Article

Journal of Educational and Behavioral Statistics
Vol. XX, No. X, pp. 1–28
DOI: 10.3102/1076998618776435

Treatment Effects on Ordinal Outcomes: Causal Estimands and Sharp Bounds

Jiannan Lu
Microsoft Corporation

Peng Ding
University of California-Berkeley

Tirthankar Dasgupta
Rutgers University

Assessing the causal effects of interventions on ordinal outcomes is an important objective of many educational and behavioral studies. Under the potential outcomes framework, we can define causal effects as comparisons between the potential outcomes under treatment and control. However, unfortunately, the average causal effect, often the parameter of interest, is difficult to interpret for ordinal outcomes. To address this challenge, we propose to use two causal parameters, which are defined as the probabilities that the treatment is beneficial and strictly beneficial for the experimental units. However, although well-defined for any outcomes and of particular interest for ordinal outcomes, the two aforementioned parameters depend on the association between the potential outcomes and are therefore not identifiable from the observed data without additional assumptions. Echoing recent advances in the econometrics and biostatistics literature, we present the sharp bounds of the aforementioned causal parameters for ordinal outcomes, under fixed marginal distributions of the potential outcomes. Because the causal estimands and their corresponding sharp bounds are based on the potential outcomes themselves, the proposed framework can be flexibly incorporated into any chosen models of the potential outcomes and is directly applicable to randomized experiments, unconfounded observational studies, and randomized experiments with noncompliance. We illustrate our methodology via numerical examples and three real-life applications related to educational and behavioral research.

Keywords: *linear programming; monotonicity; noncompliance; partial identification; potential outcome; stochastic dominance*

1. Introduction

In educational, behavioral, and public health research, a scenario frequently encountered is evaluating causal effects of interventions on ordinal (i.e., ordered

categorical) outcomes. For example, Oenema, Brug, and Lechner (2001) conducted a randomized controlled trial to access whether Web-based nutrition education changed personal awareness and intentions (e.g., negative, neutral, or positive attitudes) toward healthier diets. Hoff (2009) analyzed a data set from the 1994 General Social Survey (Smith, Marsden, Hout, & Kim, 2013), aiming to study whether the fact that parents possessing college or higher degrees affected their offspring's education level (from "less than high school" to "graduate degree"). Praet and Desoete (2014) investigated the effect of computer-aided programs on young children's proficiency in arithmetic (e.g., 0–10 scaled scores in reading, writing, and counting). To draw scientifically meaningful conclusions from such studies, it is imperative that we employ an interpretable and robust methodology for defining and inferring causal effects.

The potential outcomes framework (Neyman, 1923; Rubin, 1974) permits defining causal effects as comparisons between the potential outcomes under treatment and control. The average causal effect, generally the parameter of interest ever since the seminal work of Neyman (1923), may not be applicable to ordinal outcomes because average outcomes themselves are not well-defined substantively (although sometimes they can be well-defined mathematically), except when there are meaningful distances between outcomes (e.g., standard test scores). For example, it is difficult to interpret the "average" of "high school" and "PhD," or compare it to the average of "bachelor" and "master." Nevertheless, ordinal outcomes appear rather frequently in applied research, and the generalized linear model literature (cf. Agresti, 2010) has discussed them extensively. However, although the model parameters of the generalized linear models are useful summaries of the data, they are often not direct measures of the causal effects of interest (Freedman, 2008). More importantly, statistical inference often requires correctly specified models, and when the generalized linear model assumptions are violated, the interpretations of the parameters become obscure. Mainly focused on the classic average causal effect (and its variants), the existing causal inference literature does not thoroughly investigate ordinal outcomes. Exceptions include Rosenbaum (2001), who discussed causal inference for ordinal outcomes under the monotonicity assumption that the treatment is beneficial for all units. Cheng (2009), Agresti (2010), and Agresti and Kateri (2017) discussed various causal parameters under the assumption of independent potential outcomes. Volfovsky, Airoldi, and Rubin (2015) exploited a Bayesian strategy, requiring a full parametric model on the joint values of the potential outcomes. Diaz, Colantuoni, and Rosenblum (2016) proposed to use a causal parameter that did not rely on the assumption of the proportional odds model for ordinal outcomes.

Realizing the conceptual and theoretical gaps in this important topic, in this article we propose to use two causal parameters for ordinal outcomes, measuring the probabilities that the treatment is beneficial and strictly beneficial for the experimental units. The two parameters play important roles in decision and

policymaking for randomized evaluations with ordinal outcomes. However, because the two causal parameters depend on the association between the treatment and control potential outcomes, they are generally not identifiable from the observed data. Instead of imposing assumptions about the underlying distributions of, or the association between, the potential outcomes, we adopt the partial identification strategy (cf. Manski, 2003; Richardson, Hudgens, Gilbert, & Fine, 2014) and sharply bound the parameters by using the marginal distributions of the potential outcomes. We acknowledge concurrent work by Huang, Fang, Hanley, and Rosenblum (2017) who numerically calculated the sharp bounds of the parameters and provided their consistent estimators, allowing for potentially complex support restrictions on the marginal distributions of the potential outcomes. Compared to Huang et al. (2017), one main distinction of our work is that we focus on the identification perspective. To be specific, echoing several relevant discussions in the discrete mathematics (Williamson & Downs, 1990) and econometrics (e.g., Fan & Park, 2009; Kim, 2014; Manski, 1997; Manski & Pepper, 2009) literature, we present closed-form expressions for the sharp bounds of the causal parameters.

We believe that the mathematical practice of deriving the closed-form expressions for the sharp bounds has a 2-fold benefit. From a theoretical perspective, the closed-form expressions enable us to study when we can identify the causal parameters, that is, the lower and upper bounds collapse. At least in the context of ordinal outcomes, we believe this is a unique contribution to the existing literature. From a more practical perspective, because these bounds are defined by the potential outcomes themselves, they can be incorporated flexibly into any chosen models of the potential outcomes. Furthermore, they are directly applicable to randomized experiments, unconfounded observational studies, and randomized experiments with noncompliance. In randomized experiments, we can identify the bounds immediately, and additionally, sharpen the bounds by exploiting covariate information under certain modeling assumptions. In observational studies, if the treatment assignment is unconfounded given the observed covariates, we can identify the bounds, for example, by the propensity score weighting (Hirano, Imbens, & Ridder, 2003; Rosenbaum & Rubin, 1983). Furthermore, we extend the theory to accommodate noncompliance, which often arises in practical randomized evaluations.

This article proceeds as follows. Section 2 introduces the potential outcomes framework for causal inference for ordinal outcomes, and proposes two causal parameters that are natural measures of causal effects and are of practical importance. Section 3 presents the sharp bounds of the proposed causal parameters. Section 4 generalizes the bounds to noncompliance. Section 5 discusses statistical inference of the bounds. Sections 6 and 7 present numerical and real examples to illustrate the theoretical results. We conclude in Section 8 and give all the proofs, technical, and computational details in the Online Supplemental Material.

2. Causal Inference for Ordinal Outcomes

2.1. Potential Outcomes

We consider a study with N units, a binary treatment, and an ordinal outcome with J categories labeled as $0, \ldots, J-1$, where 0 and $J-1$ represent the worst and best categories. Under the stable unit treatment value assumption (Rubin, 1980) that there is only one version of the treatment and no interference among the units, we define the pair $\{Y_i(1), Y_i(0)\}$ as the potential outcomes of the ith unit under treatment and control, respectively. Let

$$p_{kl} = \text{pr}\{Y_i(1) = k, Y_i(0) = l\} \quad (k, l = 0, \ldots, J-1),$$

denote the proportion or probability of units whose potential outcome is k under treatment and l under control. The probability notation "pr$(\cdot)$" is either for a finite population of N units or for a super population, depending on the question of interest. The probability matrix $\boldsymbol{P} = (p_{kl})_{0 \le k,l \le J-1}$ summarizes the (unconditional) joint distribution of the potential outcomes. We denote the row and column sums of $\boldsymbol{P}$ by

$$p_{k+} = \sum_{l'=0}^{J-1} p_{kl'}, \quad p_{+l} = \sum_{k'=0}^{J-1} p_{k'l} \quad (k, l = 0, 1, \ldots, J-1).$$

The vectors $\boldsymbol{p}_1 = (p_{0+}, \ldots, p_{J-1,+})^{\mathrm{T}}$ and $\boldsymbol{p}_0 = (p_{+0}, \ldots, p_{+,J-1})^{\mathrm{T}}$ characterize the marginal distributions of the potential outcomes under treatment and control, respectively. By definition, the following constraints must hold:

$$\sum_{k=0}^{J-1} p_{k+} = 1, \quad \sum_{l=0}^{J-1} p_{+l} = 1, \quad \sum_{k=0}^{J-1}\sum_{l=0}^{J-1} p_{kl} = 1.$$

2.2. Causal Parameters for Ordinal Outcomes

We discuss the existing causal parameters for ordinal outcomes and the motivation behind proposing new ones. Any causal parameter is a function of the probability matrix $\boldsymbol{P}$. Unfortunately, the average causal effect is difficult to interpret for ordinal outcomes. Instead, we can use the distributional causal effects (cf. Ju & Geng, 2010):

$$\Delta_j = \text{pr}\{Y_i(1) \ge j\} - \text{pr}\{Y_i(0) \ge j\} = \sum_{k \ge j} p_{k+} - \sum_{l \ge j} p_{+l} \quad (j = 0, \ldots, J-1), \qquad (1)$$

to measure the difference between the marginal distributions of potential outcomes at different levels of j. Although distributional causal effects are standard and important measures for ordinal outcomes in practice, it is sometimes difficult to decide whether the treatment or the control is preferable unless they have the same sign for all j. In the presence of heterogeneous distributional treatment effects for different levels of j, we may use $\sum_{j=1}^{J-1} \omega_j \Delta_j$ to measure the treatment

effect, but such a measure depends crucially on the weights ω_j's. We illustrate this point by using the following numerical example.

Example 1: Let $\boldsymbol{p}_1 = (1/5, 3/5, 1/5)^T$ and $\boldsymbol{p}_0 = (2/5, 1/5, 2/5)^T$, with $\Delta_0 = 0$, $\Delta_1 = 1/5$, and $\Delta_2 = -1/5$. The treatment is beneficial at Level 1, but not at Level 2. In this case, distributional causal effects do not provide straightforward guidance for decision-making.

When $\Delta_j \geq 0$ for all j, $Y(1)$ stochastically dominates $Y(0)$. When this pattern appears in real data applications, practitioners often fit a proportional odds model (Agresti, 2010) and summarize the overall effectiveness of the treatment by a single odds ratio parameter. Although such summary parameter may be useful in certain cases, its causal interpretation is unclear. Moreover, when the data do not present the stochastic dominance pattern as in Example 1, summarizing the treatment effect by the single odds ratio parameter of a wrong model often gives misleading conclusions.

Volfovsky et al. (2015) studied the conditional medians:

$$m_j = \text{med}\{Y_i(1)|Y_i(0) = j\} \quad (j = 0, \ldots, J-1), \tag{2}$$

which is a set containing all values of k, such that $\sum_{k'=0}^{k} p_{k'j} \geq p_{+j}/2$ and $\sum_{k'=k}^{J-1} p_{k'j} \geq p_{+j}/2$. By definition, the conditional medians may not be unique, and they are only well-defined for j with $p_{+j} > 0$. Moreover, they are not direct measures of the treatment effect itself.

We propose to use two causal parameters that measure the probabilities that the treatment is beneficial and strictly beneficial for the experimental units:

$$\tau = \text{pr}\{Y_i(1) \geq Y_i(0)\} = \sum_{k \geq l} \sum p_{kl}, \quad \eta = \text{pr}\{Y_i(1) > Y_i(0)\} = \sum_{k > l} \sum p_{kl}. \tag{3}$$

The causal parameters τ and η are measures of causal effects that are well-defined for any types of outcomes and of particular interest to ordinal outcomes. To be more specific, they can complement the distributional causal effects and provide more information about what would happen under treatment versus control for an ordinal outcome. Similar causal measures appeared in biomedical (Demidenko, 2016; Gadbury & Iyer, 2000; Huang, Fang, Hanley, & Rosenblum, 2017; Newcombe, 2006a, 2006b; Zhou, 2008) and social sciences (Djebbari & Smith, 2008; Fan & Park, 2010; Fan, Sherman, & Shum, 2014; Heckman, Smith, & Clements, 1997). In practice, we suggest using the pair (τ, η) as measures of causal effects on ordinal outcomes. For example, if the sharp null holds, that is, $Y_i(1) = Y_i(0)$ for all units i, then $\tau = 1$ and $\eta = 0$. In this case, using only τ may be misleading. Nevertheless, we argue that the parameter τ is as important as η. Because $1 - \tau = \text{pr}\{Y_i(0) > Y_i(1)\}$, the value of τ determines the probability that the control is strictly beneficial for the experimental units. Due to the symmetry of treatment and control labels, τ and η are equally useful for real data analysis.

We use the following numerical example to show the values of m_j, τ, and η.

Example 2: Consider the following probability matrix:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1/6 & 1/6 \\ 0 & 1/6 & 0 \\ 0 & 1/3 & 1/6 \end{pmatrix}.$$

In this case, m_0 is not well-defined because $p_{k0} = 0$ for all k, m_1 is 1, and $m_2 = \{0, 1, 2\}$ by the definition of the conditional median in Equation 2. However, we have $\tau = 2/3$ and $\eta = 1/3$, that is, two thirds of the population benefit from the treatment and one third strictly benefit.

The causal parameters τ and η in Equation 3 are well-defined for both finite populations and super populations. They are functions of the potential outcomes, which distinguishe them from the parameters in superpopulation models. When the models are misspecified, the interpretations of the corresponding model parameters are often obscure. We have already discussed this issue for the proportional odds model. Our causal parameters τ and η are closely related to the relative treatment effect $\alpha = \text{pr}\{Y_i(1) > Y_i(0)\} - \text{pr}\{Y_i(1) < Y_i(0)\}$ previously studied under the assumption of independent potential outcomes (Agresti, 2010). This relative treatment effect α and the causal parameters we proposed have a simple algebraic relationship, that is, $\alpha = \tau + \eta - 1$. Therefore, our newly proposed causal parameters τ and η determine α. Furthermore, these causal parameters are also related to the notation of "probability of causation" (Pearl, 2009) because their direct interpretations are the probabilities or proportions that the treatment affects the outcome on the individual level. It is for these reasons that we advocate using τ and η as causal effect measures for ordinal outcomes.

3. Sharp Bounds on the Proposed Causal Estimands for Ordinal Outcomes

3.1. Closed-Form Expressions of Sharp Bounds

The definitions of τ and η involve the association between the treatment and control potential outcomes. Because we can never jointly measure the potential outcomes, the observed data do not provide full information about their association, rendering the causal parameters τ and η not identifiable. To partially circumvent this difficulty, we focus on the sharp bounds of τ and η, which are the minimal and maximal values of τ and η under the condition that the probability matrix $\boldsymbol{P} = (p_{kl})_{0 \le k,l \le J-1}$ is well-defined, as well as the constraints of the marginal distributions. In other words, the following needs to hold:

$$\sum_{l'=0}^{J-1} p_{kl'} = p_{k+}, \quad \sum_{k'=0}^{J-1} p_{k'l} = p_{+l}, \quad p_{kl} \ge 0 \quad (k, l = 0, \ldots, J-1). \tag{4}$$

The sharp bounds depend only on the marginal distributions of the potential outcomes. Deriving the bounds is equivalent to solving linear programming

problems, because the objective functions in Equation 3 and the constraints in Equation 4 are all linear. Previous literature (Huang et al., 2017) used a numerical method to solve the linear programming problem for η. Fortunately, as pointed out by several researchers (Fan and Park, 2009; Williamson and Downs, 1990), we can derive closed-form solutions of the above linear programming problems, for both τ and η. We first present the sharp bounds of τ, which is the foundation for the remaining of this article.

