

# A Sensitivity Analysis for Missing Outcomes Due to Truncation-by-Death under the Matched-Pairs Design

Kosuke Imai\*      Zhichao Jiang†

September 9, 2017

## Abstract

The matched-pairs design enables researchers to efficiently infer causal effects from randomized experiments. In this paper, we exploit the key feature of the matched-pairs design and develop a sensitivity analysis for missing outcomes due to truncation-by-death, in which the outcomes of interest (e.g., quality of life measures) are not even well defined for some units (e.g., deceased patients). The idea is that if two nearly identical observations are paired prior to the randomization of the treatment, the missingness of one unit's outcome is informative about the potential missingness of the other unit's outcome under an alternative treatment condition. We consider the average treatment effect among always-observed pairs (ATOP) whose units exhibit no missing outcome regardless of their treatment status. The naive estimator based on available pairs is unbiased for the ATOP if two units of the same pair are identical in terms of their missingness patterns. The proposed sensitivity analysis characterizes how the bounds of the ATOP widen as the degree of the within-pair similarity decreases. We further extend the methodology to the matched-pairs design in observational studies. Our simulation studies show that informative bounds can be obtained under some scenarios when the proportion of missing data is not too large. The proposed methodology is also applied to the randomized evaluation of the Mexican universal health insurance program.

**Keywords:** average treatment effect, bounds, causal inference, observational studies, randomized experiments

---

\*Professor, Department of Politics and Center for Statistics and Machine Learning, Princeton University, Princeton NJ 08544. Phone: 609-258-6601, Email: kimai@princeton.edu, URL: <http://imai.princeton.edu>

†Postdoctoral Fellow, Department of Politics and Center for Statistics and Machine Learning, Princeton University, Princeton NJ 08544.

# 1 Introduction

As the well known mantra “block what you can and randomize what you cannot” (Box et al., 2005, p. 93) suggests, covariate adjustment prior to the randomization of treatment assignment is an effective strategy for ascertaining the treatment effects from randomized experiments (e.g., Imai et al., 2008). In particular, it has been shown that the matched-pairs design, in which two similar observations are paired and the randomization of treatment assignment is done within each pair, can yield an efficient estimate of causal effects (e.g., Martin et al., 1993; Greevy et al., 2004; Imai et al., 2009).

However, it is well-known that the validity of experimental findings can be compromised in the presence of missing outcomes, which may arise due to drop-outs and survey non-response (e.g., Horowitz and Manski, 2000; Frangakis and Rubin, 2002). In this paper, we exploit the key feature of the matched-pairs design and develop a sensitivity analysis for missing outcomes due to truncation-by-death (Zhang and Rubin, 2003; Imai, 2008a), in which the outcomes of interest (e.g., quality of life measures) are not even well defined for some units (e.g., deceased patients). Our proposed methodology is based on the idea that if two nearly identical observations are paired prior to the randomization of the treatment, the missingness of one unit’s outcome is informative about the potential missingness of the other unit’s outcome. In other words, we may assume that similar units within a pair are comparable in terms of their outcome missingness patterns.

We consider the average treatment effect among always-observed pairs (ATOP) whose units exhibit no missing outcome regardless of their treatment status. This estimand is the matched-pairs design analogue of the the average treatment effect among always-observed units (ATOU) studied under the truncation-by-death problem in the literature. The naive estimator based on available pairs, which is often used in practice, is biased for the ATOP unless two units of the same pair are identical in terms of their potential missingness patterns. The proposed sensitivity analysis characterizes how the bounds of the ATOP widen as the degree of the within-pair similarity decreases. We further extend the methodology to the matched-pairs design in observational studies, in which the treatment assignment is assumed to be unconfounded with respect to potential outcomes but not potential values of missingness indicators. Our simulation studies show that informative bounds can be obtained under some scenarios if the proportion of missing data is not too large. The analytical results and methodology presented in this paper complement the recent work by Fukumoto (2015) who examine the bias due to missing outcomes under the matched-pairs design.

Finally, we apply the proposed methodology to the randomized evaluation of the Mexican universal health insurance program (King et al., 2009), in which randomly selected clusters of households

are encouraged to enroll in the health insurance program, Seguro Popular. We examine the outcome variable measuring the satisfaction with Seguro Popular. This outcome variable suffers from the truncation-by-death problem because it is defined and hence measured only for those households who are enrolled in the program. We conduct the proposed sensitivity analysis for the ATOP as well as the ATOU.

## 2 The Proposed Methodology

### 2.1 The Setup

Following Imai (2008b), we consider a randomized experiment with the standard matched-pairs design. Suppose that we have a sample of  $n$  pairs. Each pair consists of two units that are similar to one another in terms of their pre-treatment covariates  $\mathbf{X}$ , which is a  $K$  dimensional vector. To create such pairs, for example, researchers may employ multivariate matching based on the Mahalanobis distance measure (Greevy et al., 2004). The treatment assignment is completely randomized within each pair where one unit receives the treatment and the other unit does not. Let  $T_{ij}$  represent the binary treatment variable for unit  $i$  of pair  $j$  where  $i = 1, 2$  and  $j = 1, 2, \dots, n$ . Then, under the matched-pairs design, we have  $T_{1j} + T_{2j} = 1$  for each pair  $j$  and  $\Pr(T_{ij} = 1) = 1/2$ .

Our analysis is based on the potential outcomes framework (Rubin, 1974; Neyman, 1990), and throughout this paper, we maintain the Stable Unit Treatment Value Assumption (SUTVA). That is, we assume that there is only one version of treatment and there is no interference between units (Rubin, 1980). The SUTVA allows us to uniquely define the potential outcomes  $Y_{ij}(t)$  under the treatment condition  $T_{ij} = t$  for unit  $i$  of pair  $j$  where  $t = 0, 1$ ,  $i = 1, 2$ , and  $j = 1, 2, \dots, J$ . Since we only observe one of the two potential outcomes, we write the observed outcome as  $Y_{ij} = T_{ij}Y_{ij}(1) + (1 - T_{ij})Y_{ij}(0)$ .

In this paper, we consider the setting where the outcome variable is missing for some units due to the truncation-by-death problem, in which the outcome variable may not even be well defined under a certain circumstance. For example, in a medical trial, the quality of life measure is not well defined for deceased patients. Let  $R_{ij}(t)$  represent the potential value of response indicator for the outcome variable where  $R_{ij}(t) = 1$  ( $R_{ij}(t) = 0$ ) implies that the outcome is observed (missing) for unit  $i$  of pair  $j$  under the treatment condition  $T_{ij} = t$ . This notation relies on the SUTVA but allows the treatment to directly affect missing data pattern. Like the outcome variable, we only observe one of the two potential values of the response indicator, and hence we write the observed value of this indicator as  $R_{ij} = T_{ij}R_{ij}(1) + (1 - T_{ij})R_{ij}(0)$ . Under the truncation-by-death problem, the

potential outcome  $Y_{ij}(t)$  is not well defined when  $R_{ij}(t) = 0$ .

The randomization of treatment assignment within each pair implies the following independence relationship between the treatment assignment and the potential values of outcome and response indicator,

ASSUMPTION 1 (WITHIN-PAIR RANDOMIZATION OF TREATMENT ASSIGNMENT) *Within each pair*  
 $j = 1, 2, \dots, J$ ,

$$T_{ij} \perp\!\!\!\perp \{Y_{1j}(1), Y_{1j}(0), Y_{2j}(1), Y_{2j}(0), R_{1j}(1), R_{1j}(0), R_{2j}(1), R_{2j}(0)\}$$

and  $\Pr(T_{ij} = 1) = 1/2$  where  $i = 1, 2$ .

In observational studies, researchers often match a treated unit with a control unit based on the similarity of pre-treatment characteristics  $X$ . In this case, there are two types of confounders: the variables that confound the causal relationship between the treatment and outcome variables, and those that confound the relationship between the treatment and response indicator variables. Assumption 1 implies that neither type of confounders exist.

