\hat{\pi}_{00} = \Sigma_{i=1}^n (1-T_i)(1-R_i)/n$, $\hat{\pi}_{01} = \Sigma_{i=1}^n (1-T_i)R_i/n$, $\hat{\pi}_{10} = \Sigma_{i=1}^n T_i(1-R_i)/n$, $\hat{\pi}_{11}\Sigma_{i=1}^n T_i R_i/n$, $\hat{p}_{11} = \Sigma_{i=1}^n Y_i T_i R_i / \Sigma_{i=1}^n T_i R_i$ and $\hat{p}_{01} = \Sigma_{i=1}^n Y_i (1-T_i) R_i / \Sigma_{i=1}^n (1-T_i) R_i$. Then, by applying the multivariate central limit theorem to a vector of independent and identically distributed random variables, V_i, the result immediately follows from the use of the delta method and Slutzky's theorem after some straightforward but tedious algebra. Since Σ and δ which are defined below can be consistently estimated, the asymptotic variance can also be consistently estimated. Finally, Σ is the variance–covariance matrix of a random vector, $V_i \equiv ((1-T_i)R_i, T_i(1-R_i), T_i R_i, Y_i T_i R_i, Y_i(1-T_i)R_i)$, and the δ-vector is equal to

$$\delta = \frac{\delta^*}{\pi_{11}\pi_{01}(p_{01}-p_{11})(\pi_{00}+\pi_{01})(\pi_{10}+\pi_{11})},$$

where the elements of δ^* are given by

$$\delta_1^* = -(\pi_{10}+\pi_{11})A + \frac{A(B-A)p_{11}}{\pi_{01}(p_{01}-p_{11})},$$

$$\delta_2^* = -(p_{11}\pi_{11}+p_{01}\pi_{01})(B-2A) - (\pi_{11}+\pi_{01})A - \frac{A(B-A)(\pi_{00}+\pi_{01}-\pi_{10}-\pi_{11})}{(\pi_{00}+\pi_{01})(\pi_{10}+\pi_{11})},$$

$$\delta_3^* = -(p_{11}\pi_{11}+p_{01}\pi_{01})(B-2A) + (\pi_{00}-\pi_{11})A - A(B-A)\left\{\frac{p_{01}}{(p_{01}-p_{11})\pi_{11}} + \frac{\pi_{00}+\pi_{01}-\pi_{10}-\pi_{11}}{(\pi_{00}+\pi_{01})(\pi_{10}+\pi_{11})}\right\},$$

$$\delta_4^* = (\pi_{00}+\pi_{01})(B-2A) + \frac{A(B-A)}{(p_{01}-p_{11})\pi_{11}},$$

$$\delta_5^* = -(\pi_{10}+\pi_{11})(B-2A) - \frac{A(B-A)}{(p_{01}-p_{11})\pi_{01}},$$

where $A = \pi_{11}p_{11}(\pi_{00}+\pi_{01}) - \pi_{01}p_{01}(\pi_{10}+\pi_{11})$ and $B = \pi_{00}\pi_{11} - \pi_{10}\pi_{01}$.

A.3. Proof of proposition 3

Assumption 6 implies that $\Pr(R_i = 1 | T_i = j, Y_i = k, Z_i = 1) = \Pr(R_i = 1 | T_i = j, Y_i = k, Z_i = 0)$ for $j = 0, 1$ and $k = 0, 1$. Using the Bayes rule, it is easy to show that this constraint is sufficient to identify the unknown probabilities, i.e.

$$p_{j00} = \frac{p_{j10}\{(1-p_{j11})\pi_{j00}\pi_{j11} - (1-p_{j10})\pi_{j10}\pi_{j01}\}}{\pi_{j00}\pi_{j11}(p_{j10}-p_{j11})}, \tag{18}$$

$$p_{j01} = \frac{p_{j11}\{(1-p_{j11})\pi_{j00}\pi_{j11} - (1-p_{j10})\pi_{j10}\pi_{j01}\}}{\pi_{j10}\pi_{j01}(p_{j10}-p_{j11})}, \tag{19}$$

for $j = 0, 1$. Substituting these expressions into equation (12) gives the desired expression of the ITT effect as a function of identifiable quantities. Using the same argument as in the proof of proposition 1, assumption 6 is a minimal assumption for the identification of the ITT effect because it provides a set of four linearly independent constraints.