Proposition 1: The sharp lower and upper bounds of τ are

$$\tau_L = \max_{0\le j\le J-1}(p_{+j}+\Delta_j), \quad \tau_U = 1+\min_{0\le j\le J-1}\Delta_j. \tag{5}$$

The bounds in Equation 5 resemble Fan and Park's (2010) parallel results for continuous outcomes, where the maximum and minimum operators are replaced by supremum and infimum, respectively. As a straightforward validity check, the inequalities $0 \le \tau_L \le \tau_U \le 1$ always hold regardless of the marginal distributions of the potential outcomes. In particular, by definition in Equation 5, $\tau_L \ge p_{+0}+\Delta_0 = p_{+0}$ and $\tau_U \le 1+\Delta_0 = 1$. Moreover, the bounds in Proposition 1 are closely related to the distributional causal effects in Equation 1, and therefore, we can interpret them as the conservative and optimistic estimates of the probability that the treatment is beneficial to the outcome. Furthermore, the following corollary demonstrates that the sharp upper bound τ_U is related to the *stochastic dominance* assumption, that is, $\Delta_j \ge 0$ for all j.

Corollary 1: The causal parameter $\tau_U = 1$, if and only if the marginal probabilities $\boldsymbol{p}_1$ and $\boldsymbol{p}_0$ satisfy the stochastic dominance assumption.

The above corollary implies that for any marginal probabilities satisfying the stochastic dominance assumption, there exists a lower triangular probability matrix $\boldsymbol{P}$ that corresponds to a population satisfying the monotonicity assumption, that is, $Y_i(1) \ge Y_i(0)$ for all i. Strassen (1965) and Rosenbaum (2001) demonstrated this result, and Proposition 1 extends the previous result without imposing the stochastic dominance assumption. Moreover, Proposition 1 also justifies the use of $\min_{0\le j\le J-1}\Delta_j$ as a measure of the deviation from the stochastic dominance assumption (Scharfstein, Manski, & Anthony, 2004).

To bound η, realizing that $\eta = 1-\text{pr}\{Y_i(0) \ge Y_i(1)\}$, we can directly derive the sharp bounds for $\text{pr}\{Y_i(0) \ge Y_i(1)\}$ by switching the treatment and control labels and applying Equation 5.

Proposition 2: The sharp lower and upper bounds of η are

$$\eta_L = \max_{0\le j\le J-1}\Delta_j, \quad \eta_U = 1+\min_{0\le j\le J-1}(\Delta_j - p_{j+}). \tag{6}$$

Similar to the sharp bounds for τ in Equation 5, the inequalities $0 \le \eta_L \le \eta_U \le 1$ always hold. The bounds in Equations 5 and 6 resemble parallel results in the econometrics literature (Fan & Park, 2009, 2010; Manski, 1997; Manski & Pepper, 2000), which largely focused on continuous outcomes. In fact, deriving the sharp bounds of τ and η is related to a classical probability problem posed by

A. N. Kolmogorov (cf. Nelsen, 2006): How to bound the distribution of the sum (or difference) of two random variables with fixed marginal distributions? For continuous outcomes, because $\delta = Y(1) - Y(0)$ is well-defined, our causal parameters τ and η are determined by the distribution of the causal effect δ, the difference between the treatment and control potential outcomes. Indeed, sharp bounds on the distribution of δ have been obtained by Makarov (1982), Rüschendorf (1982), Frank, Nelsen, and Schweizer (1987), and Williamson and Downs (1990), and recently reviewed by Fan, Sherman, and Shum (2014). For ordinal outcomes, although mathematically valid, the interpretation of $Y(1) - Y(0)$ becomes more challenging, at least in many scenarios. For example, in the context of education, it is difficult to define the "difference" of PhD and master. In behavioral research, it is unclear how to compare the improvement from "negative" to "neutral" and from "neutral" to "positive."

Motivated by the above, in the Online Supplemental Material, for ordinal outcomes we provide direct proofs of Propositions 1 and 2. Our proofs directly construct the probability matrices that achieve the lower and upper bounds of τ and η. We believe that our "constructive" approach helps researchers sharply bound other causal parameters (e.g., m_j and α), at least for ordinal outcomes. It is worth mentioning that the probability matrices attaining the lower and upper bounds of τ and η correspond to negatively associated and positively associated potential outcomes. They are both "extreme" scenarios. In practice, researchers may also be interested in the case with independent potential outcomes (Agresti, 2010; Cheng, 2009; Ding & Dasgupta, 2016; Rubin, 1978), that is, $p_{kl} = p_{k+}p_{+l}$ for all k and l. With independent potential outcomes, we can identify τ and η from the marginal distributions of the potential outcomes.

Proposition 3: With independent potential outcomes,

$$\tau_I = \sum_{k \geq l}\sum p_{k+}p_{+l}, \quad \eta_I = \sum_{k > l}\sum p_{k+}p_{+l}.$$

Furthermore, $\tau_L \leq \tau_I \leq \tau_U$ and $\eta_L \leq \eta_I \leq \eta_U$.

In cases where negatively associated potential outcomes are unlikely, we can use τ_I and η_I as the lower bounds of τ and η. Below we give two numerical examples to illustrate Propositions 1 through 3.

Example 3: The marginal probabilities $\boldsymbol{p}_1 = (1/5, 3/5, 1/5)^T$ and $\boldsymbol{p}_0 = (2/5, 1/5, 2/5)^T$ do not satisfy the stochastic dominance assumption because $\Delta_0 = 0$, $\Delta_1 = 1/5 > 0$, and $\Delta_2 = -1/5 < 0$. Propositions 1 and 3 imply that $\tau_L = 2/5$, $\tau_I = 16/25$, and $\tau_U = 4/5$ The probability matrices corresponding to negatively associated, independent, and positively associated potential outcomes achieving these values are, respectively,

$$\boldsymbol{P}_1 = \begin{pmatrix} 0 & 1/5 & 0 \\ 1/5 & 0 & 2/5 \\ 2/5 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} 2/25 & 1/25 & 2/25 \\ 6/25 & 3/25 & 6/25 \\ 2/25 & 1/25 & 2/25 \end{pmatrix}, \quad \boldsymbol{P}_3 = \begin{pmatrix} 1/5 & 0 & 0 \\ 1/5 & 1/5 & 1/5 \\ 0 & 0 & 1/5 \end{pmatrix}. \quad (7)$$

Similarly, Propositions 2 and 3 imply $\eta_L = 1/5$, $\eta_I = 9/25$, and $\eta_U = 3/5$. We omit presenting the matrices attaining these values of η.

Example 4: The marginal probabilities $\boldsymbol{p}_1 = (1/5, 1/5, 3/5)^T$ and $\boldsymbol{p}_0 = (3/5, 1/5, 1/5)^T$ satisfy the stochastic dominance assumption because $\Delta_0 = 0$, $\Delta_1 = 2/5 > 0$, and $\Delta_2 = 2/5 > 0$. Propositions 1 and 3 imply $\tau_L = 3/5$, $\tau_I = 22/25$, and $\tau_U = 1$. The probability matrices corresponding to negatively associated, independent, and positively associated potential outcomes achieving these values are, respectively,

$$\boldsymbol{P}_4 = \begin{pmatrix} 0 & 1/5 & 0 \\ 0 & 0 & 1/5 \\ 3/5 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{P}_5 = \begin{pmatrix} 3/25 & 1/25 & 1/25 \\ 3/25 & 1/25 & 1/25 \\ 9/25 & 3/25 & 3/25 \end{pmatrix}, \quad \boldsymbol{P}_6 = \begin{pmatrix} 1/5 & 0 & 0 \\ 0 & 1/5 & 0 \\ 2/5 & 0 & 1/5 \end{pmatrix}. \quad (8)$$

Similarly, Propositions 2 and 3 imply $\eta_L = 2/5$, $\eta_I = 3/5$, and $\eta_U = 4/5$. We omit presenting the matrices attaining these values of η.

As demonstrated in Examples 3 and 4, the bounds of τ (or η) generally do not shrink to a point. However, there are some special cases in which the lower and upper bounds of τ (or η) are identical. The following corollary provides necessary and sufficient conditions for such cases.

Corollary 2: Let $\mathbb{K} = \{k : p_{k+} > 0\}$ and $\mathbb{L} = \{l : p_{+l} > 0\}$. The lower and upper bounds of τ are the same, if and only if there does not exist $k_1, k_2 \in \mathbb{K}$ and $l_1, l_2 \in \mathbb{L}$, such that:

$$k_2 \geq l_2 > k_1 \geq l_1 \quad \text{or} \quad l_2 > k_2 \geq l_1 > k_1. \quad (9)$$

The lower and upper bounds of η are the same, if and only if there does not exist $k_1, k_2 \in \mathbb{K}$ and $l_1, l_2 \in \mathbb{L}$ such that:

$$l_2 \geq k_2 > l_1 \geq k_1 \quad \text{or} \quad k_2 > l_2 \geq k_1 > l_1. \quad (10)$$

In the proof of Corollary 2 in the Online Supplemental Material, we provide the interpretations of the conditions in Equations 9 and 10.

3.2. Covariate Adjustment

With pretreatment covariates, it is possible to further sharpen the bounds of the causal parameters (Grilli & Mealli, 2008; Jiang & Ding, 2018; Lee, 2009; Long & Hudgens, 2013; Mealli & Pacini, 2013). Without loss of generality, we focus only on the bounds of τ. Within each level of the pretreatment covariates $\boldsymbol{X} = \boldsymbol{x}$,

$$\tau(\boldsymbol{x}) = \text{pr}\{Y(1) \geq Y(0) | \boldsymbol{X} = \boldsymbol{x}\},$$

is the conditional probability that the treatment is beneficial. We can obtain the conditional lower and upper bounds $\tau_L(\boldsymbol{x})$ and $\tau_U(\boldsymbol{x})$ given the covariate level $\boldsymbol{x}$, then average them over the covariate distribution $F(\boldsymbol{x})$, and finally obtain the adjusted bounds for τ:

$$\tau'_L = \int \tau_L(\boldsymbol{x}) F(d\boldsymbol{x}), \quad \tau'_U = \int \tau_U(\boldsymbol{x}) F(d\boldsymbol{x}). \quad (11)$$

Proposition 4: The adjusted bounds are tighter, that is, $\tau_L \leq \tau'_L \leq \tau'_U \leq \tau_U$.

Proposition 4 holds intuitively because the existence of covariates imposes more distributional restrictions on the observed data. We use the following example to illustrate Proposition 4.

Example 5: Consider a population consisting of two subpopulations of equal sizes, labeled by a binary covariate $\boldsymbol{X}$. Assume that the potential outcomes of subpopulations $\boldsymbol{X} = 1$ and $\boldsymbol{X} = 0$ are the independent potential outcomes in Example 3 and 4. Simple algebra gives the following joint distribution, marginal distributions, and τ of the potential outcomes:

$$\boldsymbol{P} = \begin{pmatrix} 1/10 & 1/25 & 3/50 \\ 9/50 & 2/25 & 7/50 \\ 11/50 & 2/25 & 1/10 \end{pmatrix}, \quad \boldsymbol{p}_1 = (1/5, 2/5, 2/5)^T, \quad \boldsymbol{p}_0 = (1/2, 1/5, 3/10)^T, \quad \tau = 19/25.$$

Without covariate information, Proposition 1 implies $\tau_L = 1/2$ and $\tau_U = 1$. However, if we first obtain the bounds for the two subpopulations and then average over them, we obtain sharper covariate adjusted bounds $\tau'_L = \tau_L(1)/2 + \tau_L(0)/2 = 1/2$ and $\tau'_U = \tau_U(1)/2 + \tau_U(0)/2 = 9/10$.

3.3. Identifying the Bounds From Observed Data

Previous subsections discussed the causal parameters τ and η and their bounds. The causal parameters depend on the joint distribution of the potential outcomes, but the bounds depend only on the marginal distributions of the potential outcomes. In practice, the observed data provide full information about only the marginal distributions. Therefore, point estimations of the bounds can be obtained, although the causal parameters themselves are only partially identified (cf. Richardson et al., 2014; Romano & Shaikh, 2008, 2010).

For unit $i = 1, \ldots, N$, let the treatment indicator be Z_i and the observed outcome be $Y_i^{\text{obs}} = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$. To avoid conceptual complications, we consider treatment assignments that satisfy the ignorability assumption (Rosenbaum & Rubin, 1983), that is, $Z \perp\!\!\!\perp \{Y(1), Y(0)\} | \boldsymbol{X}$. The ignorability assumption holds by the design of randomized experiments and cannot be validated in observational studies. Under the ignorability assumption, we define the propensity score as $e(\boldsymbol{X}) = \text{pr}(Z = 1 | \boldsymbol{X})$, which is a constant independent of $\boldsymbol{X}$ in completely randomized experiments. We can identify the marginal distributions of the potential outcomes by

$$\text{pr}\{Y(1) = k\} = E\left\{\frac{Z 1(Y^{\text{obs}} = k)}{e(\boldsymbol{X})}\right\}, \quad \text{pr}\{Y(0) = l\} = E\left\{\frac{(1 - Z) 1(Y^{\text{obs}} = l)}{1 - e(\boldsymbol{X})}\right\}.$$

By replacing the expectations by their sample analogues, we obtain the moment estimators for the marginal distributions. We defer more detailed discussion about statistical inference to Section 5.

4. Randomized Experiments With Noncompliance

4.1. Causal Effects for Compliers

Noncompliance is an important topic in practice. For instance, in clinical trials, some patients may not comply with their assigned treatments. Although noncompliance itself has been extensively investigated in the causal inference literature (e.g., Angrist, Imbens, & Rubin, 1996), there appears to be very limited discussions about causal inference of ordinal outcomes in the presence of noncompliance. To the best of our knowledge, Cheng (2009) discussed various causal parameters under the assumptions of one-sided noncompliance, and Baker (2011) generalized her results to two-sided noncompliance; both of them assumed independent potential outcomes.