Without loss of generality, we assume that within each pair the unit labels,  $i = 1, 2$ , are arbitrarily assigned and hence these two units are exchangeable (Imai, 2008b). Formally, we write this assumption as the following equality of two joint distributions,

$$\begin{aligned} & P\{Y_{1j}(1), Y_{1j}(0), Y_{2j}(1), Y_{2j}(0), R_{1j}(1), R_{1j}(0), R_{2j}(1), R_{2j}(0)\} \\ &= P\{Y_{2j}(1), Y_{2j}(0), Y_{1j}(1), Y_{1j}(0), R_{2j}(1), R_{2j}(0), R_{1j}(1), R_{1j}(0)\}. \end{aligned}$$

We will maintain this exchangeability assumption throughout this paper. In practice, the randomization of unit labels within each pair is sufficient to ensure that the assumption holds.

## 2.2 The Quantities of Interest

Typically, researchers are interested in the average treatment effect (ATE), which is defined as follows,

$$\text{ATE} = \mathbb{E}\{Y_{ij}(1) - Y_{ij}(0)\}.$$

However, when the outcome variable is missing due to truncation-by-death, researchers may focus on the average treatment effect for those units whose outcome is defined regardless of their treatment status. We define the average treatment effect for always observed units (ATOU),

$$\text{ATOU} = \mathbb{E}\{Y_{ij}(1) - Y_{ij}(0) \mid R_{ij}(1) = R_{ij}(0) = 1\},$$

To take advantage of the matched-pairs design, we consider the average treatment effect for always-observed pairs (ATOP), which is a quantity analogous to the ATOU under this design. The idea is that we focus on a subset of pairs in which the outcome variable is well defined for both units regardless of their treatment status. Formally, the ATOP is defined as follows,

$$\text{ATOP} = \mathbb{E}\{Y_{ij}(1) - Y_{ij}(0) \mid R_{1j}(1) = R_{1j}(0) = R_{2j}(1) = R_{2j}(0) = 1\}. \quad (1)$$

It is clear that if two similar units are matched such that  $R_{1j}(t) = R_{2j}(t)$  for each  $t = 0, 1$  and  $j = 1, 2, \dots, J$ , then the available-pair analysis will identify the ATOP. Specifically, under this scenario, the following difference-in-means estimator, applied to all the pairs without missing outcomes, is unbiased for the ATOP.

$$\hat{\tau}_{\text{ATOP}} = \frac{1}{\sum_{j=1}^J R_{1j}R_{2j}} \sum_{j=1}^J R_{1j}R_{2j}(2T_{1j} - 1)(Y_{1j} - Y_{2j}). \quad (2)$$

In practice, however, matching may be imperfect such that  $R_{1j}(t) \neq R_{2j}(t)$  for some  $j$  and  $t$ . Therefore, we derive the bounds for the ATOP and develop a sensitivity analysis.

### 2.3 No-assumption Bounds

We begin by deriving the bounds for the ATOP without an additional assumption. First, the following quantities are identified from the observed data,

$$\pi = P(R_{1j} = R_{2j} = 1) \quad (3)$$

$$\alpha_t = P(R_{1j} = 1 \mid T_{1j} = t) = P(R_{2j} = 1 \mid T_{2j} = t) \quad (4)$$

$$\psi_t = \mathbb{E}(Y_{1j} \mid R_{1j} = 1, T_{1j} = t) = \mathbb{E}(Y_{2j} \mid R_{2j} = 1, T_{2j} = t) \quad (5)$$

$$\omega_t = \mathbb{E}(Y_{1j} \mid T_{1j} = t, R_{1j} = R_{2j} = 1) = \mathbb{E}(Y_{2j} \mid T_{2j} = t, R_{1j} = R_{2j} = 1) \quad (6)$$

for  $t = 0, 1$  where in equations (3)–(6) the equalities hold due to the exchangeability of unit labels.

The following theorem establishes the sharp bounds for the ATOP.

#### THEOREM 1 (SHARP BOUNDS FOR THE AVERAGE TREATMENT EFFECT FOR ALWAYS-OBSERVED PAIRS)

*Suppose that the outcome variable is bounded such that  $Y_{ij} \in [l, u]$  where  $l < u$ . Then, under Assumption 1, the sharp bounds for the average treatment effect for always-observed pairs (ATOP) defined in equation (1) are given by,*

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi}), g^{[l,u]}(\omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi})] & \text{if } \pi > \frac{1}{2} \\ [l - u, u - l] & \text{if } \pi \leq \frac{1}{2} \end{cases}$$

where

$$f^{[l,u]}(x_1, y_1, x_0, y_0) = \max \left\{ l, u - \frac{u - x_1}{y_1} \right\} - \min \left\{ u, \frac{x_0 - l}{y_0} + l \right\},$$

$$g^{[l,u]}(x_1, y_1, x_0, y_0) = \min \left\{ u, \frac{x_1 - l}{y_1} + l \right\} - \max \left\{ l, u - \frac{u - x_0}{y_0} \right\}.$$

Proof is given in Appendix A.2. Theorem 1 shows that if the proportion of the observed pairs, i.e.,  $\pi$ , is high, then the bounds will be informative. The informativeness of the bounds also depends on the (expected) magnitude of naive ATOP estimator  $\hat{\tau}_{\text{ATOP}}$  given in equation (2), which is equal to  $\omega_1 - \omega_0$ . To see this, we note that the definitions of  $f^{[l,u]}$  and  $g^{[l,u]}$  imply,

$$\begin{aligned} f^{[l,u]} \left( \omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi} \right) &\geq \frac{\pi(\omega_1 - \omega_0)}{2\pi - 1} - \frac{2(1 - \pi)(u - l)}{2\pi - 1}, \\ g^{[l,u]} \left( \omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi} \right) &\leq \frac{\pi(\omega_1 - \omega_0)}{2\pi - 1} + \frac{2(1 - \pi)(u - l)}{2\pi - 1}. \end{aligned}$$

Therefore, when the naive estimator has a value away from zero, then the bounds are less likely to contain zero. The above inequalities further imply that the width of the bound is less than or equal to,

$$\frac{4(1 - \pi)}{2\pi - 1}(u - l).$$

Thus, as expected, when the proportion of observed pairs is greater, the bound is narrower.

We now compare the bounds for the ATOP with those for the ATOU. Zhang and Rubin (2003) derived the no-assumption bounds for the ATOU and the bounds under the monotonicity assumption  $R(1) \geq R(0)$ . Imai (2008a) proved that these bounds are sharp. The bounds for the ATOU below are expressed differently from those presented in the literature, but they equal to each other.

**THEOREM 2 (SHARP BOUNDS FOR THE AVERAGE TREATMENT EFFECT FOR ALWAYS-OBSERVED UNITS)**  
*If the outcome is bounded such that  $Y_{ij} \in [l, u]$  where  $l < u$ , then under Assumption 1, the lower and upper bounds for the the average treatment effect for always-observed units (ATOOU) are given by*

$$\text{ATOOU} \in \begin{cases} \left[ f^{[l,u]} \left( \psi_1, \frac{\Gamma}{\alpha_1}, \psi_0, \frac{\Gamma}{\alpha_0} \right), g^{[l,u]} \left( \psi_1, \frac{\Gamma}{\alpha_1}, \psi_0, \frac{\Gamma}{\alpha_0} \right) \right] & \text{if } \Gamma > 0 \\ [l - u, u - l] & \text{if } \Gamma \leq 0 \end{cases}$$

where  $\Gamma = \alpha_1 + \alpha_0 - 1$ .

Proof is given in Appendix A.3. Similar to the case of ATOP, the width of the bounds critically depends on  $\Gamma$ , implying that a greater missing proportion leads to wider bounds. On one hand, the bounds of the ATOP given in Theorem 1 are more likely to be  $[l - u, u - l]$  than those of the ATOU in Theorem 2. This is because  $\alpha_1 + \alpha_0 \leq 1$  implies  $\pi \leq 1/2$ . On the other hand, because in most cases,  $\Gamma/\alpha_i$  is larger than  $2 - 1/\pi$ , so the bound width in Theorem 2 is usually narrower than the bound width in Theorem 1. In a similar setting, Ding et al. (2011) point out that, even under the

monotonicity assumption ( $R(1) \leq R(0)$  or  $R(1) \geq R(0)$ ), the bounds for the ATOU may be too wide to be useful. Therefore, the bounds in Theorem 1 may give us little information about the ATOP. To address this issue, we next develop a sensitivity analysis. In particular, we exploit the fact that the assumption about within-pair similarity can be used to narrow the bounds for the ATOP.