Appendix B: The EM algorithm for the sensitivity analysis

The observed (conditional) likelihood is given by

$$\begin{aligned} L_{\text{obs}}(\phi_1, \phi_2, \eta_{00}, \eta_{01}, \eta_{10}, \eta_{11} | Y_{\text{obs}}, R, T, X) \\ = \prod_{i=1}^n \{q_{T_i}(X_i;\phi_{T_i}) r_{T_i 1}(X_i;\eta_{T_i 1})\}^{Y_i R_i} [\{1 - q_{T_i}(X_i;\phi_{T_i})\} r_{T_i 0}(X_i;\eta_{T_i 0})]^{(1-Y_i)R_i} \\ \times [q_{T_i}(X_i;\phi_{T_i})\{1 - r_{T_i 1}(X_i;\eta_{T_i 1})\} + \{1 - q_{T_i}(X_i;\phi_{T_i})\}\{1 - r_{T_i 0}(X_i;\eta_{T_i 0})\}]^{1-R_i}, \end{aligned}$$

where Y_{obs} is the observed outcome data. The complete-data (conditional) likelihood function is

$$L_{\text{com}}(\phi_0, \phi_1, \eta_{00}, \eta_{01}, \eta_{10}, \eta_{11} | Y_{\text{obs}}, Y_{\text{mis}}, R, T, X)$$
$$= \prod_{i=1}^{n} [q_{T_i}(X_i; \phi_{T_i})\, r_{T_i 1}(X_i; \eta_{T_i 1})^{R_i} \{1 - r_{T_i 1}(X_i)\}^{1-R_i}]^{Y_i}$$
$$\times [\{1 - q_{T_i}(X_i; \phi_{T_i})\}\, r_{T_i 0}(X_i; \eta_{T_i 0})^{R_i} \{1 - r_{T_i 0}(X_i; \eta_{T_i 0})\}^{1-R_i}]^{1-Y_i},$$

where Y_{mis} denotes the missing outcome data. Thus, to obtain the ML estimates of the model parameters, the following EM algorithm can be used. Let the starting values of the parameters be $(\phi_0^{(0)}, \phi_1^{(0)}, \eta_{00}^{(0)}, \eta_{01}^{(0)}, \eta_{10}^{(0)}, \eta_{11}^{(0)})$. Then, the $(t+1)$th iteration of the algorithm is given by the following steps.

(a) *E-step*: for each i, set

$$\tilde{Y}_i^{(t+1)} = \begin{cases} \dfrac{\{1 - r_{j1}(X_i; \eta_{j1}^{(t)})\}\, q_j(X_i; \phi_j^{(t)})}{\{1 - r_{j1}(X_i; \eta_{j1}^{(t)})\} q_j(X_i; \phi_j^{(t)}) + \{1 - r_{j0}(X_i; \eta_{j0}^{(t)})\}\{1 - q_j(X_i; \phi_j^{(t)})\}} & \text{if } (R_i, T_i) = (0, j), \\ Y_i & \text{if } (R_i, T_i) = (1, j), \end{cases}$$

for $j = 0, 1$.

(b) *M-step*: find the values of parameters which maximize the following function and set them as $(\phi_0^{(t+1)}, \phi_1^{(t+1)}, \eta_{00}^{(t+1)}, \eta_{01}^{(t+1)}, \eta_{10}^{(t+1)}, \eta_{11}^{(t+1)})$,

$$\sum_{i=1}^{n} \tilde{Y}_i^{(t+1)} [\log\{q_{T_i}(X_i; \phi_{T_i})\} + R_i \log\{r_{T_i 1}(X_i; \eta_{T_i 1})\} + (1 - R_i) \log\{1 - r_{T_i 1}(X_i; \eta_{T_i 1})\}]$$
$$+ (1 - \tilde{Y}_i^{(t+1)}) [\log\{1 - q_{T_i}(X_i; \phi_{T_i})\} + R_i \log\{r_{T_i 0}(X_i; \eta_{T_i 0})\} + (1 - R_i) \log\{1 - r_{T_i 0}(X_i; \eta_{T_i 0})\}].$$

This step can be accomplished by using a numerical optimization algorithm.

References

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996) Identification of causal effects using instrumental variables (with discussion). *J. Am. Statist. Ass.*, **91**, 444–455.

Barnard, J., Frangakis, C. E., Hill, J. L. and Rubin, D. B. (2003) Principal stratification approach to broken randomized experiments: a case study of school choice vouchers in New York (with discussion). *J. Am. Statist. Ass.*, **98**, 299–311.

Burden, B. C. (2000) Voter turnout and the national election studies. *Polit. Anal.*, **8**, 389–398.