Under the stable unit treatment value assumption, for unit i, let $\{D_i(1), D_i(0)\}$ be the potential values of treatment received under treatment and control; the observed treatment received is therefore $D_i^{\text{obs}} = Z_i D_i(1) + (1 - Z_i)D_i(0)$. Angrist, Imbens, and Rubin (1996) proposed to classify the units into four categories according to the joint values of $D_i(1)$ and $D_i(0)$:

$$G_i = \begin{cases} a, & \text{if } D_i(1) = 1, D_i(0) = 1, \\ c, & \text{if } D_i(1) = 1, D_i(0) = 0, \\ d, & \text{if } D_i(1) = 0, D_i(0) = 1, \\ n, & \text{if } D_i(1) = 0, D_i(0) = 0, \end{cases} \tag{12}$$

and referred to the subgroups defined in Equation 12 as always-takers (a), compliers (c), defiers (d), and never-takers (n). Let $\pi_g = \text{pr}(G = g)$ denote the probability of the stratum $g \in \{a, c, d, n\}$ and

$$g_{kl} = \text{pr}\{Y(1) = k, Y(0) = l | G = g\},$$

be the probability of potential outcome k under treatment and potential outcome l under control within stratum g. The $J \times J$ probability matrix $\{g_{kl}\}_{0 \le k,l \le J-1}$ summarizes the joint distribution of the potential outcomes for stratum g. Define

$$g_{k+} = \sum_{l'=0}^{J-1} g_{kl'}, \quad g_{+l} = \sum_{k'=0}^{J-1} g_{k'l} \quad (k, l = 0, 1, \ldots, J-1), \tag{13}$$

the vectors $(g_{0+}, \ldots, g_{J-1,+})^T$ and $(g_{+0}, \ldots, g_{+,J-1})^T$ characterize the marginal distributions of the potential outcomes under treatment and control. By the law of total probability,

$$p_{kl} = \sum_g \pi_g g_{kl}, \quad p_{k+} = \sum_g \pi_g g_{k+}, \quad p_{+l} = \sum_g \pi_g g_{+l}. \tag{14}$$

We define the subgroup causal parameters within stratum g as

$$\tau_g = \text{pr}\{Y_i(1) \ge Y_i(0) | G = g\} = \sum_{k \ge l} \sum g_{kl}, \ \eta_g = \text{pr}\{Y_i(1) > Y_i(0) | G = g\} = \sum_{k > l} \sum g_{kl}.$$

Again, these parameters are particularly useful for ordinal outcomes, but are also applicable to general continuous outcomes.

Following Angrist et al. (1996), we invoke the following "standard" assumptions: (1) complete randomization, that is, $Z \perp\!\!\!\perp \{D(1), D(0), Y(1), Y(0), \boldsymbol{X}\}$, (2) monotonicity, that is, $D_i(1) \geq D_i(0)$ for all i, and (3) exclusion restriction, that is, $D_i(1) = D_i(0)$ implies $Y_i(1) = Y_i(0)$. Monotonicity rules out the defiers with $G = d$, and strong monotonicity further rules out the always-takers with $G = a$. Exclusion restriction implies that $\tau_n = 1$, $\eta_n = 0$, $\tau_a = 1$, and $\eta_a = 0$. Therefore, we discuss only the causal effects for the compliers, that is, τ_c and η_c.

4.2. Bounds on the Causal Effects for Compliers

We focus only on the monotonicity assumption because it is more general than strong monotonicity. Under monotonicity and exclusion restriction, we identify the probabilities of always-takers, compliers, and never-takers, that is, (π_a, π_c, π_n), and the distributions of the potential outcomes conditional on G (Angrist et al., 1996; Baker, 2011; Cheng, 2009), that is, the g_{k+}'s and g_{+l}'s. Below, we establish the relationships between the causal parameters τ and τ_c and between η and η_c.

Proposition 5: $\tau_c = \tau/\pi_c - (1 - \pi_c)/\pi_c$ and $\eta_c = \eta/\pi_c$.

Therefore, we can plug in the upper and lower bounds of τ and η to obtain the bounds of τ_c and η_c, using the relationships in Proposition 5. However, these bounds are not sharp, and the following bounds, implied by Propositions 1 and 2, are narrower.

Corollary 3: The sharp lower and upper bounds of τ_c are

$$\tau_{c,L} = \max_{0 \leq j \leq J-1} (c_{+j} + \Delta_{c,j}), \quad \tau_{c,U} = 1 + \min_{0 \leq j \leq J-1} \Delta_{c,j},$$

and the sharp lower and upper bounds of η_c are

$$\eta_{c,L} = \max_{0 \leq j \leq J-1} \Delta_{c,j}, \quad \eta_{c,U} = 1 + \min_{0 \leq j \leq J-1} (\Delta_{c,j} - c_{j+}).$$

Similar to Section 3.2, we can use covariates to sharpen the bounds of τ_c Within each level of the pretreatment covariates $\boldsymbol{X} = \boldsymbol{x}$, we define the conditional probabilities that the treatment is beneficial for compliers as

$$\tau_c(\boldsymbol{x}) = \text{pr}\{Y(1) \geq Y(0) | G = c, \boldsymbol{X} = \boldsymbol{x}\},$$

and obtain their conditional sharp upper and lower bounds $\tau_{c,L}(\boldsymbol{x})$ and $\tau_{c,U}(\boldsymbol{x})$. Because

$$\tau_c = \frac{\int \tau_c(\boldsymbol{x}) \pi_c(\boldsymbol{x}) dF(\boldsymbol{x})}{\int \pi_c(\boldsymbol{x}) dF(\boldsymbol{x})},$$

the bounds for τ_c become

$$\tau'_{c,L} = \frac{\int \tau_{c,L}(\boldsymbol{x})\pi_c(\boldsymbol{x})dF(\boldsymbol{x})}{\int \pi_c(\boldsymbol{x})dF(\boldsymbol{x})}, \quad \tau'_{c,U} = \frac{\int \tau_{c,U}(\boldsymbol{x})\pi_c(\boldsymbol{x})dF(\boldsymbol{x})}{\int \pi_c(\boldsymbol{x})dF(\boldsymbol{x})}. \tag{15}$$

Similar to Proposition 4, the adjusted bounds are tighter, that is, $\tau_{c,L} \leq \tau'_{c,L} \leq \tau'_{c,U} \leq \tau_{c,U}$.

4.3. Using Noncompliance to Sharpen Bounds for the Whole Population

Proposition 5 and Corollary 3 imply two new sets of bounds for τ and η, which are tighter than those in Propositions 1 and 2.

Corollary 4: Under monotonicity and exclusion restriction, we can bound τ from below and above using

$$\tau''_L = \pi_c\tau_{c,L} + 1 - \pi_c, \quad \tau''_U = \pi_c\tau_{c,U} + 1 - \pi_c,$$

and bound η from below and above using

$$\eta''_L = \pi_c\eta_{c,L}, \quad \eta''_U = \pi_c\eta_{c,U}.$$

These new bounds above are narrower than those in Propositions 1 and 2 because they satisfy $\tau_L \leq \tau''_L$, $\tau_U = \tau''_U$, $\eta_L = \eta''_L$, and $\eta_U \geq \eta''_U$.

There are two reasons that we can obtain tighter bounds. First, we use the partially observed variable G as a pretreatment variable. Second, the monotonicity and exclusion restriction assumptions further restrict the probability structure of the potential outcomes.

5. Statistical Inference of the Bounds

5.1. Point Estimation

In practice, we need to use the observed data to estimate the marginal probabilities of the potential outcomes and the bounds. To save space for the main text, we discuss only the bounds of τ and τ_c We describe the point estimation procedures for the three scenarios mentioned in the previous sections—completely randomized experiments with or without noncompliance and unconfounded observational studies.

First, we consider completely randomized experiments without noncompliance. To estimate the unadjusted bounds, we replace p_{k+} and p_{+l} in Proposition 1 with their sample analogues

$$\widehat{p}_{k+} = N^{-1}\sum_{i=1}^{N} Z_i 1(Y_i^{\text{obs}} = k), \quad \widehat{p}_{+l} = N^{-1}\sum_{i=1}^{N}(1 - Z_i)1(Y_i^{\text{obs}} = l).$$

To estimate the covariate adjusted bounds in Equation 11, we first estimate the marginal probabilities of the potential outcomes of unit i given covariates $\boldsymbol{x}_i$. Following Imbens and Rubin (2015), for low-dimensional and discrete covariates, we can still use sample analogues. For high-dimensional and continuous covariates, we can invoke parametric models such as proportional odds models. We then use the estimates, denoted by $\widehat{p}_{k+}(\boldsymbol{x}_i)$ and $\widehat{p}_{+l}(\boldsymbol{x}_i)$, respectively, to estimate the sharp lower and upper bounds of $\tau(\boldsymbol{x}_i)$, denoted by $\widehat{\tau}_L(\boldsymbol{x}_i)$ and $\widehat{\tau}_U(\boldsymbol{x}_i)$, respectively. Consequently, the estimated adjusted bounds of τ are as follows:

$$\widehat{\tau}'_L = N^{-1}\sum_{i=1}^{N}\widehat{\tau}_L(\boldsymbol{x}_i), \quad \widehat{\tau}'_U = N^{-1}\sum_{i=1}^{N}\widehat{\tau}_U(\boldsymbol{x}_i).$$

Second, we consider unconfounded observational studies. If we have propensity score estimator $\widehat{e}(\boldsymbol{x}_i)$ for unit i, then we can estimate the marginal probabilities by:

$$\widehat{p}_{k+} = N^{-1}\sum_{i=1}^{N} Z_i \frac{1(Y_i^{\text{obs}} = k)}{\widehat{\boldsymbol{e}}(\boldsymbol{x}_i)}, \quad \widehat{p}_{+l} = N^{-1}\sum_{i=1}^{N}(1 - Z_i)\frac{1(Y_i^{\text{obs}} = l)}{1 - \widehat{\boldsymbol{e}}(\boldsymbol{x}_i)},$$

and then estimate the bounds accordingly.

Third, we consider completely randomized experiments with noncompliance. Without covariates, we use the expectation maximization (EM) algorithm by Dempster, Laird, and Rubin (1976) to estimate π_c, c_{k+}, and c_{+l} and then estimate the unadjusted bounds in Corollary 3. For a more detailed description of the EM algorithm, see Baker (2011). With covariates, we need to invoke parametric models for G (e.g., multinomial logistic model given $\boldsymbol{X}$) and the marginal probabilities of the potential outcomes and use the EM algorithm to compute the maximum likelihood of the model parameters. For more details, see Online Supplemental Material, and Zhang, Rubin, and Mealli (2009) and Frumento, Mealli, Pacini, and Rubin (2012). After obtaining the sample analogues of $\tau_{c,L}(\boldsymbol{x})$, $\tau_{c,U}(\boldsymbol{x})$, and $\pi_c(\boldsymbol{x})$, we estimate the covariate adjusted bounds defined in Equation 15 using a plug-in approach.

5.2. Finite-Sample Bias and Bias Correction

As pointed out by several researchers (e.g., Liu & Brown, 1993; Manski & Pepper, 2000, 2009), the minimum and maximum operators in the closed-form expressions of the sharp bounds usually complicate the estimation procedure, by introducing finite-sample biases to the corresponding plug-in estimators. For example, even with unbiased estimators of the marginal probabilities (e.g., in completely randomized experiments), the estimated lower bound is positively biased. In the existing literature, this non-smoothness-induced bias has been recognized and discussed by Laber and Murphy (2011), Hirano and Porter (2012), and Luedtke and Van der Laan (2016), under various settings. However, fortunately, such biases tend to diminish as the sample size increases, due to the consistency of the plug-in estimators. More importantly, as pointed out by

TABLE 1.
Numerical Examples

Case	N	τ	τ_L	Bias_p	Bias_b
1	100	0.640	.400	.023	.005
1	200	0.640	.400	.016	.004
1	500	0.640	.400	.009	.001
2	100	0.800	.400	.017	−.002
2	200	0.800	.400	.014	.001
2	500	0.800	.400	.007	−.001
3	100	0.880	.600	.037	.010
3	200	0.880	.600	.026	.007
3	500	0.880	.600	.016	.004
4	100	1.000	.600	.036	.009
4	200	1.000	.600	.026	.007
4	500	1.000	.600	.013	.001

Note. The first four columns contain the case label, sample size, and true values of τ and τ_L. The last two columns contain the biases of the plug-in (labeled "*p*") estimator and the bootstrap bias-corrected (labeled "*b*") estimators, calculated by 1,000 repeat samplings and 200 bootstraps for each sample.

Kreider and Pepper (2007), it is possible to effectively reduce such biases by a nonparametric bootstrap correction (Efron & Tibshirani, 1994; Parr, 1983). To be more specific, let $\hat{\tau}_L$ denote the point estimator of τ_L, and the corresponding bias-corrected estimator is therefore $2\widehat{\tau}_L - \mathrm{E}_B(\widehat{\tau}_L)$, where "$E_B$" denotes the expectation induced by the bootstrap distribution.

The following numerical example demonstrates the magnitude of the bias associated with the plug-in estimator and the performance of the bias-correction estimator.

Example 6: Consider a completely randomized experiment without noncompliance. To save space, we focus only on τ and its unadjusted lower bound τ_L in Equation 5. We choose the sample size $N \in \{100, 200, 500\}$ and consider four different probability matrices. Cases 1 and 2 correspond to matrices $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ in Equation 7, that is, the independent and positively associated potential outcomes, which share the same marginal distribution but do not satisfy the stochastic dominance assumption. Cases 3 and 4 correspond to matrices $\boldsymbol{P}_5$ and $\boldsymbol{P}_6$ in Equation 8, that is, the independent and positively associated potential outcomes, which share the same marginal distribution and satisfy the stochastic dominance assumption. Columns 3 and 4 of Table 1 summarize the true values of τ and τ_L for all four cases.

For each case and fixed sample size, we independently draw 1,000 treatment assignments from a balanced completely randomized experiment. For each observed data set, we calculate point estimates of τ_L using the plug-in estimator and the bias-correction estimator based on 200 bootstraps. In Columns 5 and 6 of Table 1, we report the biases of the two point estimators, from which we can draw

two conclusions. First, for each case, the bias of the plug-in estimator decreases as the sample size increases. Second, the bias-corrected estimator greatly reduces (in most cases by over 60%) the bias of the plug-in estimator.

5.3. Confidence Intervals (CIs)

We discuss the construction of CIs for the aforementioned causal parameters and their unadjusted or covarite adjusted bounds. For illustration, we again use τ as an example. From a practical (e.g., decision-making) perspective, we aim to construct a CI that covers τ at least $100(1-\alpha)\%$ of the times, for prespecified significance level α. Because the casual parameter is only partially identifiable, it is difficult to do so directly without additional assumptions or information. A common approach to address this challenge is to instead construct a $100(1-\alpha)\%$ CI for the sharp bounds $[\tau_L, \tau_U]$. Because $\tau \in [\tau_L, \tau_U]$, the resulted interval automatically guarantees at least $100(1-\alpha)\%$ coverage rate for τ itself.

Similar to the point estimation procedure because both the upper and the lower bounds involve the maximum and minimum operators, their asymptotic distributions become nonnormal, rendering the construction of CIs covering the bounds extremely challenging (Hirano & Porter, 2012). Consequently, in practice, statisticians (Cheng & Small, 2006; Yang & Small, 2016) often employed bootstrap methods (e.g., Beran, 1988, 1990; Bickel, Gotze, & Van Zwet, 1997; Bickel & Sakov, 2008) to construct CIs for partially identified parameters. Among numerous proposals, the most conceptually straightforward and transparent one is arguably the "standard" bootstrap procedure advocated by Horowitz and Manski (2000), for which $1-\alpha$ CI is simply $\{\widehat{\tau}_L - z_B(\alpha), \widehat{\tau}_U + z_B(\alpha)\}$, where the threshold value $z_B(\alpha)$ can be obtained by solving the equation:

$$\Pr_B\{\widehat{\tau}_L - z_B(\alpha) \leq \widehat{\tau}_L, \widehat{\tau}_U \leq \widehat{\tau}_U + z_B(\alpha)\} = 1-\alpha.$$

In the above equation, "$\Pr_B$" is the probability measure induced by bootstrap. Recently, several researchers (e.g., Chernozhukov, Lee, & Rosen, 2013; Romano & Shaikh, 2008, 2010) proposed more delicate methods to construct CIs for partially identified parameters. Although the theoretical guarantees of the classic bootstrapped CIs (Horowitz & Manski, 2000) are not completely established, several researchers (e.g., Fan & Park, 2010; Yang, 2014) have evaluated them via extensive simulation studies and found that they achieve nominal coverage rates in many realistic scenarios.