## 2.4 Sensitivity Analysis

The goal of the matched-pairs design is to find pairs of similar observations based on a set of pre-treatment covariates. The idea is that if matching is successful, then the observed outcome of one unit can be used as a good estimate of the unobserved potential outcome of the other unit within the same pair, i.e.,  $Y_{1j}(t) \approx Y_{2j}(t)$ . We apply this logic to the missingness pattern and develop a sensitivity analysis. That is, if matching is successful, the observed value of one unit's response indicator variable should be a reasonable estimate of the unobserved value of the response indicator for the other unit within the same pair, i.e.,  $R_{1j}(t) \approx R_{2j}(t)$ . We formalize this assumption as follows:

ASSUMPTION 2 (SIMILARITY OF MISSINGNESS PATTERN WITHIN A PAIR)

$$P\{R_{2j}(t) = r \mid R_{1j}(t) = r\} = P\{R_{1j}(t) = r \mid R_{2j}(t) = r\} \geq \gamma$$

for  $r = 0, 1$  and  $t = 0, 1$ .

Under Assumption 2, if  $\gamma = 1$ , then we have  $R_{1j}(t) = R_{2j}(t)$  for  $t = 0, 1$  and therefore the naive estimator  $\hat{\tau}_{\text{ATOP}}$  given in equation (2) is unbiased for the ATOP. By moving the sensitivity parameter  $\gamma$  away from 1, we can characterize the maximal magnitude of bias induced in the estimation of the ATOP. The next theorem formally characterizes the bounds for the ATOP as a function of the sensitivity parameter  $\gamma$ :

THEOREM 3 (SENSITIVITY ANALYSIS FOR THE AVERAGE TREATMENT EFFECT FOR ALWAYS-OBSERVED PAIRS)

*Under Assumptions 1 and 2, the sharp bounds for the ATOP are*

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta), g^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta)] & \text{if } \Delta > 0 \\ [l - u, u - l] & \text{if } \Delta \leq 0 \end{cases}$$

where  $\Delta = \max_{i \in \{0,1\}} \{2\pi - 1 + \gamma(1 - \alpha_i), \pi - (1 - \gamma)(\alpha_1 + \alpha_0), 2\pi - (2 - \gamma)\alpha_i, \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0), \pi - (1 - \gamma)(1 - |\alpha_1 - \alpha_0|)\} / \pi$ .

Proof is given in Appendix A.4. In typical situations with  $\alpha_i > 1/2$  for  $i = 0, 1$ , we can simplify the expression of  $\Delta$  as  $\Delta = \max_{i \in \{0,1\}} \{2\pi - 1 + \gamma(1 - \alpha_i), \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0)\} / \pi$ .

From the expressions of the bounds, we have

$$\begin{aligned} f^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta) &\geq \frac{\omega_1 - \omega_0}{\Delta} - \left(\frac{1}{\Delta} - 1\right)(u - l), \\ g^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta) &\leq \frac{\omega_1 - \omega_0}{\Delta} + \left(\frac{1}{\Delta} - 1\right)(u - l). \end{aligned}$$

Thus, similar to Theorem 1, the informativeness of these sensitivity bounds depends on the (expected) magnitude of the naive ATOP estimator  $\hat{\tau}_{\text{ATOP}}$ . In addition, the width of the bounds is less than or equal to,

$$2(u - l) \left(\frac{1}{\Delta} - 1\right),$$

which is decreasing  $\Delta$ . When  $\gamma = 1$ , we have  $\Delta = 1$ , reducing the bounds to a point and identifying the ATOP as  $\omega_1 - \omega_0$ . When  $\alpha_i > 1/2$  for  $i = 0, 1$ , the width of the bounds is decreasing in  $\gamma$ . This is consistent with our expectation that when the units within each pair are similar the bounds should be narrower. Interestingly, the width of the bounds is not necessarily decreasing in  $\alpha_i$  unless the within-pair similarity is high as indicated by a large value of  $\gamma$ .

## 2.5 Extension to Observational Studies

Finally, we extend our results to observational studies and develop a sensitivity analysis for pair-matching. We consider the setting where a set of observed confounders are used to create pairs of similar observations. A variety of matching methods such as Mahalanobis and propensity score matching can be used for the construction of pairs, and we make the standard within-pair unconfoundedness assumption (e.g., Rubin, 2006; Ho et al., 2007; Stuart, 2010).

**ASSUMPTION 3 (WITHIN-PAIR UNCONFOUNDEDNESS OF TREATMENT ASSIGNMENT)** *Within each pair,  $j = 1, 2, \dots, J$ ,*

$$T_{ij} \perp\!\!\!\perp \{Y_{1j}(0), Y_{1j}(1), Y_{2j}(0), Y_{2j}(1)\}$$

*and  $\Pr(T_{ij} = 1) = 1/2$  where  $i = 1, 2$ .*

Unlike Assumption 1, we do not assume that the treatment assignment is independent of potential response indicator variables. This allows for the possible existence of unobserved confounders that are associated with the treatment assignment and potential response indicators even within each pair.

Assumption 3 alone leads to non-informative bounds, and yet, researchers may feel that Assumption 1 is not credible in observational studies. Thus, we consider the following conditional independence assumption, which still allows for the correlation between  $T_{ij}$  and  $\{R_{1j}(0), R_{1j}(1), R_{2j}(0), R_{2j}(1)\}$  as well as the correlation between  $\{Y_{1j}(0), Y_{1j}(1), Y_{2j}(0), Y_{2j}(1)\}$  and  $\{R_{1j}(0), R_{1j}(1), R_{2j}(0), R_{2j}(1)\}$ .



ASSUMPTION 4 (PATTERN MIXTURE MODEL) *For each pair  $j = 1, 2, \dots, J$ , we assume the following conditional independence,*

$$Y_{ij}(t) \perp\!\!\!\perp T_{ij} \mid \{R_{ij}(t), R_{i'j}(1-t)\}$$

for  $i \neq i'$  and  $t = 0, 1$ .

In other words, we assume that the within-pair unconfoundedness of treatment assignment even after conditioning on the potential values of response indicator variables.

We motivate Assumption 4 by pattern mixture models that are widely used in the missing data literature (e.g., Little, 1993). In the causal inference literature, for example, D'Agostino and Rubin (2000) used a pattern mixture model to account for partially observed covariates in the estimation of propensity scores. In the current context, we consider the setting where we construct matched pairs based on a set of observed confounders  $X$  for the treatment-outcome relationship (such that Assumption 3 holds) but there exists another set of unobserved pre-treatment covariates that confound the relationship between the treatment and missingness (such that  $T_{ij} \not\perp\!\!\!\perp \{R_{1j}(0), R_{1j}(1), R_{2j}(0), R_{2j}(1)\}$ ). Then, the pattern mixture decomposition is given by,

$$P(Y_{1j}, Y_{2j}, R_{1j}, R_{2j}, T_{1j}) = P(Y_{1j}, Y_{2j} \mid R_{1j}, R_{2j}, T_{1j})P(R_{1j}, R_{2j}, T_{1j})$$

Assumption 4 implies that within each missingness pattern the treatment assignment is independent of potential outcomes.

Under Assumptions 3 and 4, we develop a sensitivity analysis by operationalizing the possible existence of unobserved confounders for the treatment and missingness relationship as

$$\frac{\Pr(R_{ij}(t) = 1 \mid T_{ij} = 1)}{\Pr(R_{ij}(t) = 1 \mid T_{ij} = 0)} = \kappa_t \tag{7}$$

for  $t = 0, 1$ . It is clear that when  $\kappa_1 = \kappa_0 = 1$ , we have  $\{R_{1j}(0), R_{1j}(1), R_{2j}(0), R_{2j}(1)\} \perp\!\!\!\perp T_{ij}$ . Because of the relationships,  $\Pr(R_{ij}(t) = 1 \mid T_{ij} = 0) = \alpha_t / \kappa_t$  for  $t = 0, 1$ , the observed data have some restrictions on the possible ranges of  $\kappa_1$  and  $\kappa_0$ , i.e.,  $\kappa_1 \geq \alpha_1$  and  $\kappa_0 \leq 1/\alpha_0$ . Furthermore, we can write the proportion of missingness using these sensitivity parameters,

$$\Pr(R_{ij}(1) = 1) = \frac{\alpha_1(1 + \kappa_1)}{2\kappa_1}, \quad \Pr(R_{ij}(0) = 1) = \frac{\alpha_0(1 + \kappa_0)}{2} \tag{8}$$

for  $t = 0, 1$ . Using this relationship, the next theorem shows that under Assumptions 3 and 4 the sharp bounds for the ATOP can be expressed as a function of these sensitivity parameters, i.e.,  $\kappa_0$  and  $\kappa_1$ .