Cox, D. R. (1958) *Planning of Experiments*. New York: Wiley.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Diggle, P. J. and Kenward, M. G. (1994) Informative drop-out in longitudinal data analysis (with discussion). *Appl. Statist.*, **43**, 49–93.

Frangakis, 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.

Gerber, A. S. and Green, D. P. (2000) The effects of canvassing, telephone calls, and direct mail on voter turnout: a field experiment. *Am. Polit. Sci. Rev.*, **94**, 653–663.

Goldstein, D. G., Imai, K. and Göritz, A. S. (2007) The subtle psychology of voter turnout. *Technical Report*. Department of Politics, Princeton University, Princeton. (Available from `http://ssrn.com/abstract=977000`.)

Gollwitzer, P. M. (1999) Implementation intentions. *Am. Psychol.*, **54**, 493–503.

Hirano, K., Imbens, G., Ridder, G. and Rubin, D. B. (2001) Combining panel data sets with attrition and refreshment samples. *Econometrica*, **69**, 1645–1659.

Holland, P. W. (1986) Statistics and causal inference (with discussion). *J. Am. Statist. Ass.*, **81**, 945–960.

Horiuchi, Y., Imai, K. and Taniguchi, N. (2007) Designing and analyzing randomized experiments: application to a Japanese election survey experiment. *Am. J. Polit. Sci.*, **51**, 669–687.

Horowitz, J. L. and Manski, C. F. (2000) Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Statist. Ass.*, **95**, 77–84.

Imai, K. (2005) Do get-out-the-vote calls reduce turnout?: the importance of statistical methods for field experiments. *Am. Polit. Sci. Rev.*, **99**, 283–300.

Imai, K. (2007) Statistical analysis of randomized experiments with non-ignorable missing binary outcomes. *Technical Report*. Department of Politics, Princeton University, Princeton. (Available from `http://ssrn.com/abstract=1014849`.)

Imai, K. (2008) Experiment: R package for designing and analyzing randomized experiments. (Available from `http://cran.r-project.org`.)
Imbens, G. W. (2004) Nonparametric estimation of average treatment effects under exogeneity: a review. *Rev. Econ. Statist.*, **86**, 4–29.
King, G. and Zeng, L. (2002) Estimating risk and rate levels, ratios, and differences in case-control studies. *Statist. Med.*, **21**, 1409–1427.
Lin, D. Y., Psaty, B. M. and Kronmal, R. A. (1998) Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, **54**, 948–963.
Little, R. J. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.
Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc.* B, **44**, 226–233.
Mealli, F., Imbens, G. W., Ferro, S. and Biggeri, A. (2004) Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, **5**, 207–222.
O'Malley, A. J. and Normand, S.-L. T. (2005) Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. *Biometrics*, **61**, 325–334.
Orbell, S., Hodgkins, S. and Sheeran, P. (1997) Implementation intensions and the theory of planned behavior. *Personlty Socl Psychol. Bull.*, **23**, 945–954.
Robins, J. M., Rotnitzky, A. and Scharfstein, D. (1999) Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology* (eds E. Halloran and D. Berry), pp. 1–92. New York: Springer.
Rosenbaum, P. R. (1987) Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, **74**, 13–26.
Rosenbaum, P. R. (2002) *Observational Studies*, 2nd edn. New York: Springer.
Rosenbaum, P. R. and Rubin, D. B. (1983a) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.
Rosenbaum, P. R. and Rubin, D. B. (1983b) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. R. Statist. Soc.* B, **45**, 212–218.
Rubin, D. B. (1990) Comments on "On the application of probability theory to agricultural experiments: Essay on principles, Section 9", by J. Splawa-Neyman. *Statist. Sci.*, **5**, 472–480.
Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999) Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *J. Am. Statist. Ass.*, **94**, 1096–1120.
Stubbendick, A. L. and Ibrahim, J. G. (2003) Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics*, **59**, 1140–1150.
Vansteelandt, S. and Goetghebeur, E. (2001) Analyzing the sensitivity of generalized linear models to incomplete outcomes via the IDE algorithm. *J. Computnl Graph. Statist.*, **10**, 656–672.
Verzilli, C. and Carpenter, J. R. (2002) Assessing uncertainty about parameter estimates with incomplete repeated ordinal data. *Statist. Modllng*, **2**, 203–215.
Yau, L. H. Y. and Little, R. J. (2001) Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *J. Am. Statist. Ass.*, **96**, 1232–1244.