To empirically illustrate the validity of our inferential procedure, in Online Appendix C, we compare Horowitz and Manski's (2000) method to a more theoretically rigorous one, under a wide range of settings. The results suggest that, at least in our context, Horowitz and Manski's (2000) bootstrap interval performs equally well, if not slightly more "conservative." Therefore, for simplicity in simulations and transparency in applications, we still use bootstrap to

construct CIs. We provide the code to implement the above construction approach; more sophisticated users can straightforwardly modify our code and explore more advanced methods.

6. Simulation Studies

6.1. Without Noncompliance

To save space for the main text, we focus only on τ and its sharp bounds in Proposition 1. For illustration, we first adopt the settings (i.e., sample sizes and probability matrices) in Example 6. It is worth mentioning that, for Cases 1 and 3 with independent potential outcomes, $\tau_L < \tau < \tau_U$. For Cases 2 and 4 with positively associated potential outcomes, $\tau = \tau_U$. In addition, by symmetry Cases 1 through 4, only consider $\tau > .5$.

For each case, we independently draw 1,000 treatment assignments from a balanced completely randomized experiment. For each observed data set, we obtain bias-corrected estimates of τ_L and τ_U and construct a 95% CI for $[\tau_L, \tau_U]$ using 200 bootstrapped samples. In Columns 5 through 8 of Table 2, we report the biases and standard errors of the point estimators $\widehat{\tau}_L$ and $\widehat{\tau}_U$; in Column 9, we report the coverage rates of the CIs on the bounds $[\tau_L, \tau_U]$ and τ itself. We can draw several conclusions from the simulation results. First, the point estimators have small biases and standard errors. Second, the CIs achieve reasonable coverage rates for the bounds $[\tau_L, \tau_U]$, although always overcover τ, especially in cases with independent potential outcomes.

As mentioned previously, in Online Appendix C, we conduct additional simulation studies to further examine the performance of Horowitz and Manski's (2000) bootstrap CI. The simulation results suggest that it achieves nearly nominal coverage rates for the bounds $[\tau_L, \tau_U]$, except for certain "edge cases" (e.g., when $\tau \approx \tau_U \approx 1$), and as expected usually overcover τ.

6.2. With Noncompliance

To evaluate the finite-sample performances of the estimators and the CIs of the bounds, we conduct simulation studies under different model specifications. To save space, we focus only on the parameter τ_c and consider six simulation cases. Cases 1 through 3 are indexed by the parameter $\beta \in \{1, 1/2, 0\}$ and Cases 4 through 6 by $\xi \in \{1, 1/2, 0\}$. We postpone the interpretations of β and ξ until afterward. For each case, let the pretreatment covariates $\boldsymbol{X} = (1, X_1, X_2)$, where $X_1 \sim N(0, 1)$ and $X_2 \sim \text{Bern}(1/2)$. For fixed $\boldsymbol{X} = \boldsymbol{x}$, we generate the variable G from a multinomial logit model:

$$\pi_g(\boldsymbol{x}) = \exp(\eta_g^{\mathrm{T}}\boldsymbol{x}) / \left\{ \sum_{g'} \exp(\eta_{g'}^{\mathrm{T}}\boldsymbol{x}) \right\} \quad (g = a, c, n),$$

TABLE 2.
Simulated Examples Without Noncompliance

Case	N	τ	τ_L	τ_U	Bias_L	se_L	Bias_U	se_U	Coverage_1	Coverage_2
1	100	0.640	.400	0.800	.005	.056	.001	.067	.989	1.000
1	200	0.640	.400	0.800	.004	.040	−.000	.044	.989	1.000
1	500	0.640	.400	0.800	.001	.025	−.003	.029	.982	1.000
2	100	0.800	.400	0.800	−.002	.063	−.001	.082	.969	0.979
2	200	0.800	.400	0.800	.001	.044	−.000	.057	.966	0.976
2	500	0.800	.400	0.800	−.001	.027	−.002	.035	.968	0.979
3	100	0.880	.600	1.000	.010	.049	.000	.000	.959	1.000
3	200	0.880	.600	1.000	.007	.035	.000	.000	.965	1.000
3	500	0.880	.600	1.000	.004	.022	.000	.000	.969	1.000
4	100	1.000	.600	1.000	.009	.053	.000	.000	.940	1.000
4	200	1.000	.600	1.000	.007	.035	.000	.000	.967	1.000
4	500	1.000	.600	1.000	.001	.021	.000	.000	.983	1.000

Note. The first five columns contain the case number, sample size, and true values of the parameter and its sharp lower and upper bounds. The next four columns contain the biases and standard errors of the point estimators of the bounds, and the last two columns contain the coverage properties of the confidence intervals for the bounds (labeled "coverage_1") and the true parameter itself (labeled "coverage_2").

where $\eta_c = \mathbf{0}$, $\eta_a = (1/2, 1, 0)$, and $\eta_n = (-1/2, 1, 0)$. We generate the potential outcomes from proportional odds models.

1. For always-takers, let $Y_i(1) = Y_i(0)$, and their marginal distributions be:

$$\text{logit}\left\{\sum_{k\le j} a_{k+}(\boldsymbol{x})\right\} = \text{logit}\left\{\sum_{l\le j} a_{+l}(\boldsymbol{x})\right\} = \alpha_{a,j} - 2\boldsymbol{x}_1,$$

   where $\alpha_{a,0} = -1/2$ and $\alpha_{a,1} = 1$.

2. For never-takers, let $Y_i(1) = Y_i(0)$, and their marginal distributions be

$$\text{logit}\left\{\sum_{k\le j} n_{k+}(\boldsymbol{x})\right\} = \text{logit}\left\{\sum_{l\le j} n_{+l}(\boldsymbol{x})\right\} = \alpha_{n,j},$$

   where $\alpha_{a,0} = -3/2$ and $\alpha_{a,1} = 0$

3. For compliers let $Y_i(1)$ and $Y_i(0)$ be independent, and the values of the parameters be $\alpha_{c,0} = -1$, $\alpha_{c,1} = 1/2$, $\gamma_{c,0} = 1/2$, and $\gamma_{c,1} = 2$.

a. For Cases 1–3, let the marginal distributions be

TABLE 3.
Simulated Examples With Noncompliance

Unadjusted Bounds

Case	τ_c	$\tau_{c,L}$	$\tau_{c,U}$	Bias_L	Bias_U	Length	Coverage
1	.685	.488	0.971	−.003	−.005	.659	.945
2	.770	.553	1.000	−.008	.006	.574	.973
3	.856	.622	1.000	.013	.001	.489	.966
4	.782	.589	1.000	.000	.006	.523	.957
5	.736	.540	1.000	−.003	.003	.593	.975
6	.686	.488	0.970	−.001	−.004	.655	.945

Adjusted bounds

	τ_c	$\tau'_{c,L}$	$\tau'_{c,U}$	Bias_L	Bias_U	Length	Coverage
1	.685	.503	0.772	−.001	.003	.466	.968
2	.770	.563	0.935	−.006	.001	.530	.968
3	.856	.622	1.000	.001	.002	.489	.959
4	.782	.602	0.846	−.002	.017	.436	.960
5	.738	.556	0.817	−.001	.004	.447	.965
6	.686	.503	0.772	.008	−.006	.466	.968

Note. In each subtable, the first three columns contain the true values of the causal parameter τ_c and its lower and upper bounds, the next two columns contain the biases of the point estimators of the lower and upper bounds, and the last two columns contain the lengths and coverage rates of the 95% confidence intervals for the bounds.

$$\text{logit}\left\{\sum_{k\le j} c_{k+}(\boldsymbol{x})\right\} = \alpha_{c,j} - 2\beta\boldsymbol{x}_1, \quad \text{logit}\left\{\sum_{l\le j} c_{+l}(\boldsymbol{x})\right\} = \gamma_{c,j} + \beta\boldsymbol{x}_1.$$

b. For Cases 4–6, let the marginals distributions be:

$$\text{logit}\left\{\sum_{k\le j} c_{k+}(\boldsymbol{x})\right\} = \alpha_{c,j} - 2\boldsymbol{x}_1 - \xi\boldsymbol{x}_2, \quad \text{logit}\left\{\sum_{l\le j} c_{+l}(\boldsymbol{x})\right\} = \gamma_{c,j} + \boldsymbol{x}_1 + \xi\boldsymbol{x}_2.$$

For the above six cases, their true values of τ_c unadjusted and adjusted bounds are in Columns 2 through 4 of each subtable of Table 3. For Cases 1 through 3, the parameter β quantifies the association between the covariates and the potential outcomes. As β decreases, the covariate adjusted bounds become closer to the unadjusted bounds. For Cases 4 through 6, the parameter ξ quantifies the association between the binary covariate X_2 and the potential outcomes of compliers.

We conduct inference without the binary covariate X_2. This does not affect Cases 1 through 3 because X_2 is irrelevant in the data-generating process but does affect Cases 4 through 6. We purposefully design the data-generating process in this way to examine the performance of our estimators under correct and incorrect model specifications. For each case, we choose the sample size to be 1,000 and independently draw 1,000 treatment assignments from a balanced completely randomized experiment. For each observed data set, based on based on 100 bootstrapped samples, we first obtain the bias-corrected estimates of $\tau_{c,L}$ and $\tau_{c,U}$ and construct a 95% CI for $[\tau_{c,L}, \tau_{c,U}]$; we then estimate the bounds $\tau'_{c,L}$ and $\tau'_{c,U}$ and construct a 95% CI for $[\tau'_{c,L}, \tau'_{c,U}]$.

We report the simulation results in Table 3, in which Columns 4 through 7 of each subtable include the biases of the point estimators, the average lengths, and coverage rates of the 95% CIs on the bounds. First, the point estimators of the bounds have small biases. Second, when the pretreatment covariates are associated with the potential outcomes, the CIs of the bounds $[\tau_{c,L}, \tau_{c,U}]$ are longer than those of $[\tau'_{c,L}, \tau'_{c,U}]$ on average. Third, the CIs for the bounds $[\tau_{c,L}, \tau_{c,U}]$ and $[\tau'_{c,L}, \tau'_{c,U}]$ achieve reasonable coverage rates. Fourth, the performance of the bounds is robust to the missingness of the binary covariate, or, equivalently, a misspecification of the outcome models.

7. Applications

7.1. A Taste-Testing Experiment Without Noncompliance

We use the taste-testing experiment data in Bradley, Katti, and Coons (1962) to demonstrate the estimation and inference of the proposed causal parameters. The outcome of interest Y is ordinal with five categories, from "terrible" with $Y = 0$ to "excellent" with $Y = 4$. We consider only three treatments C, D, and E, and summarize the data and results in Table 4. Because negative associated potential outcomes appear unlikely in practice (Ding & Dasgupta, 2016), that is, the three treatments are not drastically different (e.g., $Y_i(C) = 4$ and $Y_i(E) = 0$), we focus on the interpretations of the cases with independent and positive correlated potential outcomes, for example, τ_I and τ_U. First, treatment E stochastically dominates treatment C, and the CIs for $[\tau_I, \tau_U]$ and $[\eta_I, \eta_U]$ are [0.914, 1.000] and [0.651, 0.997]. The results suggest that treatment E is indeed better than treatment C because both lower confidence limits are greater than 0.5. Second, although Treatments E and D do not stochastically dominate each other, the CIs for $[\tau_I, \tau_U]$ and $[\eta_I, \eta_U]$ are [0.656, 0.982] and [0.510, 0.886], suggesting that Treatment E is better than Treatment D. Therefore, the proposed causal parameters τ and η are useful for decision-making, especially when the stochastic dominance assumption does not hold.

TABLE 4.
Analysis of a Taste-Testing Experiment

Data From Bradley et al. (1962)

	Outcome Categories					
Treatment	0	1	2	3	4	Row Sum
C	14	13	6	7	0	40
D	11	15	3	5	8	42
E	0	2	10	30	2	44

Results for τ: Point estimators and confidence intervals (CIs)

	$\widehat{\tau}_L$	$\widehat{\tau}_I$	$\widehat{\tau}_U$	CI for $[\tau_L, \tau_U]$	CI for $[\tau_I, \tau_U]$
E vs. C	.765	.946	1.000	[0.667, 1.000]	[0.914, 1.000]
E vs. D	.630	.782	0.856	[0.503, 0.997]	[0.656, 0.982]

Results for η: Point estimators and confidence intervals (CIs)

	$\widehat{\eta}_L$	$\widehat{\eta}_I$	$\widehat{\eta}_U$	CI for $[\eta_L, \eta_U]$	CI for $[\eta_I, \eta_U]$
E vs. C	.623	.780	.870	[0.480, 1.000]	[0.651, 0.997]
E vs. D	.573	.659	.738	[0.413, 0.896]	[0.510, 0.886]

Note. CI = confidence interval.

7.2. A Sexual Assault Education Program Without Noncompliance

Between September 2011 and February 2013, three universities in Canada (Windsor, Guelph, and Calgary) conducted the Sexual Assault Resistance Education (SARE) trial. The SARE trial investigates whether the enhanced Assess, Acknowledge, and Act (AAA) program, which consist of numerous activities (e.g., lectures, discussions, and practices) can help prevent sexual assaults. Four hundred and fifty-one first-year female students from the above universities where randomly assigned to the treatment group ($Z = 1$) with access to AAA, and 442 were randomly assigned to the control group ($Z = 0$) with brochures containing general information on sexual assault. The primary outcome Y is ordinal with six categories, from "complete rape" with $Y = 0$ to "no reporting of any non-consensual sexual contact" with $Y = 5$.

We summarize the data and results in Table 5. Because both the treatment and control groups receive useful information on sexual assault prevention, negatively associated potential outcomes seem unlikely. Therefore, we again focus on independent and positively correlated potential outcomes. The CIs for $[\tau_I, \tau_U]$ and $[\eta_I, \eta_U]$ are [0.758, 1.000] and [0.554, 0.999], suggesting that AAA is indeed beneficial because both lower confidence limits are greater than 0.5. Our findings corroborate the recommendations by Senn et al. (2015).

TABLE 5.
Analysis of the SARE Trial

Data From Senn et al. (2015)							
Outcome Categories							
	0	1	2	3	4	5	Row Sum
Treatment	23	15	48	67	121	177	451
Control	42	40	62	103	184	11	442

Results for τ and η: point estimators and confidence intervals (CIs)

	Lower bound	Indep.	Upper bound	CI for $[L, U]$	CI for $[I, U]$
τ	.636	.783	1.000	[0.598, 1.000]	[0.758, 1.000]
η	.368	.604	0.962	[0.311, 1.000]	[0.554, 0.999]

Note. SARE = Sexual Assault Resistance Education; CI = confidence interval.

7.3. A Job Training Program With Noncompliance

In the mid-1990s, Mathematica Policy Research conducted an experiment that randomly enrolled eligible applicants into the Job Corps program (Lee, 2009; Schochet et al., 2003). We reanalyzed the data set from 1995 with 13,499 units. For more detailed descriptions of the data set, see Zhang et al. (2009) and Frumento et al. (2012). In the following analysis, $Z = 1$ if an applicant was enrolled in the program and $Z = 0$ otherwise; $D = 1$ if an applicant actually participated in the program, and $D = 0$ otherwise. The strong monotonicity assumption with $D_i(0) = 0$ for all i holds by design. Using the hourly wage after 52 weeks of enrollment, we create a three-level ordinal outcome Y as follows: $Y = 0$ for zero wage because of unemployment, $Y = 1$ for low wage (no more than 4.25 U.S. dollars, 150% of the minimal wage at the time the data were collected), and $Y = 2$ for high wage (more than 4.25 U.S. dollars). In the following analysis, we take into account covariates such as gender, age, education, and marital status.