**THEOREM 4 (SENSITIVITY ANALYSIS FOR OBSERVATIONAL STUDIES)** *Suppose that the outcome variable is bounded such that  $Y_{ij} \in [l, u]$  where  $l < u$ . Then, under Assumptions 3 and 4, the sharp bounds for the average treatment effect for always-observed pairs (ATOP) are*

$$ATOP \in \begin{cases} [f^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*), g^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*)] & \text{if } \Delta^* > 0 \\ [l - u, u - l] & \text{if } \Delta^* \leq 0 \end{cases}$$

where  $\Delta^* = \max_{t \in \{0,1\}} \{2\pi - 2 + 2\gamma(1 - \alpha_t^*), \pi - 2(1 - \gamma)(\alpha_1^* + \alpha_0^*), 2\pi - 2(2 - \gamma)\alpha_t^*, \pi - 2(1 - \gamma)(2 - \alpha_1^* - \alpha_0^*), \pi - 2(1 - \gamma)(1 - |\alpha_1^* - \alpha_0^*|)\} / \pi$ , and  $\alpha_t^* = \Pr(R(t) = 1)$  defined in equation (8).

Proof is given in Appendix A.5. Because  $\Delta^* \leq \Delta$ , the bounds in Theorem 4 are always wider than those in Theorem 3. In typical situations with  $\alpha_i^* > 1/2$  for  $i = 0, 1$ , we can simplify the expression of  $\Delta^*$  as  $\Delta^* = \max_{i \in \{0,1\}} \{2\pi - 2 + 2\gamma(1 - \alpha_i^*), \pi - 2(1 - \gamma)(2 - \alpha_1^* - \alpha_0^*)\} / \pi$ . From the expressions of the bounds, we have

$$\begin{aligned} f^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*) &\geq \frac{\omega_1 - \omega_0}{\Delta^*} - \left(\frac{1}{\Delta^*} - 1\right)(u - l), \\ g^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*) &\leq \frac{\omega_1 - \omega_0}{\Delta^*} + \left(\frac{1}{\Delta^*} - 1\right)(u - l). \end{aligned}$$

Thus, similar to Theorems 1 and 3, the informativeness of these sensitivity bounds depends on the (expected) magnitude of the naive ATOP estimator  $\hat{\tau}_{ATOP}$ . In addition, the width of the bounds is less than or equal to,

$$2(u - l) \left(\frac{1}{\Delta^*} - 1\right),$$

which is decreasing  $\Delta^*$ . When  $\gamma = 1$ , we have  $\Delta^* = 1$ , reducing the bounds to a point and identifying the ATOP as  $\omega_1 - \omega_0$ . When  $\alpha_i^* > 1/2$  for  $i = 0, 1$ , the width of the bounds is decreasing in  $\gamma$ , and is decreasing in  $\alpha_i^*$  if the within-pair similarity is high.

### 3 Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of our approach under several different missing data mechanisms. To be consistent with Assumption 1, we consider randomized experiments where the treatment is independently and randomly assigned to one of the units within each pair, i.e.,  $\Pr(T_{1j} = 1) = 0.5$  and  $T_{2j} = 1 - T_{1j}$  for each  $j$ . Next, we generate the potential response indicator variables, i.e.,  $\{R_1(1), R_1(0), R_2(1), R_2(0)\}$ . Let us denote  $\zeta_{ijkl} = \Pr\{R_1(1) = i, R_1(0) = j, R_2(1) = k, R_2(0) = l\}$ . Because the labels are arbitrary within each pair, we have  $\zeta_{1011} = \zeta_{1110}$ ,  $\zeta_{1001} = \zeta_{0110}$ ,  $\zeta_{1100} = \zeta_{0011}$  and  $\zeta_{0001} = \zeta_{0100}$ . We generate  $\{R_1(1), R_1(0), R_2(1), R_2(0)\}$  from a multinomial distribution with probabilities  $\zeta =$

( $\zeta_{1111}, \zeta_{1101}, \zeta_{0111}, \zeta_{1011}, \zeta_{1110}, \zeta_{1001}, \zeta_{0110}, \zeta_{1010}, \zeta_{0101}, \zeta_{1100}, \zeta_{0011}, \zeta_{0001}, \zeta_{0100}, \zeta_{1000}, \zeta_{0010}, \zeta_{0000}$ ). Based on the values of  $\zeta$ , we can calculate the proportions of missingness under the treatment and control conditions as follows,

$$\Pr\{R_{ij}(1) = 1\} = \sum_{j,k,l=0,1} \zeta_{1jkl} \quad \text{and} \quad \Pr\{R_{ij}(0) = 1\} = \sum_{i,k,l=0,1} \zeta_{i1kl}.$$

We consider the following three scenarios with different missing proportions under the treatment and control conditions:

1. **Low level of missingness:**  $\Pr\{R_{ij}(1) = 1\} = 0.85$  and  $\Pr\{R_{ij}(0) = 1\} = 0.9$  where  $\zeta = (3/4, 1/40, 1/40, 1/80, 1/80, 1/80, 1/80, 1/80, 1/80, 1/20, 1/80, 1/80, 1/80, 1/80, 1/80, 1/80, 1/80)$ .
2. **Moderate level of missingness:**  $\Pr\{R_{ij}(1) = 1\} = 0.7$  and  $\Pr\{R_{ij}(0) = 1\} = 0.75$  where  $\zeta = (39/80, 1/40, 1/40, 1/32, 1/32, 1/32, 1/32, 1/32, 7/80, 1/32, 1/32, 1/32, 1/32, 1/32, 1/32, 1/32)$ .
3. **High level of missingness:**  $\Pr\{R_{ij}(1) = 1\} = 0.55$  and  $\Pr\{R_{ij}(0) = 1\} = 0.6$  where  $\zeta = (9/40, 1/40, 1/40, 1/20, 1/20, 1/20, 1/20, 1/20, 1/8, 1/20, 1/20, 1/20, 1/20, 1/20, 1/20, 1/20)$ .

Finally, we generate potential outcomes conditional on the potential response indicator so that missing data are not ignorable. Specifically, for each of the above missingness mechanism, we use,

$$\begin{aligned} \Pr\{Y_{ij}(1) = 1 \mid R_{ij}(1) = 0, R_{1-i,j}(0) = 0\} &= 0.2, & \Pr\{Y_{ij}(1) = 1 \mid R_{ij}(1) = 0, R_{1-i,j}(0) = 1\} &= 0.7, \\ \Pr\{Y_{ij}(1) = 1 \mid R_{ij}(1) = 1, R_{1-i,j}(0) = 0\} &= 0.3, & \Pr\{Y_{ij}(1) = 1 \mid R_{ij}(1) = 1, R_{1-i,j}(0) = 1\} &= 0.6, \\ \Pr\{Y_{ij}(0) = 1 \mid R_{ij}(0) = 0, R_{1-i,j}(1) = 0\} &= 0.7, & \Pr\{Y_{ij}(0) = 1 \mid R_{ij}(0) = 0, R_{1-i,j}(1) = 1\} &= 0.6, \\ \Pr\{Y_{ij}(0) = 1 \mid R_{ij}(0) = 1, R_{1-i,j}(1) = 0\} &= 0.4, & \Pr\{Y_{ij}(0) = 1 \mid R_{ij}(0) = 1, R_{1-i,j}(1) = 1\} &= 0.2. \end{aligned}$$

According to this data generating process, all scenarios have an identical true value of 0.4 for the ATOP. For each scenario, we generate 1,000 pairs of units and calculate the average bounds across 500 simulated datasets. Note that the bounds for the ATOU are (0.259, 0.594), (−0.105, 0.909), and (−0.998, 1) for the scenarios with low, moderate, and high levels of missingness, respectively. The bounds are particularly non-informative under moderate and high missing proportions.