We report the results in Table 6. Similar as before, we focus on independent and positively correlated potential outcomes. For both causal parameters τ_c and η_c, the CIs for the lower and upper bounds become narrower when we take covariates into account. Similarly as the previous example, we focus on the interpretations of the cases with independent and positive correlated potential outcomes. The CIs with or without covariates for $[\tau_I, \tau_U]$ suggest that the hourly wages of more than 70% of participants do not decrease because of the job training program. Additionally, the CIs with or without covariates for $[\eta_I, \eta_U]$

TABLE 6.
Analysis of the Job Corps Program

Results for τ: point estimators and confidence intervals (CIs)

	$\widehat{\tau}_{c,L}$	$\widehat{\tau}_{c,I}$	$\widehat{\tau}_{c,U}$	CI for $[\tau_{c,L}, \tau_{c,U}]$	CI for $[\tau_{c,I}, \tau_{c,U}]$
Without covariates	.561	.707	.912	[.536, .938]	[.687, .934]
With covariates	.592	.723	.912	[.570, .932]	[.700, .932]

Results for η: point estimators and confidence intervals (CIs)

	$\widehat{\eta}_{c,L}$	$\widehat{\eta}_{c,I}$	$\widehat{\eta}_{c,U}$	CI for $[\eta_{c,L}, \eta_{c,U}]$	CI for $[\eta_{c,I}, \eta_{c,U}]$
Without covariates	.005	.209	.351	[.000, .362]	[.199, .363]
With covariates	.004	.193	.320	[.000, .331]	[.180, .331]

Note. CI = confidence interval.

suggest that the hourly wages of roughly 20% to 30% of participants strictly increase because of the job training program.

As a final note, we use this example to illustrate Corollary 4. First, without the noncompliance information, the estimators of the bounds of τ are $\widehat{\tau}_L = .558$ and $\widehat{\tau}_U = .937$, with 95% CI [0.541, 0.954]; the estimators of the bounds of η are $\widehat{\eta}_L = .004$ and $\widehat{\eta}_U = .379$, with 95% CI [0.000, 0.388]. With the noncompliance information, the estimators of the bounds of τ are $\widehat{\tau}''_L = .683$ and $\widehat{\tau}''_U = .937$, with 95% CI [0.666, 0.954]; the estimators of the bounds of η are $\widehat{\eta}''_L = .004$ and $\widehat{\eta}''_U = .254$, with 95% CI [0.000, 0.262]. Therefore, the noncompliance information in return improves the inferences of τ and η for the whole population.

8. Concluding Remarks

We proposed to use two causal parameters to evaluate treatment effect on ordinal outcomes, and derived the explicit forms of their sharp bounds by using only the marginal distributions of the potential outcomes. Although we advocate the use of parameters τ and η to measure treatment effects, we acknowledge that some other causal parameters may also provide information in practice (e.g., Agresti, 2010; Volfovsky, Airoldi, & Rubin, 2015). For general parameters, although deriving the explicit forms of the bounds may be difficult, we can use numerical methods. For instance, for another widely used parameter, the relative treatment effect $\alpha = \tau + \eta - 1$ (Agresti, 2010), we can use numerical linear programs to calculate its maximum and minimum values under the constraints in Equation 4.

Acknowledgments

The authors thank Drs. Avi Feller at Berkeley and Luke W. Miratrix at Harvard for various helpful suggestions. JL is grateful to several colleagues at Microsoft Corporation, especially Dr. Alex Deng, for continuous encouragement and support. Thoughtful comments from the coeditor, Dr. Dan McCaffrey, and three anonymous reviewers have helped improve the quality and presentation of our article significantly.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: PD gratefully acknowledges financial support from the Institute for Education Science (Grant No. R305D150040) and the National Science Foundation (Grant No. DMS-1713152).

ORCID iD

J. Lu http://orcid.org/0000-0002-8839-6024

References

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: John Wiley.

Agresti, A., & Kateri, M. (2017). Ordinal probability effect measures for group comparisons in multinomial cumulative link models. *Biometrics*, *73*, 214–219.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, *91*, 444–455.

Baker, S. G. (2011). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome: An alternative approach. *Biometrics*, *67*, 319–323.

Beran, R. (1988). Balanced simultaneous confidence sets. *Journal of the American Statistical Association*, *83*, 679–697.

Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *Journal of the American Statistical Association*, *85*, 517–426.

Bickel, P. J., Gotze, F., & Van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, *7*, 1–31.

Bickel, P. J., & Sakov, A. (2008). On the choice of m in the m-out-of-n bootstrap and confidence bounds for extrema. *Statistica Sinica*, *18*, 967–985.

Bradley, R. A., Katti, S. K., & Coons, I. J. (1962). Optimal scaling for ordered categories. *Psychometrika*, *27*, 355–374.

Cheng, J. (2009). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics*, *65*, 96–103.

Cheng, J., & Small, D. S. (2006). Bounds on causal effects in three-arm trials with noncompliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*, 815–836.

Chernozhukov, V., Lee, S., & Rosen, A. (2013). Intersection bounds: Estimation and inference. *Econometrica*, *81*, 667–737.

Demidenko, E. (2016). The *p*-value you can't buy. *The American Statistician*, *70*, 33–38.

Dempster, A. P., Laird, N., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data using the EM algorithm (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *39*, 1–38.

Diaz, I., Colantuoni, E., & Rosenblum, M. (2016). Enhanced precision in the analysis of randomized trials with ordinal outcomes. *Biometrics*, *72*, 422–431.

Ding, P., & Dasgupta, T. (2016). A potential tale of two by two tables from completely randomized experiments. *Journal of the American Statistical Association*, *111*, 157–168.

Djebbari, H., & Smith, J. A. (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, *145*, 64–80.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall.

Fan, Y., & Park, S. S. (2009). Partial identification of the distribution of treatment effects and its confidence sets. *Advanced Economies*, *25*, 3–70.

Fan, Y., & Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, *26*, 931–951.

Fan, Y., Sherman, R., & Shum, M. (2014). Identifying treatment effects under data combination. *Econometrica*, *82*, 811–822.

Frank, M. J., Nelsen, R. B., & Schweizer, B. (1987). Best-possible bounds for the distribution of a sum: A problem of Kolmogorov. *Probability Theory and Related Fields*, *74*, 199–211.

Freedman, D. A. (2008). Randomization does not justify logistic regression. *STASIS Trailer*, *23*, 237–249.

Frumento, P., Mealli, F., Pacini, B., & Rubin, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *Journal of the American Statistical Association*, *107*, 450–466.

Gadbury, G. L., & Iyer, H. K. (2000). Unit-treatment interaction and its practical consequences. *Biometrics*, *56*, 882–885.

Grilli, L., & Mealli, F. (2008). Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. *Journal of Educational and Behavioral Statistics*, *33*, 111–130.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, *64*, 487–535.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*, 1161–1189.

Hirano, K., & Porter, J. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, *80*, 1769–1790.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. New York, NY: Springer Science & Business Media.

Horowitz, J. L., & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, *95*, 77–84.

Huang, E. J., Fang, E. X., Hanley, D. F., & Rosenblum, M. (2017). Inequality in treatment benefits: Can we determine if a new treatment benefits the many or the few. *Biostatistics*, *18*, 308–324.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York, NY: Cambridge University Press.

Jiang, Z., & Ding, P. (2018). Using missing types to improve partial identification with missing binary outcomes. *Annals of Applied Statistics*, in press.

Ju, C., & Geng, Z. (2010). Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*, 129–142.

Kim, J. H. (2014). Identifying the distribution of treatment effects under support restrictions. *arXiv*. Retrieved from https://arxiv.org/abs/1410.5885

Kreider, B., & Pepper, J. V. (2007). Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, *102*, 432–441.

Laber, E. B., & Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association*, *106*, 904–913.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*, 1071–1102.

Liu, R. C., & Brown, L. D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Annals of Statistics*, *21*, 1–13.

Long, D. M., & Hudgens, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics*, *69*, 812–819.

Luedtke, A. R., & Van der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, *44*, 713–742.

Makarov, G. D. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and Its Applications*, *26*, 803–806.

Manski, C. F. (1997). Monotone treatment response. *Econometrica*, *65*, 1311–1334.

Manski, C. F. (2003). *Partial identification of probability distributions*. New York, NY: Springer.

Manski, C. E., & Pepper, J. V. (2000). Monotone instrumental variables with an application to the returns to schooling. *Econometrica*, *68*, 997–1010.

Manski, C. F., & Pepper, J. V. (2009). More on monotone instrumental variables. *The Econometrics Journal*, *12*, 200–216.

Mealli, F., & Pacini, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *Journal of the American Statistical Association*, *108*, 1120–1131.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). New York, NY: Springer.

Newcombe, R. G. (2006a). Confidence intervals for an effect size measure based on the mann–whitney statistic. Part 1: General issues and tail-area-based methods. *Statistics in Medicine*, *25*, 543–557.

Newcombe, R. G. (2006b). Confidence intervals for an effect size measure based on the mann–whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, *25*, 559–573.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. Section 9. *Statistical Science*, *5*, 465–472.

Oenema, A., Brug, J., & Lechner, L. (2001). Web-based tailored nutrition education: Results of a randomized controlled trial. *Health Education Research*, *16*, 647–660.

Parr, W. C. (1983). A note on the jackknife, the bootstrap and the delta method estimators of bias and variance. *Biometrika*, *70*, 719–722.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). New York, NY: Cambridge University Press.

Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: A randomised controlled study. *Teaching and Teacher Education*, *39*, 56–65.

Richardson, A., Hudgens, M. G., Gilbert, P. B., & Fine, J. P. (2014). Nonparametric bounds and sensitivity analysis of treatment effects. *Statistical Science*, *29*, 596–618.

Romano, J. P., & Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, *138*, 2786–2807.

Romano, J. P., & Shaikh, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica*, *78*, 169–211.

Rosenbaum, P. R. (2001). Effects attributable to treatment: Inference in experiments and observational studies within a discrete pivot. *Biometrika*, *88*, 219–231.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, *6*, 34–58.

Rubin, D. B. (1980). Comment on "Randomization analysis of experimental data: The Fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, *75*, 591–593.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, *14*, 623–632.

Scharfstein, D. O., Manski, C. F., & Anthony, J. C. (2004). On the construction of bounds in prospective studies with missing ordinal outcomes: Application to the good behavior game trial. *Biometrics*, *60*, 154–164.

Schochet, P. Z., Cao, J. B. R., Glazerman, S., Grady, A., Gritz, M., McConnell, S., Johnson, T., & Burghardt, J. (2003). *National job corps study: Data documentation and public use files, Volume I. Documentation*. Washington, DC: Mathematica Policy Research.

Senn, C. Y., Eliasziw, M., Barata, P. C., Thurston, W. E., Newby-Clark, I. R., Radtke, H. L., & Hobden, K. L. (2015). Efficacy of a sexual assault resistance program for university women. *The New England Journal of Medicine*, *372*, 2326–2335.

Smith, T. W., Marsden, P., Hout, M., & Kim, J. (2013). *General social surveys, 1972–2010*. Chicago, IL: National Opinion Research Center.

Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, *36*, 423–439.

Volfovsky, A., Airoldi, E. M., & Rubin, D. B. (2015). Causal inference for ordinal outcomes. *arXiv*. Retrieved from https://arxiv.org/abs/1501.01234

Williamson, R. C., & Downs, T. (1990). Probabilistic arithmetic. i. numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, *4*, 89–158.

Yang, F. (2014). *Causal inference methods for addressing censoring by death and unmeasured confounding using instrumental variables* (PhD thesis). University of Pennsylvania. Retrieved from https://repository.upenn.edu/dissertations/AAI3622136/

Yang, F., & Small, D. S. (2016). Using post-quality of life measurement information in censoring by death problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*, 299–318.

Zhang, J. L., Rubin, D. B., & Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, *104*, 166–176.

Zhou, W. (2008). Statistical inference for $P(X < Y)$. *Statistics in Medicine*, *27*, 257–279.

Authors

JIANNAN LU is a data scientist at Microsoft Corporation (Analysis and Experimentation), One Microsoft Way, Redmond, WA 98052, USA; email: jiannl@microsoft.com. His research interests are causal inference, design of experiments, and nonparametric statistics.

PENG DING is an assistant professor in the Department of Statistics, University of California-Berkeley, 425 Evans Hall, Berkeley, CA 94720, USA; email: pengdingpku@berkeley.edu. His research interests are causal inference in experiments and observational studies, missing data.

TIRTHANKAR DASGUPTA is an associate professor in the Department of Statistics and Biostatistics, Rutgers University, 110 Frelinghuysen Rd, Piscataway, NJ 08854, USA; email: tirthankar.dasgupta@rutgers.edu. His research interests are design of experiments, causal inference, and statistical applications in the physical sciences and engineering.

Manuscript received May 10, 2017
First revision received December 8, 2017
Second revision received March 17, 2018
Accepted April 4, 2018

Web-based supplementary material for "Treatment effects on ordinal outcomes: Causal estimands and sharp bounds"

Jiannan Lu, Peng Ding and Tirthankar Dasgupta

The Supplementary Material consists of three parts. Part A contains all the proofs. Part B gives the details of the EM algorithm with noncompliance. Part C presents additional simulation results.

A. Proofs of Lemma, Propositions and Corollaries

A.1. A lemma and its proof

We first state a lemma extending a result in Strassen (1965). This lemma plays a central role in our later proofs, and is also of independent interest.

Lemma 1. Assume that $(x_0, \ldots, x_{n-1})$ and $(y_0, \ldots, y_{n-1})$ are nonnegative constants.

(a) If $\sum_{r=s}^{n-1} x_r \geq \sum_{r=s}^{n-1} y_r$ for all $s = 0, \ldots, n-1$, there exists an $n \times n$ lower triangular matrix $\boldsymbol{A}_n = (a_{kl})_{0 \leq k,l \leq n-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{n-1} a_{kl'} \leq x_k, \quad \sum_{k'=0}^{n-1} a_{k'l} = y_l \quad (k, l = 0, \ldots, n-1). \tag{A.1}$$

(b) If $\sum_{r=s}^{n-1} x_r \leq \sum_{r=s}^{n-1} y_r$ for all $s = 0, \ldots, n-1$, there exists an $n \times n$ upper triangular matrix $\boldsymbol{B}_n = (b_{kl})_{0 \leq k,l \leq n-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{n-1} b_{kl'} = x_k, \quad \sum_{k'=0}^{n-1} b_{k'l} \leq y_l \quad (k, l = 0, \ldots, n-1). \tag{A.2}$$

(c) If $\sum_{r=0}^{s} x_r \le \sum_{r=0}^{s} y_r$ for all $s = 0, \ldots, n-1$, there exists an $n \times n$ lower triangular matrix $\boldsymbol{C}_n = (p_{kl})_{0 \le k,l \le n-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{n-1} p_{kl'} = x_k, \quad \sum_{k'=0}^{n-1} p_{k'l} \le y_l \quad (k, l = 0, \ldots, n-1). \tag{A.3}$$

(d) If $\sum_{r=0}^{s} x_r \ge \sum_{r=0}^{s} y_r$ for all $s = 0, \ldots, n-1$, there exists an $n \times n$ upper triangular matrix $\boldsymbol{D}_n = (d_{kl})_{0 \le k,l \le n-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{n-1} d_{kl'} \le x_k, \quad \sum_{k'=0}^{n-1} d_{k'l} = y_l \quad (k, l = 0, \ldots, n-1). \tag{A.4}$$

(e) If we further assume $\sum_{r=0}^{n-1} y_r = \sum_{r=0}^{n-1} x_r$, the above inequalities in (A.1)–(A.4) all reduce to equalities, i.e., the matrices $\boldsymbol{A}_n$, $\boldsymbol{B}_n$, $\boldsymbol{C}_n$ and $\boldsymbol{D}_n$ have $(x_0, \ldots, x_{n-1})$ and $(y_0, \ldots, y_{n-1})$ as their row and column sums.