We now report the simulation results concerning the bounds of the ATOP. The no-assumption bounds of the ATOP are (0.066, 0.667), (−0.772, 1), and (−1, 1) for the scenarios with the low, moderate, and high levels of missing data, respectively. As expected, these bounds are wider than those of the ATOU reported above. To narrow these bounds, we conduct sensitivity analysis by varying the sensitivity parameter  $\gamma$ . The left plot of Figure 1 shows the results of the sensitivity

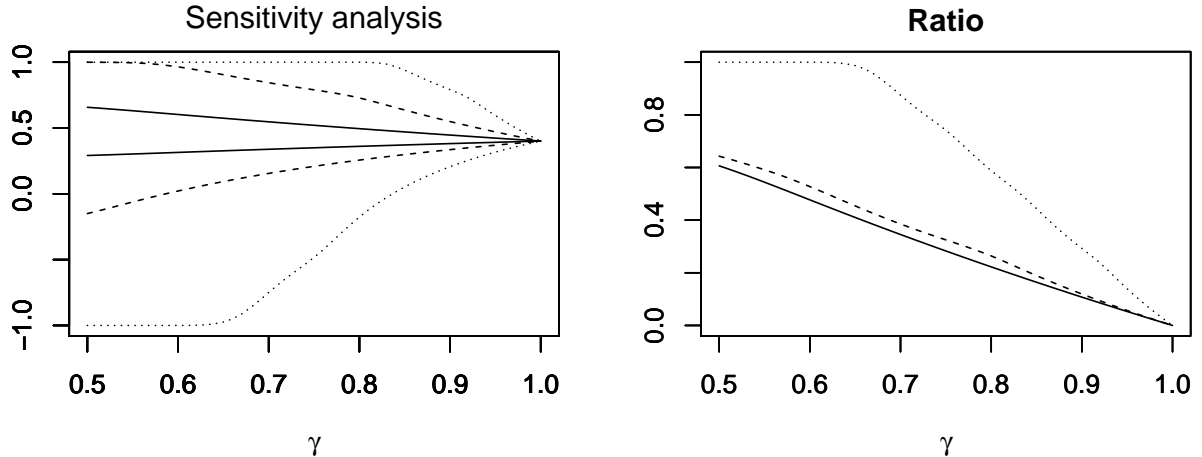


Figure 1: Sensitivity analysis with different levels of missing proportions. In the left plot, the solid, dashed, and dotted lines represent the bounds under the scenarios of the low, moderate, and high levels of missing proportions, respectively. The sensitivity parameter  $\gamma$ , which represents the within-pair similarity of missingness pattern, is shown on the horizontal axis. In the right plot, we plot the ratio of bound widths under different values of the sensitivity parameter relative to the widths of the no-assumption bounds. Again, the solid, dashed, and dotted lines denote the scenarios of the low, moderate, and high levels of missing proportions, respectively.

analysis with different values of  $\gamma$ , whereas the right plot shows the ratio of the bound widths under different values of the sensitivity parameter to the widths of the no-assumption bounds. Under the scenarios of the low (solid lines) and moderate (dashed lines) levels of missing proportions, the bounds become narrower as  $\gamma$  increases from 0.5 to 1. In contrast, the bounds under the scenario of the high level of missing proportion remain unchanged when  $\gamma$  is less than 0.65. Similarly, when the level of missing proportion is moderate, the bounds do not cover zero so long as  $\gamma \geq 0.6$ , whereas when the missing proportion is high, the bounds cover zero unless  $\gamma \geq 0.85$ . Thus, as expected, when the amount of missing data is great, informative bounds require the assumption of greater within-pair similarity.

## 4 Empirical Illustration

In this section, we apply the sensitivity analysis developed above to the data from a matched-pairs randomized experiment that was conducted to evaluate the efficacy of the Mexican universal health insurance program, Seguro Popular (King et al., 2009).

## 4.1 Data and Quantities of Interest

The evaluation study employed the cluster matched-pairs design where within each of 50 pairs of health clusters, the treatment was randomly assigned to one health cluster. Within a treatment cluster, households were encouraged to enroll in Seguro Popular. In contrast, the households of a control cluster did not receive such an encouragement. We analyze the household-level satisfaction of Segro popular as an outcome. Unfortunately, this variable suffers from the truncation-by-death problem because the satisfaction variable is not well defined for those households who did not enroll in the program.

For the sake of empirical illustration, we create pairs at the individual household level. Specifically, we conduct one-to-one nearest neighbor propensity score matching without replacement by pairing each household of one treatment cluster with a similar household from its paired control cluster. We use R package `MatchIt` (Ho et al., 2011) for implementation. For some cluster pairs, the number of the households in a control cluster is different from that in a treated cluster of the same pair. In those cases, we simply remove unmatched households. As a result, we have a total of 14,902 matched household pairs. Although, unlike cluster-level pairs, these household pairs were created retrospectively, we emphasize that Assumption 1 still holds within each pair.

We use the logistic regression model for estimating the propensity score and include assets, age, sex, education, household assets, household average age, household sex, and household education as predictors. To deal with the missing values of the covariates, we include a response indicator, which is equal to 1 if all the covariates are observed and 0 otherwise, as well as the original variable where the missing values are replaced with zero. The latter variable represents the value of the corresponding variable conditional on being observed. The outcome of interest is the satisfaction with Seguro Popular, which ranges from 1 (very good) to 5 (very bad). For the purpose of illustration, we dichotomize the outcome, i.e., we denote  $Y_{ij} = 1$  if the score is less than or equal to 3 and  $Y_{ij} = 0$  otherwise. For those who did not enroll in the program, we do not observe their satisfaction level. As a result, we have a high missing proportion of 79.4%.

## 4.2 Sensitivity Analysis

We begin by computing the no-assumption bounds for the ATOU and ATOP. To account for sample variability, we construct confidence intervals for each set of bounds. Because the lower and upper bounds correspond to the minimum and maximum values for some parameters, we cannot directly use traditional techniques (Andrews, 2000; Imbens and Manski, 2004; Vansteelandt et al., 2006;

Satisfaction with Segro Popular

Naive estimates using available units	
Estimates	−0.023
95% confidence intervals	(−0.038, −0.010)
Naive estimates using available pairs ( $\hat{\tau}_{\text{ATOP}}$ )	
Estimates	−0.075
95% confidence intervals	(−0.118, −0.039)
Average Treatment Effect for Always-observed Units (ATOU)	
Bounds	(−1, 1)
95% confidence intervals	(−1, 1)
Average Treatment Effect for Always-observed Pairs (ATOP)	
Bounds	(−1, 1)
95% confidence intervals	(−1, 1)

Table 1: The naive estimates using available units and pairs for the average treatment effect on the satisfaction with Segro Popular. The no-assumption bounds for the average treatment effect for always-observed units (ATOU) and for always-observed pairs (ATOP) are also shown. Given a high proportion of missing values, these bounds are not informative.

Romano and Shaikh, 2010). Instead, we use the method proposed by Jiang and Ding (2016), which gives the confidence intervals for a partial identification parameter  $\pi$  of the form,

$$\max\{L(1), \dots, L(Q)\} \leq \pi \leq \min\{U(1), \dots, U(R)\},$$

where  $L(\cdot)$  and  $U(\cdot)$  are functions of observed probabilities. To apply this procedure, we transform the bounds in Theorem 3 as,

$$\max_k \left\{ f^{[l,u]}(\omega_1, \Delta_k, \omega_0, \Delta_k) \right\} \leq \text{ATOP} \leq \min_k \left\{ g^{[l,u]}(\omega_1, \Delta_k, \omega_0, \Delta_k) \right\}$$

where  $\Delta_k$  is the  $k$ -th term in the expression of  $\Delta$ . A similar transformation can be applied to the other bounds as well.

Table 1 presents the naive estimates using available units and available pairs ( $\hat{\tau}_{\text{ATOP}}$ ) as well as the no-assumption bounds for the ATOU and ATOP. For each estimate, we also report its corresponding 95% confidence interval. Although the naive estimates are statistically significantly negative, the no-assumption bounds are not informative at all due to the high missing proportion.

Next, we conduct the proposed sensitivity analysis for the ATOP’s with different values of  $\gamma$ . Figure 2 shows the results. The 95% confidence interval for the bounds do not cover zero unless  $\gamma$  is very close to 1. Therefore, unless pairing is done near perfectly, we cannot conclude that the ATOP is statistically significantly negative. However, when pairing is done well, the bounds become much narrower than the no-assumption bounds, giving us some information about the ATOP.

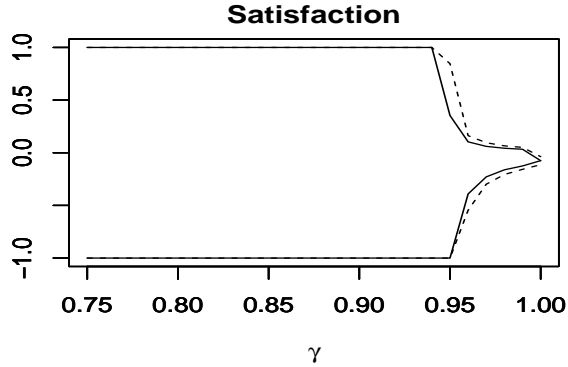


Figure 2: Sensitivity analysis for the ATOP. The solid lines are the bounds and the dashed lines are the confidence intervals.

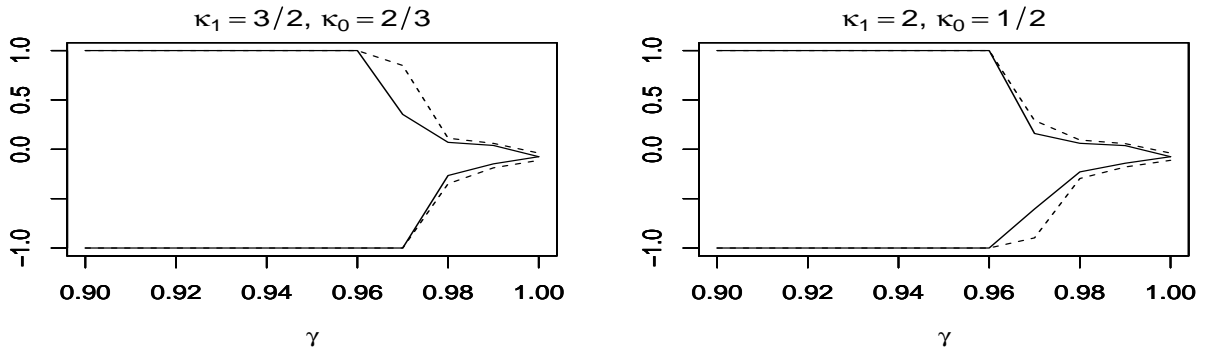


Figure 3: Sensitivity analysis for the ATOP in observational studies. The solid lines are the bounds and the dashed lines are the confidence intervals. When  $\kappa_1 = 3/2$  and  $\kappa_0 = 2/3$ ,  $\Pr\{R_{ij}(1) = 1\} = 0.297$  and  $\Pr\{R_{ij}(0) = 1\} = 0.046$ , and when  $\kappa_1 = 2$  and  $\kappa_0 = 1/2$ ,  $\Pr\{R_{ij}(1) = 1\} = 0.268$  and  $\Pr\{R_{ij}(0) = 1\} = 0.042$ .

Finally, even though this is an experimental study, for the sake of illustration, we conduct the proposed sensitivity analysis for observational studies using Theorem 4. We fix the values of the two sensitivity parameters,  $\kappa_1$  and  $\kappa_0$ , and vary the remaining sensitivity parameter  $\gamma$ . We choose two levels for both  $\kappa_1$  and  $\kappa_0$ :  $(\kappa_1, \kappa_0) = (3/2, 2/3)$  and  $(\kappa_1, \kappa_0) = (2, 1/2)$ . Figure 3 shows the results. As expected, compared with Figure 2, the bounds in Figure 3 are wider. However, as in the case of randomized experiments, the bound widths become substantially narrower as we increase the value of  $\gamma$ . This implies that the within-pair similarity can help narrow the bounds in observational studies as well.

## 5 Concluding Remarks

In this paper, we propose a sensitivity analysis for the matched-pairs design in the presence of missing outcome due to the truncation-by-death problem. We begin with the development of such

methodology for randomized experiments and extend it to observational studies. Design considerations are essential for valid and efficient causal inference, and we believe that sensitivity analyses should also be developed by taking into consideration the features of study designs. This paper illustrates this new direction of research by exploiting the fact that under the matched-pairs design similar units are used to form pairs: if matching is successful, the missingness pattern of one unit should be similar to that of the other unit within the same pair. Future research might explore a similar approach to more general randomized block designs in randomized experiments as well as stratification in observational studies.

## References

- Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.
- Box, G. E., J. S. Hunter, and W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.). New York: Wiley-Interscience.
- Cheng, J. and D. S. Small (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(5), 815–836.
- D’Agostino, Jr., R. B. and D. B. Rubin (2000). Estimating and using propensity scores with partially missing data. *Journal of the American Statistical Association* 95(451), 749–759.
- Dantzig, G. B. and M. N. Thapa (2006). *Linear programming 2: theory and extensions*. Springer Science & Business Media.
- Ding, P., Z. Geng, W. Yan, and X. H. Zhou (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association* 106, 1578–1591.
- Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics* 58(1), 21–29.
- Fukumoto, K. (2015). Missing data under the matched-pair design: A practical guide. Technical report, Presented at the 32nd Annual Summer Meeting of Society for Political Methodology.
- Greevy, R., B. Lu, J. H. Silber, and P. Rosenbaum (2004). Optimal multivariate matching before randomization. *Biostatistics* 5(2), 263–275.



- Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007, Summer). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3), 199–236.
- Ho, D. E., K. Imai, G. King, and E. A. Stuart (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* 42(8), 1–28.
- Horowitz, J. L. and C. F. Manski (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association* 95(449), 77–84.
- Imai, K. (2008a). Sharp bounds on the causal effects in randomized experiments with truncation-by-death. *Statistics & probability letters* 78(2), 144–149.
- Imai, K. (2008b). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in medicine* 27(24), 4857–4873.
- Imai, K., G. King, and C. Nall (2009, February). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation (with discussions). *Statistical Science* 24(1), 29–53.
- Imai, K., G. King, and E. A. Stuart (2008, April). Misunderstandings among experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 171(2), 481–502.
- Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.
- Jiang, Z. and P. Ding (2016). Using the information of missing types and monotonicity to improve partial identification with binary outcomes missing not at random. *arXiv preprint arXiv:1610.01198*.
- King, G., E. Gakidou, K. Imai, J. Lakin, R. T. Moore, C. Nall, N. Ravishankar, M. Vargas, M. M. Téllez-Rojo, J. E. H. Ávila, et al. (2009). Public policy for the poor? a randomised assessment of the mexican universal health insurance programme. *The lancet* 373(9673), 1447–1454.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88, 125–134.

- Martin, D. C., P. Diehr, E. B. Perrin, and T. D. Koepsell (1993). The effect of matching on the power of randomized community intervention studies. *Statistics in Medicine* 12, 329–338.
- Neyman, J. (1923[1990]). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translation of excerpts by D. Dabrowska and T. Speed. *Statistical Science* 6, 462–47.
- Romano, J. P. and A. M. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica* 78(1), 169–211.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688–701.
- Rubin, D. B. (1980). Comment on “Randomization analysis of experimental data: The Fisher randomization test”. *Journal of the American Statistical Association* 75, 591–593.
- Rubin, D. B. (2006). *Matched Sampling for Causal Effects*. Cambridge: Cambridge University Press.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1), 1–21.
- Vansteelandt, S., E. Goetghebeur, M. G. Kenward, and G. Molenberghs (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 953–979.
- Zhang, J. L. and D. B. Rubin (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. *Journal of Educational and Behavioral Statistics* 28, 353–368.

# A Mathematical Appendix

## A.1 Preliminaries

We need the following lemma to simplify the derivation of the bounds.