Note that if all the x_i's are the same as the y_i's, we can simply construct a diagonal matrix with elements x_i's or y_i's. The following proof deals with general cases.

Proof of Lemma 1(a). We prove by induction. When $n = 1$, we let $\boldsymbol{A}_1 = y_0 \ge 0$, and Lemma 1(a) holds because $y_0 \le x_0$. When $n \ge 2$, suppose Lemma 1(a) holds for $n-1$. In particular, for any $(x_1, \ldots, x_{n-1})$ and $(y_1, \ldots, y_{n-1})$ such that $\sum_{r=s}^{n-1} x_r \ge \sum_{r=s}^{n-1} y_r$ for all $s = 1, \ldots, n-1$, there exists a lower triangular matrix $\boldsymbol{A}_{n-1} = (a_{kl})_{1 \le k,l \le n-1}$ with nonnegative elements such that

$$\sum_{l'=1}^{n-1} a_{kl'} \le x_k, \quad \sum_{k'=1}^{n-1} a_{k'l} = y_l \quad (k, l = 1, \ldots, n-1). \tag{A.5}$$

To prove that Lemma 1(a) holds for n, we let

$$\boldsymbol{A}_n = \begin{pmatrix} a_{00} & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{a} & \boldsymbol{A}_{n-1} \end{pmatrix},$$

where a_{00} and $\boldsymbol{a} = (a_{10}, \ldots, a_{n-1,0})^{\mathrm{T}}$ are defined for two separate cases below.

(1) $y_0 < x_0$. We let $a_{00} = y_0$, and $a_{k0} = 0$ for all $k = 1, \ldots, n-1$. Clearly, $\boldsymbol{A}_n$ has nonnegative elements, and satisfies the row and column sum conditions in Lemma 1(a) holds;

(2) $y_0 \geq x_0$. We let $a_{00} = x_0$, and

$$a_{k0} = (y_0 - a_{00}) \frac{x_k - \sum_{l'=1}^{n-1} a_{kl'}}{\sum_{k'=1}^{n-1} \left(x_{k'} - \sum_{l'=1}^{n-1} a_{k'l'} \right)} \geq 0 \quad (k = 1, \ldots, n-1). \tag{A.6}$$

This construction guarantees that the column sums of $\boldsymbol{A}_n$ are y_l's. Furthermore, because $\boldsymbol{A}_{n-1}$ satisfies (A.5), we have

$$\begin{aligned} \sum_{k'=1}^{n-1} \left(x_{k'} - \sum_{l'=1}^{n-1} a_{k'l'} \right) &= \sum_{k'=1}^{n-1} x_{k'} - \sum_{k'=1}^{n-1} \sum_{l'=1}^{n-1} a_{k'l'} = \sum_{k'=1}^{n-1} x_{k'} - \sum_{l'=1}^{n-1} \sum_{k'=1}^{n-1} a_{k'l'} \\ &= \sum_{k'=1}^{n-1} x_{k'} - \sum_{k'=1}^{n-1} y_{k'} \geq y_0 - x_0 = y_0 - a_{00} > 0. \end{aligned} \tag{A.7}$$

Formulas (A.6) and (A.7) imply that $a_{k0} \leq x_k - \sum_{l'=1}^{n-1} a_{kl'}$ and therefore $\sum_{l'=0}^{n-1} a_{kl'} \leq x_k$ for $k = 1, \ldots, n-1$.

Therefore Lemma 1(a) holds for n, and the proof is complete. □

Proof of Lemma 1(b). By applying Lemma 1(a) to $(y_0, \ldots, y_{n-1})$ and $(x_0, \ldots, x_{n-1})$, we obtain a lower triangular matrix $\widetilde{\boldsymbol{B_n}} = (\tilde{b}_{kl})_{0 \leq k,l \leq n-1}$ with nonnegative elements such that

$$\sum_{k'=0}^{n-1} \tilde{b}_{k'l} = x_k, \quad \sum_{l'=0}^{n-1} \tilde{b}_{kl'} \leq y_k \quad (k, l = 0, \ldots, n-1).$$

Let $\boldsymbol{B}_n = \widetilde{\boldsymbol{B_n}}^{\mathrm{T}}$, and the proof is complete. □

Proof of Lemma 1(c). By applying Lemma 1(a) to $(y_{n-1}, \ldots, y_0)$ and $(x_{n-1}, \ldots, x_0)$, we obtain a lower triangular matrix $\widetilde{\boldsymbol{C_n}} = (\tilde{p}_{kl})_{0 \leq k,l \leq n-1}$ with nonnegative elements such that

$$\sum_{k'=0}^{n-1} \tilde{c}_{k'l} = x_{n-l-1}, \quad \sum_{l'=0}^{n-1} \tilde{c}_{kl'} \leq y_{n-k-1} \quad (k, l = 0, \ldots, n-1).$$

Let $\boldsymbol{C}_n = (\tilde{c}_{n-l-1,n-k-1})_{0 \leq k,l \leq n-1}$, and the proof is complete. □

Proof of Lemma 1(d). By applying Lemma 1(c) to $(y_0, \ldots, y_{n-1})$ and $(x_0, \ldots, x_{n-1})$, we obtain a

lower triangular matrix $\widetilde{\boldsymbol{D_n}} = (\tilde{d}_{kl})_{0\le k,l\le n-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{k} \tilde{d}_{kl'} = y_k, \quad \sum_{k'=l}^{n-1} \tilde{d}_{k'l} \le x_k \quad (k, l = 0, \ldots, n-1).$$

Let $\boldsymbol{D}_n = \widetilde{\boldsymbol{D_n}}^{\mathrm{T}}$, and the proof is complete. □

Proof of Lemma 1(e). In addition to the proof of Lemma 1(a), we further need to show that if $\sum_{r=0}^{n-1} y_r = \sum_{r=0}^{n-1} x_r$, the row sums of the constructed matrix $\boldsymbol{A}_n$ are x_k's. In the induction of the proof of Lemma 1(a), if we have constructed matrix A_{n-1}, the case with $y_0 < x_0$ would not happen. We consider only the case with $y_0 \ge x_0$. Because the lower triangular matrix A_{n-1} has the column sums y_l's, and $\sum_{r=0}^{n-1} y_r = \sum_{r=0}^{n-1} x_r$, we have

$$\sum_{k'=1}^{n-1} \left(x_{k'} - \sum_{l'=1}^{n-1} a_{k'l'} \right) = \sum_{k'=1}^{n-1} x_{k'} - \sum_{k'=1}^{n-1} y_{k'} = y_0 - x_0 = y_0 - a_{00} > 0.$$

The above formula, coupled with the construction of the first column of $\boldsymbol{A}_n$ in (A.6), gives $a_{k0} = x_k - \sum_{l'=1}^{n-1} a_{kl'}$ and thus $\sum_{l'=0}^{n-1} a_{kl'} = x_k$ for all k. □

A.2. Proof of Proposition 1

Now we prove the main Proposition 1, and the proofs for other propositions and corollaries are relatively straightforward.

Proof of Proposition 1. For all $j = 0, 1, \ldots, J-1$,

$$\begin{aligned}
\tau &= \sum\sum_{k\ge l} p_{kl} = 1 - \sum\sum_{k<l} p_{kl} \\
&\le 1 - \sum_{k<j}\sum_{l\ge j} p_{kl} = 1 - \left(\sum_{k=0}^{J-1}\sum_{l\ge j} p_{kl} - \sum_{k\ge j}\sum_{l\ge j} p_{kl} \right) \qquad \text{(A.8)} \\
&\le 1 - \left(\sum_{k=0}^{J-1}\sum_{l\ge j} p_{kl} - \sum_{k\ge j}\sum_{l=1}^{J-1} p_{kl} \right) = 1 - \left(\sum_{l\ge j} p_{+l} - \sum_{k\ge j} p_{k+} \right) \qquad \text{(A.9)} \\
&= 1 + \Delta_j,
\end{aligned}$$

and

$$\begin{aligned}
\tau &= \sum_{k\geq l}\sum p_{kl} \\
&\geq \sum_{k\geq j}\sum_{l\leq j} p_{kl} = \sum_{k\geq j}\sum_{l=0}^{J-1} p_{kl} - \sum_{k\geq j}\sum_{l>j} p_{kl} \qquad \text{(A.10)} \\
&\geq \sum_{k\geq j}\sum_{l=0}^{J-1} p_{kl} - \sum_{k=0}^{J-1}\sum_{l>j} p_{kl} = \sum_{k\geq j} p_{k+} - \sum_{l>j} p_{+l} \qquad \text{(A.11)} \\
&= p_{+j} + \Delta_j,
\end{aligned}$$

which implies that $\tau_L \leq \tau \leq \tau_U$.

We now construct two probability matrices attaining the lower and upper bounds respectively, using Lemma 1.

We first construct a probability matrix attaining the upper bound τ_U. Let

$$j_1 = \min\left\{0 \leq j' \leq J-1 : \Delta_{j'} = \min_{0\leq j\leq J-1} \Delta_j\right\}$$

be the minimum index j that attains the minimum value of Δ_j's. To attain τ_U, the equalities in (A.8) and (A.9) must hold, i.e.,

$$\sum_{k<l}\sum p_{kl} = \sum_{k<j_1}\sum_{l\geq j_1} p_{kl}, \quad \sum_{k\geq j_1}\sum_{l\geq j_1} p_{kl} = \sum_{k\geq j_1}\sum_{l=1}^{J-1} p_{kl}. \qquad \text{(A.12)}$$

If $j_1 = 0$, $\min_{0\leq j\leq J-1}\Delta_j = \Delta_0 = 0$, implying that $\Delta_j = \sum_{k=j}^{J-1} p_{k+} - \sum_{l=j}^{J-1} p_{+l} \geq 0$ for all j, i.e., the marginal probabilities satisfy the stochastic dominance assumption. According to Lemma 1(e), there exists a lower triangular probability matrix $\boldsymbol{P}$ with marginal probabilities $\boldsymbol{p}_1 = (p_{0+}, \ldots, p_{J-1,+})^{\mathrm{T}}$ and $\boldsymbol{p}_0 = (p_{+0}, \ldots, p_{+,J-1})^{\mathrm{T}}$. Correspondingly, $\tau = 1 + \Delta_0 = 1$.

If $j_1 > 0$, the constraints in (A.12) force some elements of the probability matrix to be zeros. To be more specific, the constraints in (A.12) imply that the probability matrix has the following block structure:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{\mathrm{tl}} & \boldsymbol{P}_{\mathrm{tr}} \\ \boldsymbol{0} & \boldsymbol{P}_{\mathrm{br}} \end{pmatrix}, \qquad \text{(A.13)}$$

where the $j_1 \times j_1$ sub-matrix $\boldsymbol{P}_{\rm tl}$ on top left and the $(J-j_1)\times(J-j_1)$ sub-matrix $\boldsymbol{P}_{\rm br}$ on bottom right are both lower triangular, and the $j_1\times(J-j_1)$ sub-matrix $\boldsymbol{P}_{\rm tr}$ on top right has no restrictions.

Because $\Delta_{j_1} \le \Delta_j$ for all $j = 0, 1, \ldots, J-1$, we have

$$\sum_{k=j}^{j_1-1} p_{k+} \ge \sum_{l=j}^{j_1-1} p_{+l} \quad (j = 0, \ldots, j_1 - 1); \quad \sum_{k=j_1}^{j} p_{k+} \le \sum_{l=j_1}^{j} p_{+l} \quad (j = j_1, \ldots, J-1).$$

Given the above two sets of constraints on the marginal probabilities, we construct the probability matrix $\boldsymbol{P}$ in three steps.

(1) We apply Lemma 1(a) to $(p_{0+}, \ldots, p_{j_1-1,+})$ and $(p_{+0}, \ldots, p_{+,j_1-1})$, and obtain a lower triangular matrix $\boldsymbol{P}_{\rm tl} = (p_{kl})_{0\le k,l\le j_1-1}$ with nonnegative elements such that

$$\sum_{l'=0}^{j_1-1} p_{kl'} \le p_{k+}, \quad \sum_{k'=0}^{j_1-1} p_{k'l} = p_{+l} \quad (k, l = 0, \ldots, j_1 - 1).$$

(2) We apply Lemma 1(c) to $(p_{j_1+}, \ldots, p_{J-1,+})$ and $(p_{+j_1}, \ldots, p_{+,J-1})$, and obtain a lower triangular matrix $\boldsymbol{P}_{\rm br} = (p_{kl})_{j_1\le k,l\le J-1}$ with nonnegative elements such that

$$\sum_{l'=j_1}^{J-1} p_{kl'} = p_{k+}, \quad \sum_{k'=j_1}^{J-1} p_{k'l} \le p_{+l} \quad (k, l = j_1, \ldots, J-1).$$

(3) We construct $\boldsymbol{P}_{\rm tr} = (p_{kl})_{0\le k\le j_1-1, j_1\le l\le J-1}$ by letting

$$p_{kl} = \left(p_{k+} - \sum_{l'=0}^{j_1-1} p_{kl'}\right)\left(p_{+l} - \sum_{k'=j_1}^{J-1} p_{k'l}\right) \ge 0 \quad (k = 0, \ldots, j_1 - 1;\ l = j_1, \ldots, J-1).$$

The constructed probability matrix $\boldsymbol{P}$ has marginal probabilities $\boldsymbol{p}_1 = (p_{0+}, \ldots, p_{J-1,+})^{\rm T}$ and $\boldsymbol{p}_0 = (p_{+0}, \ldots, p_{+,J-1})^{\rm T}$. What is more, by (A.13) the τ of $\boldsymbol{P}$ is the sum of all the elements in $\boldsymbol{P}_{\rm tl}$ and $P_{\rm br}$, which we construct in the above (1) and (2). Therefore, we have

$$\tau = \sum_{l'=0}^{j_1-1} p_{+l'} + \sum_{k'=j_1}^{J-1} p_{k'+} = 1 + \Delta_{j_1},$$

which implies that the probability matrix $\boldsymbol{P}$ attains τ_U.

We then construct a probability matrix attaining the lower bound in τ_L. Let

$$j_2 = \min\left\{ j' : p_{+j'} + \Delta_{j'} = \max_{0 \le j \le J-1} (p_{+j} + \Delta_j) \right\}$$

be the minimum index j that attains the maximum value of $(p_{+j}+\Delta_j)$'s. To attain τ_L, the equalities in (A.10) and (A.11) must hold, i.e.,

$$\sum_{k \ge l} \sum p_{kl} = \sum_{k \ge j_2} \sum_{l \le j_2} p_{kl}, \quad \sum_{k \ge j_2} \sum_{l > j_2} p_{kl} = \sum_{k=0}^{J-1} \sum_{l > j_2} p_{kl}. \tag{A.14}$$

If $j_2 = 0$, from (A.14) we know that the elements in the lower triangular part but not in the first column of the probability matrix $\boldsymbol{P}$ are all zeros, i.e.,

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{p} & \boldsymbol{P}_{\rm tr} \\ p_{J-1,0} & 0^{\rm T} \end{pmatrix}, \tag{A.15}$$

where $\boldsymbol{p} = (p_{0,0}, \ldots, p_{J-2,0})^{\rm T}$, and the $(J-1) \times (J-1)$ sub-matrix $P_{\rm tr}$ on top right is upper triangular. Because $p_{+0} + \Delta_0 \ge p_{+j} + \Delta_j$ for all j, we have

$$\sum_{k=0}^{j} p_{k+} \ge \sum_{l=0}^{j} p_{+,l+1} \quad (j = 0, \ldots, J-2).$$

Applying Lemma 1(d) to $(p_{0+}, \ldots, p_{J-2,+})$ and $(p_{+1}, \ldots, p_{+,J-1})$, we obtain an upper triangular matrix $\boldsymbol{P}_{\rm tr} = (p_{kl})_{0 \le k \le J-2, 1 \le l \le J-1}$ with nonnegative elements such that

$$\sum_{l'=1}^{J-1} p_{kl'} \le p_{k+}, \quad \sum_{k'=0}^{J-2} p_{k'l} = p_{+l} \quad (k = 0, \ldots, J-2;\ l = 1, \ldots, J-1).$$

To complete the construction, let $p_{J-1,0} = p_{J-1,+}$, and

$$p_{k0} = p_{k+} - \sum_{l'=1}^{J-1} p_{kl'} \ge 0 \quad (k = 0, \ldots, J-2).$$

The constructed probability matrix $\boldsymbol{P}$ has marginal probabilities $\boldsymbol{p}_1 = (p_{0+}, \ldots, p_{J-1,+})^{\rm T}$ and $\boldsymbol{p}_0 = (p_{+0}, \ldots, p_{+,J-1})^{\rm T}$. Moreover, by (A.15) the τ of $\boldsymbol{P}$ is the sum of all the elements in the first

column. Therefore $\tau = p_{+0} = p_{+0} + \Delta_0$, which implies that $\boldsymbol{P}$ attains τ_L.

If $j_2 = J-1$, the proof is similar to the above case with $j_2 = 0$. If $0 < j_2 < J-1$, because the first equality in (A.14) is equivalent to

$$\sum_{k<j_2}\sum_{l\le k} p_{kl} + \sum_{k\ge j_2}\sum_{l\le k} p_{kl} = \sum_{k\ge j_2}\sum_{l\le j_2} p_{kl},$$

the probability matrix $\boldsymbol{P}$ must satisfy the following constraints:

(C1) For all $k = 0, \ldots, j_2 - 1$, $p_{kl} = 0$ for all $l = 0, \ldots, k$.

(C2) For all $k = j_2 + 1, \ldots, J-1$, $p_{kl} = 0$ for all $l = j_2 + 1, \ldots, k$.

Similarly, because the second equality in (A.14) is equivalent to

$$\sum_{k\ge j_2}\sum_{l>j_2} p_{kl} = \sum_{k\ge j_2}\sum_{l>j_2} p_{kl} + \sum_{k<j_2}\sum_{l>j_2} p_{kl},$$

the probability matrix $\boldsymbol{P}$ must further satisfy the following constraint:

(C3) $p_{kl} = 0$, for all $k = 0, \ldots, j_2 - 1$ and $l = j_2 + 1, \ldots, J-1$.

The constraints in (C1), (C2) and (C3) imply that $\boldsymbol{P}$ must have the following block structure:

$$\boldsymbol{P} = \begin{pmatrix} (\mathbf{0}, \boldsymbol{P}_{\mathrm{tl}}) & \mathbf{0} \\ \boldsymbol{P}_{\mathrm{bl}} & \begin{pmatrix} \boldsymbol{P}_{\mathrm{br}} \\ \mathbf{0}^{\mathrm{T}} \end{pmatrix} \end{pmatrix} \tag{A.16}$$

where the $j_2 \times j_2$ sub-matrix $\boldsymbol{P}_{\mathrm{tl}}$ and the $(J - j_1 - 1) \times (J - j_1 - 1)$ sub-matrix $\boldsymbol{P}_{\mathrm{br}}$ are both upper triangular, and the $(J - j_2) \times (j_2 + 1)$ sub-matrix $\boldsymbol{P}_{\mathrm{bl}}$ on bottom left has no restrictions.

Because $p_{+j_2} + \Delta_{j_2} \ge p_{+j} + \Delta_j$ for all j, we have

$$\sum_{k=j}^{j_2-1} p_{k+} \le \sum_{l=j}^{j_2-1} p_{+,l+1} \quad (j = 0, \ldots, j_2 - 1); \quad \sum_{k=j_2}^{s} p_{k+} \ge \sum_{l=j_2}^{s} p_{+,l+1} \quad (j = j_2, \ldots, J-2).$$

Given the above two sets of constraints for the marginal probabilities, we construct the probability matrix $\boldsymbol{P}$ in three steps.

(1) We apply Lemma 1(b) to $(p_{0+}, \ldots, p_{j_2-1,+})$ and $(p_{+1}, \ldots, p_{+,j_2})$, and obtain an upper triangular matrix $\boldsymbol{P}_{\mathrm{tl}} = (p_{kl})_{0 \le k \le j_2-1, 1 \le l \le j_2}$ with nonnegative elements such that

$$\sum_{l'=1}^{j_2} p_{kl'} = p_{k+}, \quad \sum_{k'=0}^{j_2-1} p_{k'l} \le p_{+l} \quad (k = 0, \ldots, j_2 - 1;\ l = 1, \ldots, j_2).$$

(2) We apply Lemma 1(d) to $(p_{j_2+}, \ldots, p_{J-2,+})$ and $(p_{+,j_2+1}, \ldots, p_{+,J-1})$, and obtain an upper triangular matrix $\boldsymbol{P}_{\mathrm{br}} = (p_{kl})_{j_2 \le k \le J-2, j_2+1 \le l \le J-1}$ with nonnegative elements such that

$$\sum_{l'=j_2+1}^{J-1} p_{kl'} \le p_{k+}, \quad \sum_{k'=j_2}^{J-2} p_{k'l} = p_{+l} \quad (k = j_2, \ldots, J-2;\ l = j_2+1, \ldots, J-1).$$

(3) We construct $\boldsymbol{P}_{\mathrm{bl}} = (p_{kl})_{j_2 \le k \le J-1, 0 \le l \le j_2}$ by letting

$$p_{kl} = \left(p_{k+} - \sum_{l'=j_2+1}^{J-1} p_{kl'} \right) \left(p_{+l} - \sum_{k'=0}^{j_2-1} p_{k'l} \right) \ge 0 \quad (k = j_2, \ldots, J-1;\ l = 0, \ldots, j_2).$$

The constructed probability matrix $\boldsymbol{P}$ has marginal probabilities $\boldsymbol{p}_1 = (p_{0+}, \ldots, p_{J-1,+})^{\mathrm{T}}$ and $\boldsymbol{p}_0 = (p_{+0}, \ldots, p_{+,J-1})^{\mathrm{T}}$. Moreover, by (A.16) the corresponding τ is the sum of all the elements in $\boldsymbol{P}_{\mathrm{bl}}$, which we construct in the above (3). Therefore,

$$\tau = 1 - \sum_{k'=0}^{j_2-1} p_{k'+} - \sum_{l'=j_2+1}^{J-1} p_{+l'} = \sum_{k'=j_2}^{J-1} p_{k'+} - \sum_{l'=j_2+1}^{J-1} p_{+l'} = p_{+j_2} + \Delta_{j_2},$$

which implies that $\boldsymbol{P}$ attains τ_L. □

A.3. Proofs of other propositions

Proof of Proposition 2. Because $\eta = 1 - \mathrm{pr}\{Y_i(0) \ge Y_i(1)\}$, its lower bound is one minus the upper bound of $\mathrm{pr}\{Y_i(0) \ge Y_i(1)\}$. By switching the treatment and control labels, we can bound $\mathrm{pr}\{Y_i(0) \ge Y_i(1)\}$ from the above by

$$\mathrm{pr}\{Y_i(0) \ge Y_i(1)\} \le 1 - \max_{0 \le j \le J-1} \Delta_j,$$

which implies that $\eta_L = \max_{0 \le j \le J-1} \Delta_j$.

Similarly, the upper bound of η equals one minus the lower bound of $\text{pr}\{Y_i(0) \geq Y_i(1)\}$. By switching the treatment and control labels, we can bound $\text{pr}\{Y_i(0) \geq Y_i(1)\}$ from below by

$$\text{pr}\{Y_i(0) \geq Y_i(1)\} \geq \max_{0 \leq j \leq J-1} (p_{j+} - \Delta_j),$$

which implies that $\eta_U = 1 + \min_{0 \leq j \leq J-1} (\Delta_j - p_{j+})$. □

Proof of Proposition 3. With independent potential outcomes, the probability matrix P has elements $p_{kl} = p_{k+}p_{+l}$ for k and l. We obtain τ_I and η_I by their definitions. Obviously, they are between their lower and upper bounds, i.e., $\tau_L \leq \tau_I \leq \tau_U$ and $\eta_L \leq \eta_I \leq \eta_U$. □

Proof of Proposition 4. The proof follows Lee (2009). Because any value of τ within the covariate adjusted bounds $[\tau'_L, \tau'_U]$ must be compatible with the distributions of $\{Y(1), \boldsymbol{X}\}$ and $\{Y(0), \boldsymbol{X}\}$, it must also be compatible with the distributions of $Y(1)$ and $Y(0)$ by discarding $\boldsymbol{X}$. Therefore, any value of τ within the adjusted bounds $[\tau'_L, \tau'_U]$ must also be within the unadjusted bounds $[\tau_L, \tau_U]$. Consequently, the adjusted bounds are tighter, i.e., $[\tau'_L, \tau'_U] \subset [\tau_L, \tau_U]$. Similar arguments apply to the covariate adjusted bounds and the unadjusted bounds for τ_c. □

Proof of Proposition 5. Under monotonicity, by the law of total probability, we have

$$\tau = \pi_c \tau_c + \pi_a \tau_a + \pi_n \tau_n.$$

Under exclusion restriction, we have $\tau_a = 1$ and $\tau_n = 1$, yielding

$$\tau = \pi_c \tau_c + 1 - \pi_c,$$

which implies that

$$\tau_c = \tau/\pi_c - (1 - \pi_c)/\pi_c.$$

Analogously, we have $\eta = \pi_c \eta_c$, which implies that $\eta_c = \eta/\pi_c$. □

A.4. Proofs of the corollaries

Proof of Corollary 1. By Proposition 1, $\tau = 1$ if and only if $\min_{0\leq j\leq J-1} \Delta_j = 0$. Because $\Delta_0 = 0$, this is equivalent to $\Delta_j \geq 0$ for all j, i.e., the stochastic dominance assumption holds. □

Proof of Corollary 2. Similar to the proof of Proposition 2, because $\eta = 1 - \text{pr}\{Y_i(0) \geq Y_i(1)\}$, (9) immediately implies (10). Therefore, we need only to prove that (9) is the sufficient and necessary condition that the lower and upper bounds of τ are the same, i.e., $\tau_L = \tau_U$.

First we prove the necessity of the condition. Assume that it does not hold, i.e., there does exist $k_1, k_2 \in \mathbb{K}$ and $l_1, l_2 \in \mathbb{L}$ such that (9) holds. In this case we construct two probability matrices with the same marginal probabilities but different values of τ. The first probability matrix is $\boldsymbol{P} = (p_{k+}p_{+l})_{0\leq k,l\leq J-1}$. For the second probability matrix, let $\xi = \min(p_{k_1,+}p_{+,l_1},\ p_{k_2,+}p_{+,l_2})$, which is a positive constant. We then apply the following matrix operation to the 2×2 sub-matrix of the first probability matrix:

$$\begin{pmatrix} p_{k_1l_1} & p_{k_1l_2} \\ p_{k_2l_1} & p_{k_2l_2} \end{pmatrix} \longrightarrow \begin{pmatrix} p_{k_1l_1} - \xi & p_{k_1l_2} + \xi \\ p_{k_2l_1} + \xi & p_{k_2l_2} - \xi \end{pmatrix}$$

The above operation preserves the marginal probabilities, and the difference of τ between the first and second probability matrices is ξ, if $k_2 \geq l_2 > k_1 \geq l_1$, and $-\xi$, if $l_2 > k_2 \geq l_1 > k_1$.

Second, we prove the sufficiency of the condition. If $|\mathbb{K}| = 1$ or $|\mathbb{L}| = 1$, the probability matrix degenerates and consequently we have $\tau_L = \tau_{c,U}$. If $|\mathbb{K}| \geq 2$ and $|\mathbb{L}| \geq 2$, let $k_* = \min_{k\in\mathbb{K}} k$ and $k^* = \max_{k\in\mathbb{K}} k$ be the minimal and maximal indices of nonzero p_{k+}'s, and $l_* = \min_{l\in\mathbb{L}} l$ and $l^* = \max_{l\in\mathbb{L}} l$ the minimal and maximal indices of nonzero p_{+l}'s. A useful fact that we repeatedly use is that if $p_{k+} = 0$, then $p_{kl} = 0$ for all l. Similarly, if $p_{+l} = 0$, then $p_{kl} = 0$ for all k.

Because k_*, k^* and l_*, l^* cannot satisfy (9), we discuss the two following cases based on the relative locations of the two intervals $[k_*, k^*]$ and $[l_*, l^*]$:

1. "Non-overlapping," i.e., $k_* \geq l^*$ or $k^* < l_*$:

   (a) If $k_* \geq l^*$, we prove that $p_{kl} = 0$ for all $k < l$. Assume the claim does not hold, then there exists $k' < l'$ such that $p_{k'l'} > 0$, then $p_{k'+} > 0$ and $p_{+l'} > 0$. This implies that $k_* \leq k' < l' \leq l^*$, contradicting the initial assumption. Therefore, $\tau_L = \tau_U = 1$.

(b) If $k^* < l_*$, similarly $p_{kl} = 0$ for all $k \geq l$, implying that $\tau_L = \tau_U = 0$.

2. "Inclusive," i.e., $l^* > k^* > k_* \geq l_*$ or $k^* \geq l^* > l_* > k_*$:

(a) If $l^* > k^* > k_* \geq l_*$, and furthermore if there exists $l' \in \mathbb{L}$ such that $k_* < l' \leq k^*$, then $l' \neq l_*$ and $l' \neq l^*$. Moreover, k_*, k^* and l', l^* satisfy (9), contradicting the initial assumption. Therefore for all $l \in \mathbb{L}$, $l \leq k_*$ or $l > k^*$. Consequently,

$$\begin{aligned}\tau &= \sum_{k \geq l} p_{kl} 1(k \in \mathbb{K}, l \in \mathbb{L}) = \sum_{k \geq l} p_{kl} 1(k \in \mathbb{K}, l \in \mathbb{L}, l \leq k_*) \\ &= \sum p_{kl} 1(k \in \mathbb{K}, l \in \mathbb{L}, l \leq k_*) = \sum_{l \leq k_*, l \in \mathbb{L}} p_{+l}\end{aligned}$$

is identifiable, which implies that $\tau_L = \tau_U$.