LEMMA 1 (CHENG AND SMALL (2006)) *Let  $X_0$  be a mixture of two variables  $X_1$  and  $X_2$  with  $X_0 \sim qX_1 + (1 - q)X_2$ ,  $X_i \in [l, u]$ , and a known mixing proportion  $q$ . Then we have*

$$\begin{aligned} \max\left(l, u - \frac{u - \mathbb{E}(X_0)}{q}\right) &\leq \mathbb{E}(X_1) \leq \min\left(u, \frac{\mathbb{E}(X_0) - l}{q} + l\right), \\ \max\left(l, u - \frac{u - \mathbb{E}(X_0)}{1 - q}\right) &\leq \mathbb{E}(X_2) \leq \min\left(u, \frac{\mathbb{E}(X_0) - l}{1 - q} + l\right). \end{aligned}$$

We also introduce a general linear programming problem:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^\top \mathbf{q}, \\ \text{Subject to} \quad & \mathbf{A}\mathbf{q} = \mathbf{b}, \quad \mathbf{q} \geq \mathbf{0}. \end{aligned} \tag{9}$$

The dual problem is

$$\begin{aligned} \text{Maximize} \quad & \mathbf{b}^\top \mathbf{p}, \\ \text{Subject to} \quad & \mathbf{A}^\top \mathbf{p} \leq \mathbf{c}. \end{aligned}$$

The set  $\{\mathbf{p} \mid \mathbf{A}^\top \mathbf{p} \leq \mathbf{c}\}$  is a polyhedron and  $\mathbf{b}^\top \mathbf{p}$  reaches its maximum at the vertexes of the polyhedron. According to Dantzig and Thapa (2006), an optimal solution to the dual problem corresponds to an optimal solution to the original problem. Therefore, the solution to the original is

$$L = \max(\mathbf{b}^\top \mathbf{p}_1^*, \dots, \mathbf{b}^\top \mathbf{p}_K^*), \tag{10}$$

where  $\mathbf{p}_k^*$  is the vertex of the polyhedron  $\{\mathbf{p} \mid \mathbf{A}^\top \mathbf{p} \leq \mathbf{c}\}$ , and  $K$  is the number of the vertexes.

To derive the bounds, we recall some notation in the main text and introduce some new notation.

Denote

$$\begin{aligned} \pi &= P(R_{1j} = R_{2j} = 1) \\ \alpha_t &= P(R_{1j} = 1 \mid T_{1j} = t) = P(R_{2j} = 1 \mid T_{2j} = t) \\ \psi_t &= \mathbb{E}(Y_{1j} \mid R_{1j} = 1, T_{1j} = t) = \mathbb{E}(Y_{2j} \mid R_{2j} = 1, T_{2j} = t) \\ \omega_t &= \mathbb{E}(Y_{1j} \mid T_{1j} = t, R_{1j} = R_{2j} = 1) = \mathbb{E}(Y_{2j} \mid T_{2j} = t, R_{1j} = R_{2j} = 1), \\ \phi &= P\{R_1(1) = R_2(0) = 1 \mid R_1(0) = R_2(1) = 1\} \end{aligned}$$

$$\begin{aligned}
&= P\{R_1(0) = R_2(1) = 1 \mid R_1(1) = R_2(0) = 1\}, \\
\theta_t &= E\{Y(t) \mid R_1(1) = R_2(0) = R_2(1) = R_1(0) = 1\}, \\
\lambda_1 &= P\{R(0) = 1 \mid R(1) = 1\}, \quad \lambda_0 = P\{R(1) = 1 \mid R(0) = 1\}.
\end{aligned}$$

## A.2 Proof of Theorem 1

By the law of total probability, we have

$$\delta_t \phi + E\{Y(t) \mid R_1(1)R_2(0) = 1, R_2(1)R_1(0) = 0\}(1 - \phi) = \omega_t.$$

Then according to Lemma 1, we have

$$\max\left(l, u - \frac{u - \omega_t}{\phi}\right) \leq \theta_t \leq \min\left(u, \frac{\omega_t - l}{\phi} + l\right).$$

Because  $\text{ATOP} = \theta_1 - \theta_0$ , we can express the bounds for  $\text{ATOP}$  in terms of  $\phi$ :

$$\text{OPLB}^{[l,u]}(\phi) = \max\left(l, u - \frac{u - \omega_1}{\phi}\right) - \min\left(u, \frac{\omega_0 - l}{\phi} + l\right), \quad (11)$$

$$\text{OPUB}^{[l,u]}(\phi) = \min\left(u, \frac{\omega_1 - l}{\phi} + l\right) - \max\left(l, u - \frac{u - \omega_0}{\phi}\right). \quad (12)$$

From the law of total probability, we have

$$\phi\pi + (1 - \pi)P(R_2(1)R_1(0) = 1 \mid R_1(1)R_2(0) = 0) = \pi.$$

According to Lemma 1, we have

$$\max\left(0, \frac{2\pi - 1}{\pi}\right) \leq \phi \leq 1.$$

From (11) and (12),  $\text{OPLB}^{[l,u]}(\phi)$  is non-decreasing and  $\text{OPUB}^{[l,u]}(\phi)$  is non-increasing in  $\phi$ , respectively. Thus, we can obtain the bounds for  $\text{ATOP}$  by setting  $\phi$  to its lower bound:

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi}), g^{[l,u]}(\omega_1, 2 - \frac{1}{\pi}, \omega_0, 2 - \frac{1}{\pi})] & \text{if } \pi > \frac{1}{2} \\ [l - u, u - l] & \text{if } \pi \leq \frac{1}{2} \end{cases}$$

where

$$\begin{aligned}
f^{[l,u]}(x_1, y_1, x_0, y_0) &= \max\left\{l, u - \frac{u - x_1}{y_1}\right\} - \min\left\{u, \frac{x_0 - l}{y_0} + l\right\}, \\
g^{[l,u]}(x_1, y_1, x_0, y_0) &= \min\left\{u, \frac{x_1 - l}{y_1} + l\right\} - \max\left\{l, u - \frac{u - x_0}{y_0}\right\}.
\end{aligned}$$

□

### A.3 Proof of Theorem 2

Under Assumption 1, we have  $P\{R(t) = 1\} = \alpha_1$ . Thus,  $\lambda_1\alpha_1 = \lambda_0\alpha_0$ . By the law of total probability, we can obtain

$$\Delta\lambda_1 + E\{Y(1) \mid R(1) = 1, R(0) = 0\}(1 - \lambda_1) = E\{Y(1) \mid R(1) = 1\} = \psi_1.$$

Then according to Lemma 1, we have

$$\max\left(l, u - \frac{u - \psi_1}{\lambda_1}\right) \leq \Delta \leq \min\left(u, \frac{\psi_1 - l}{\lambda_1} + l\right).$$

Similarly, we have

$$\max\left(l, u - \frac{u - \psi_0}{\lambda_0}\right) \leq \delta_0 \leq \min\left(u, \frac{\psi_0 - l}{\lambda_0} + l\right).$$

Thus, we can write the bounds in terms of  $\lambda_1$ .

$$\begin{aligned} \text{OUUB}^{[l,u]}(\lambda_1) &= \min\left(u, \frac{\psi_1 - l}{\lambda_1} + l\right) - \max\left(l, u - \frac{\alpha_0(u - \psi_0)}{\alpha_1\lambda_1}\right), \\ \text{OULB}^{[l,u]}(\lambda_1) &= \max\left(l, u - \frac{u - \psi_1}{\lambda_1}\right) - \min\left(u, \frac{\alpha_0(\psi_0 - l)}{\alpha_1\lambda_1} + l\right). \end{aligned}$$

By the law of total probability,  $\lambda_1\alpha_1 + P\{R(0) = 1 \mid R(1) = 0\}(1 - \alpha_1) = \alpha_0$ . According to Lemma 1,

$$\lambda_1 \geq \max\left(0, \frac{\alpha_1 + \alpha_0 - 1}{\alpha_1}\right).$$

We can obtain the bounds by maximizing  $\text{OUUB}^{[l,u]}(\lambda_1)$  and minimizing  $\text{OULB}^{[l,u]}(\lambda_1)$  over  $\lambda_1$ . Because  $\text{OUUB}^{[l,u]}(\lambda_1)$  is decreasing in  $\lambda_1$  and  $\text{OULB}^{[l,u]}(\lambda_1)$  is increasing in  $\lambda_1$ , we can obtain the bounds for ATOP by setting  $\lambda_1$  to its lower bound:

$$\text{ATOU} \in \begin{cases} \left[ f^{[l,u]}(\psi_1, \frac{\Gamma}{\alpha_1}, \psi_0, \frac{\Gamma}{\alpha_0}), g^{[l,u]}(\psi_1, \frac{\Gamma}{\alpha_1}, \psi_0, \frac{\Gamma}{\alpha_0}) \right] & \text{if } \Gamma > 0 \\ [l - u, u - l] & \text{if } \Gamma \leq 0 \end{cases}$$

where  $\Gamma = \alpha_1 + \alpha_0 - 1$ . □

For the proof of Theorem 3, we use linear programming. Denote  $\zeta_{ijkl} = P\{R_1(1) = i, R_1(0) = j, R_2(1) = k, R_2(0) = l\}$ . Because the labels are randomly assigned, some of  $\zeta_{ijkl}$  are equal. To make it clear for linear programming, we denote

$$\begin{aligned} x_1 &= \zeta_{1111}, x_2 = \zeta_{1101} = \zeta_{0111}, x_3 = \zeta_{1011} = \zeta_{1110}, x_4 = \zeta_{1001} = \zeta_{0110}, x_5 = \zeta_{1010}, \\ x_6 &= \zeta_{0101}, x_7 = \zeta_{1100} = \zeta_{0011}, x_8 = \zeta_{0001} = \zeta_{0100}, x_9 = \zeta_{1000} = \zeta_{0010}, x_{10} = \zeta_{0000}. \end{aligned}$$

We specify the constraint imposed by Assumptions 1 and 2. By the law of total probability, we have the following constraints.

$$\begin{aligned}
x_1 + x_2 + x_3 + x_4 &= \pi, \\
x_1 + 2x_3 + x_5 &= P\{R_2(1)R_1(1) = 1\} \geq \gamma\alpha_1, \\
x_1 + 2x_2 + x_6 &= P\{R_2(0)R_1(0) = 1\} \geq \gamma\alpha_0, \\
x_6 + 2x_8 + x_{10} &= P\{R_2(1) = R_1(1) = 0\} \geq \gamma(1 - \alpha_1), \\
x_5 + 2x_9 + x_{10} &= P\{R_2(0) = R_1(0) = 0\} \geq \gamma(1 - \alpha_0), \\
x_3 + x_5 + x_7 + x_9 &= P\{R_1(1) = 1, R_2(0) = 0\} = \alpha_1 - \pi, \\
x_2 + x_6 + x_7 + x_8 &= P\{R_1(1) = 0, R_2(0) = 1\} = \alpha_0 - \pi, \\
x_4 + x_8 + x_9 + x_{10} &= P\{R_1(1) = 0, R_2(0) = 1\} = 1 + \pi - \alpha_1 - \alpha_0, \\
x_i &\geq 0 \quad \text{for } i = 1, \dots, 10.
\end{aligned} \tag{13}$$

We can introduce additional parameter to transform the constraints from inequalities in (13) to equalities:

$$\begin{aligned}
x_1 + x_2 + x_3 + x_4 &= \pi, \\
x_1 + 2x_3 + x_5 - x_{11} &= \gamma\alpha_1, \\
x_1 + 2x_2 + x_6 - x_{12} &= \gamma\alpha_0, \\
x_6 + 2x_8 + x_{10} - x_{13} &= \gamma(1 - \alpha_1), \\
x_5 + 2x_9 + x_{10} - x_{14} &= \gamma(1 - \alpha_0), \\
x_3 + x_5 + x_7 + x_9 &= \alpha_1 - \pi, \\
x_2 + x_6 + x_7 + x_8 &= \alpha_0 - \pi, \\
x_4 + x_8 + x_9 + x_{10} &= 1 + \pi - \alpha_1 - \alpha_0, \\
x_i &\geq 0 \quad \text{for } i = 1, \dots, 14.
\end{aligned} \tag{14}$$

We will use these constraints to derive the bounds for ATOP.

#### A.4 Proof of Theorem 3

From equations (11) and (12),  $\text{OPLB}^{[l,u]}(\phi)$  is non-decreasing and  $\text{OPUB}^{[l,u]}(\phi)$  is non-increasing in  $\phi$ , respectively. Thus, we need only to obtain the sharp lower bound of  $\phi$ . By definition  $\phi = x_1/\pi$ . Therefore, we need only to minimize  $x_1$  subject to equation (14). By solving the dual problem, we have

$$x_1 \geq \max_{i=0,1} \{2\pi - 1 + \gamma(1 - \alpha_i), \pi - (1 - \gamma)(\alpha_1 + \alpha_0), 2\pi - (2 - \gamma)\alpha_i, \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0)\},$$

$$\pi - (1 - \gamma)(1 - |\alpha_1 - \alpha_0|).$$

Thus, we can obtain the bounds for ATOP:

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta), g^{[l,u]}(\omega_1, \Delta, \omega_0, \Delta)] & \text{if } \Delta > 0 \\ [l - u, u - l] & \text{if } \Delta \leq 0 \end{cases}$$

where  $\Delta = \max_{i=0,1} \{2\pi - 1 + \gamma(1 - \alpha_i), \pi - (1 - \gamma)(\alpha_1 + \alpha_0), 2\pi - (2 - \gamma)\alpha_i, \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0), \pi - (1 - \gamma)(1 - |\alpha_1 - \alpha_0|)\} / \pi$ .

If  $\alpha_1 \geq 1/2$  and  $\alpha_0 \geq 1/2$ , then

$$\begin{aligned} & 2\pi - 1 + \gamma(1 - \alpha_i) - \{2\pi - (2 - \gamma)\alpha_i\} = (1 - \gamma)(2\alpha_i - 1) \geq 0, \\ & \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0) - \{\pi - (1 - \gamma)(1 - |\alpha_1 - \alpha_0|)\}, \\ & = (1 - \gamma)(\alpha_1 + \alpha_0 - 1 - |\alpha_1 - \alpha_0|) \geq 0. \end{aligned}$$

Therefore, in this case, we can simplify the expression of  $\Delta$  as  $\Delta = \max_{i=0,1} \{2\pi - 1 + \gamma(1 - \alpha_i), \pi - (1 - \gamma)(2 - \alpha_1 - \alpha_0)\} / \pi$ .  $\square$

### A.5 Proof of Theorem 4

Denote  $\alpha_t^* = \Pr\{R(t) = 1\}$ . According to Assumption 4, we have  $\omega_t = E\{Y_1(t) \mid R_1(t) = 1, R_2(1 - t) = 1\}$  and

$$\alpha_1^* = \frac{\alpha_1(1 + \kappa_1)}{2\kappa_1}, \quad \alpha_0^* = \frac{\alpha_0(1 + \kappa_0)}{2}.$$

If we view  $\pi^*$  and  $\alpha_t^*$  as known, then following the proof of Theorem 3, we have

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, \Delta(\pi^*), \omega_0, \Delta(\pi^*)), g^{[l,u]}(\omega_1, \Delta(\pi^*), \omega_0, \Delta(\pi^*))] & \text{if } \Delta(\pi^*) > 0 \\ [l - u, u - l] & \text{if } \Delta(\pi^*) \leq 0 \end{cases}$$

where  $\Delta(\pi^*) = \max_{i \in \{0,1\}} \{2\pi^* - 1 + \gamma(1 - \alpha_i^*), \pi - (1 - \gamma)(\alpha_1^* + \alpha_0^*), 2\pi - (2 - \gamma)\alpha_i^*, \pi - (1 - \gamma)(2 - \alpha_1^* - \alpha_0^*), \pi - (1 - \gamma)(1 - |\alpha_1^* - \alpha_0^*|)\} / \pi$ . Because  $f^{[l,u]}(\omega_1, \Delta(\pi^*), \omega_0, \Delta(\pi^*))$  is decreasing in  $\Delta(\pi^*)$ ,  $g^{[l,u]}(\omega_1, \Delta(\pi^*), \omega_0, \Delta(\pi^*))$  is increasing in  $\Delta(\pi^*)$  and  $\Delta(\pi^*)$  is decreasing in  $\pi^*$ , we can obtain the bounds for ATOP by taking  $\pi^*$  to its lower bound,  $\pi/2$ :

$$\text{ATOP} \in \begin{cases} [f^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*), g^{[l,u]}(\omega_1, \Delta^*, \omega_0, \Delta^*)] & \text{if } \Delta^* > 0 \\ [l - u, u - l] & \text{if } \Delta^* \leq 0 \end{cases}$$

where  $\Delta^* = \max_{i \in \{0,1\}} \{2\pi - 2 + 2\gamma(1 - \alpha_i^*), \pi - 2(1 - \gamma)(\alpha_1^* + \alpha_0^*), 2\pi - 2(2 - \gamma)\alpha_i^*, \pi - 2(1 - \gamma)(2 - \alpha_1^* - \alpha_0^*), \pi - 2(1 - \gamma)(1 - |\alpha_1^* - \alpha_0^*|)\} / \pi$ .  $\square$