(b) If $k^* \geq l^* > l_* > k_*$, similarly as above for all $k \in \mathbb{K}$, $k < l_*$ or $k \geq l^*$. Consequently,

$$\tau = \sum_{k \geq l} p_{kl} 1(k \in \mathbb{K}, l \in \mathbb{L}) = \sum_{k \geq l} p_{kl} 1(k \in \mathbb{K}, l \in \mathbb{L}, k \geq l^*) = \sum_{k \geq l^*, k \in \mathbb{K}} p_{k+}$$

is identifiable, which implies that $\tau_L = \tau_U$.

□

Proof of Corollary 3. The proof follows directly from Propositions 1 and 2. □

Proof of Corollary 4. The closed-form expressions for $\tau''_{c,L}$, $\tau''_{c,U}$, $\eta''_{c,L}$ and $\eta''_{c,U}$ follow directly from Proposition 5 and Corollary 2. Furthermore, under the monotonicity and exclusion restriction assumptions, we have

$$\Delta_j = \pi_c \Delta_{c,j} \quad (j = 0, \ldots, J-1).$$

Therefore, for the upper bound of τ, we have

$$\tau_U = 1 - \pi_c + \pi_c(1 + \min \Delta_{c,j}) = \tau''_U,$$

and for the lower bound, we have

$$\tau_L \leq \max(p_{+j} - 1 + \pi_c + \pi_c \Delta_{c,j}) = 1 - \pi_c + \pi_c \max(c_{+j} + \Delta_{c,j}) = \tau''_L.$$

The first step holds because under the strong monotonicity assumption $n_{+j}\pi_n \leq \pi_n$, and under the monotonicity assumption $a_{+j}\pi_a + n_{+j}\pi_n \leq \pi_a + \pi_n$. Similar arguments apply to the bounds of η. □

B. Details of the EM Algorithm with Noncompliance

Let $\boldsymbol{X}_i = \boldsymbol{x}_i$, $Z_i = z_i$, $D_i^{\text{obs}} = d_i$ and $Y_i^{\text{obs}} = y_i$ be the values of the pretreatment covariates, treatment assigned, treatment received and observed outcome of the ith unit. We write the likelihood function as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i:z_i=1,d_i=1} \{\pi_a(\boldsymbol{x}_i)\, a_{y_i,+}(\boldsymbol{x}_i) + \pi_c(\boldsymbol{x}_i)\, c_{y_i,+}(\boldsymbol{x}_i)\} \times \prod_{i:z_i=1,d_i=0} \{\pi_n(\boldsymbol{x}_i)\, n_{y_i,+}(\boldsymbol{x}_i)\} \\ &\times \prod_{i:z_i=0,d_i=1} [\pi_a(\boldsymbol{x}_i)\, a_{+,y_i}(\boldsymbol{x}_i)] \times \prod_{i:z_i=0,d_i=0} \{\pi_n(\boldsymbol{x}_i)\, n_{+,y_i}(n;\boldsymbol{x}_i) + \pi_c(\boldsymbol{x}_i)\, c_{+,y_i}(\boldsymbol{x}_i)\}. \end{aligned}$$

Let $G_i = g_i$ be the value of the principal stratification variable of the ith unit. By treating it as missing data, we write the complete-data log-likelihood as:

$$\begin{aligned} l_{\text{C}}(\boldsymbol{\theta}) &= \sum_{i:g_i=a,z_i=1} [\log\{\pi_a(\boldsymbol{x}_i)\} + \log\{a_{y_i,+}(\boldsymbol{x}_i)\}] + \sum_{i:g_i=c,z_i=1} [\log\{\pi_c(\boldsymbol{x}_i)\} + \log\{c_{y_i,+}(\boldsymbol{x}_i)\}] \\ &+ \sum_{i:g_i=n,z_i=1} [\log\{\pi_n(\boldsymbol{x}_i)\} + \log\{n_{y_i,+}(\boldsymbol{x}_i)\}] + \sum_{i:g_i=a,z_i=0} [\log\{\pi_a(\boldsymbol{x}_i)\} + \log\{a_{+,y_i}(\boldsymbol{x}_i)\}] \\ &+ \sum_{i:g_i=c,z_i=0} [\log\{\pi_c(\boldsymbol{x}_i)\} + \log\{c_{+,y_i}(\boldsymbol{x}_i)\}] + \sum_{i:g_i=n,z_i=0} [\log\{\pi_n(\boldsymbol{x}_i)\} + \log\{n_{+,y_i}(\boldsymbol{x}_i)\}], \end{aligned}$$

We denote the realizations of (4) and (5) for the ith unit when evaluated at the true parameter value $\boldsymbol{\theta}$ as $\pi_g(\boldsymbol{x}_i)$, $g_{k+}(\boldsymbol{x}_i)$ and $g_{+l}(\boldsymbol{x}_i)$, and those when evaluated at the tth iteration of the parameter estimate $\boldsymbol{\theta}^{(t)}$ as $\pi_g^{(t)}(\boldsymbol{x}_i)$, $g_{k+}^{(t)}(\boldsymbol{x}_i)$ and $g_{+l}^{(t)}(\boldsymbol{x}_i)$. The EM algorithm proceeds as follows. Given the current tth iteration of the parameter estimate $\boldsymbol{\theta}^{(t)}$, we obtain the updated $(t+1)$th iteration $\boldsymbol{\theta}^{(t+1)}$ as follows:

1. E-Step: obtain the conditional expectation of the complete-data log-likelihood, given observed data and the current parameter estimate $\boldsymbol{\theta}^{(t)}$, by finding the (current) conditional probabilities of the principal stratum g, denoted as $\pi_{g,i}^{(t)}$.

(a) For all i such that $z_i = 1$ and $d_i = 1$, let $\pi_{n,i}^{(t)} = 0$, and

$$\pi_{g,i}^{(t)} = \frac{\pi_g^{(t)}(\boldsymbol{x}_i)\, g_{y_i,+}^{(t)}(\boldsymbol{x}_i)}{\pi_a^{(t)}(\boldsymbol{x}_i)\, a_{y_i,+}^{(t)}(\boldsymbol{x}_i) + \pi_c^{(t)}(\boldsymbol{x}_i)\, c_{y_i,+}^{(t)}(\boldsymbol{x}_i)} \quad (g = a, c).$$

(b) For all i such that $z_i = 1$ and $d_i = 0$, let $\pi_{a,i}^{(t)} = 0$, $\pi_{c,i}^{(t)} = 0$, and $\pi_{n,i}^{(t)} = 1$.

(c) For all i such that $z_i = 0$ and $d_i = 1$, let $\pi_{a,i}^{(t)} = 1$, $\pi_{c,i}^{(t)} = 0$, and $\pi_{n,i}^{(t)} = 0$.

(d) For all i such that $z_i = 0$ and $d_i = 0$, let $\pi_{a,i}^{(t)} = 0$ and

$$\pi_{g,i}^{(t)} = \frac{\pi_g^{(t)}(\boldsymbol{x}_i)\, g_{+,y_i}^{(t)}(\boldsymbol{x}_i)}{\pi_c^{(t)}(\boldsymbol{x}_i)\, c_{+,y_i}^{(t)}(\boldsymbol{x}_i) + \pi_n^{(t)}(\boldsymbol{x}_i)\, n_{+,y_i}^{(t)}(\boldsymbol{x}_i)} \quad (g = c, n).$$

2. M-Step: obtain the updated parameter estimate $\boldsymbol{\theta}^{(t+1)}$, by maximizing the conditional expectation with respect to $\boldsymbol{\theta}$. To do this, we adopt the following two-step procedure:

(a) Obtain $\boldsymbol{\theta}_{\text{PS}}^{(t+1)}$, the updated estimates of the parameters in the model for the principal strata, by maximizing the following objective function:

$$\begin{aligned} F(\boldsymbol{\theta}_{\text{PS}}) \;=\; & \sum_{i:z_i=1,d_i=1} \left\{ \pi_{a,i}^{(t)} \log \pi_a(\boldsymbol{x}_i) + \pi_{c,i}^{(t)} \log \pi_c(\boldsymbol{x}_i) \right\} \;+ \sum_{i:z_i=1,d_i=0} \pi_{n,i}^{(t)} \log \pi_n(\boldsymbol{x}_i) \\ & + \sum_{i:z_i=0,d_i=1} \pi_{a,i}^{(t)} \log \pi_a(\boldsymbol{x}_i) \;+ \sum_{i:z_i=0,d_i=0} \left\{ \pi_{c,i}^{(t)} \log \pi_c(\boldsymbol{x}_i) + \pi_{n,i}^{(t)} \log \pi_n(\boldsymbol{x}_i) \right\}. \end{aligned}$$

The optimization problem is equivalent to fitting the following weighted multinomial logistic regression:

i. For i such that $z_i = 1$ and $d_i = 1$, create two new observations for the regression: one always-taker with weight $\pi_{a,i}^{(t)}$ and one complier with weight $\pi_{c,i}^{(t)}$.

ii. For i such that $z_i = 0$ and $d_i = 0$, create two new observations: one complier with weight $\pi_{c,i}^{(t)}$ and one never-taker with weight $\pi_{n,i}^{(t)}$.

iii. For i such that $z_i = 1$ and $d_i = 0$, create one never-taker with weight 1.

iv. For i such that $z_i = 0$ and $d_i = 1$, create one always-taker with weight 1.

(b) Similarly, obtain $\boldsymbol{\theta}_{\text{PO}}^{(t+1)}$, the updated estimates of the parameters in the model for the potential outcomes, by fitting weighted proportional odds models:

i. For $g = a$, use all i such that $z_i = 1$ and $d_i = 1$ with weight $\pi_{a,i}^{(t)}$, and all i such that $z_i = 0$ and $d_i = 1$ with weight 1.

ii. For $g = c$, use all i such that $z_i = 1$ and $d_i = 1$ and all i such that $z_i = 0$ and $d_i = 0$ with weight $\pi_{c,i}^{(t)}$.

iii. For $g = n$, use all i such that $z_i = 1$ and $d_i = 0$ with weight 1, and all i such that $z_i = 0$ and $d_i = 0$ with weight $\pi_{n,i}^{(t)}$.

C. Additional Simulation Studies

C.1. The alternative method

We further examine the performances of Horowitz and Manski (2000)'s bootstrap method to construct confidence intervals for partially identified parameters, by comparing it to a more rigorous approach proposed by Jiang and Ding (2018) as follows:

1. Denote $p_{+j} + \Delta_j$ and $1 + \Delta_j$, the "building blocks" of τ_L, as L_j and U_j respectively. Obtain their finite-sample estimates $\hat{L}_j$ and $\hat{U}_j$, and let

$$\hat{q} = \text{argmax}_{0 \le j \le J-1} \hat{L}_j, \quad \hat{r} = \text{argmin}_{0 \le j \le J-1} \hat{U}_j$$

   be the minimum indices attaining the maximum value of $\hat{L}_j$'s, and the minimum value of $\hat{U}_j$'s, respectively;

2. Estimate the standard deviations of $\hat{L}_{\hat{q}}$ and $\hat{U}_{\hat{r}}$, via standard bootstrap. Denote the resulted standard errors as $\hat{\sigma}_{\hat{q}}$ and $\hat{\sigma}_{\hat{r}}$, respectively;

3. Correspondingly, the $(1 - \alpha)$ confidence interval for τ is $(\hat{L}_{\hat{q}} - C\hat{\sigma}_{\hat{q}}, \hat{U}_{\hat{r}} + C\hat{\sigma}_{\hat{r}})$, where we the threshold value C by solving the equation

$$\Phi\left\{C + \frac{\hat{U}_{\hat{r}} - \hat{L}_{\hat{q}}}{\max(\hat{\sigma}_{\hat{q}}, \hat{\sigma}_{\hat{r}})}\right\} - \Phi(-C) = 1 - \alpha.$$

By utilizing the fact that $\hat{L}_j$'s and $\hat{U}_j$'s are jointly asymptotically normal, Jiang and Ding (2018) generalized previous results by Imbens and Manski (2004) and proved that the resulted confidence interval achieves nominal coverage rate for τ.

C.2. Simulation results

We choose the sample size $N = 200$, and generate 50 simulation cases (i.e., probability matrices) by repeating the following procedure 50 times:

1. Let $U_r \overset{i.i.d}{\sim} \text{Unif}(0,1)$ for all $r = 1, \ldots, 6$;

2. Let the marginal probabilities

$$\boldsymbol{p}_1 = (U_1, U_2, U_3) / \sum_{k=1}^{3} U_k, \quad \boldsymbol{p}_0 = (U_4, U_5, U_6) / \sum_{l=4}^{6} U_l;$$

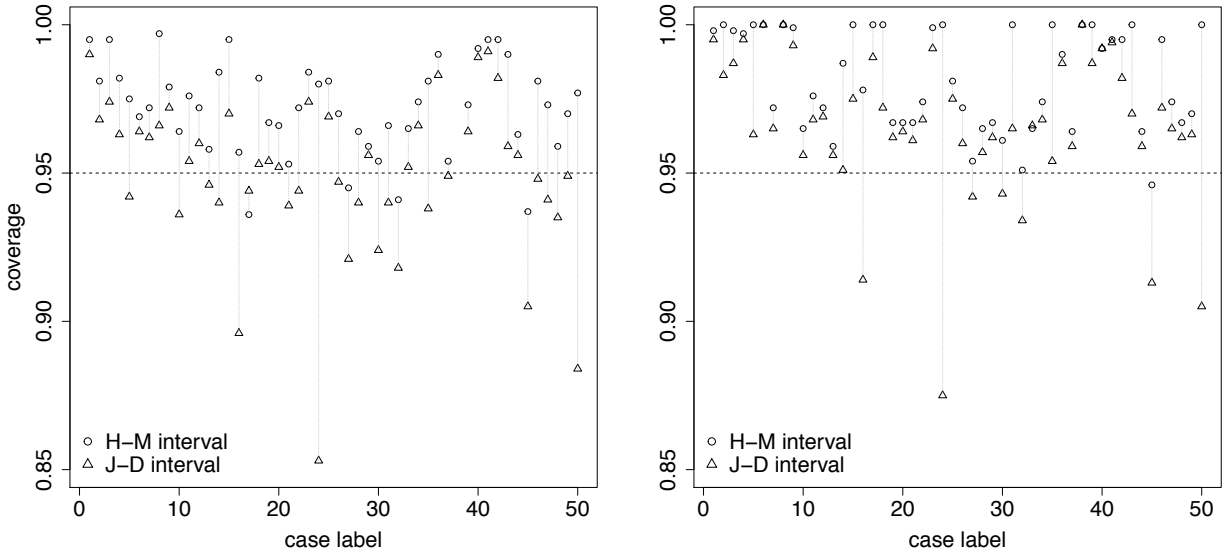
3. Based on $\boldsymbol{p}_1$ and $\boldsymbol{p}_0$ construct a probability matrix corresponding to positively correlated potential outcomes (i.e., $\tau = \tau_U$).

It is worth mentioning that, we intentionally use uniform random variables for the marginal probabilities, to "objectively" explore a wide range of potential outcome distributions. Additionally, we choose the case with positively correlated potential outcomes, because previous simulations suggest that it appear to be the most challenging.

We only focus on the coverage properties of the confidence intervals by Horowitz and Manski (2000) and Jiang and Ding (2018), respectively. Following the main text, for each of the 100 probability matrices, we independently draw 1000 treatment assignments from a balanced completely randomized experiment, and calculate two interval estimates of the bounds (τ_L, τ_U). In Figure 1 we report the coverage rates of the two confidence intervals, for both the bounds (τ_L, τ_U) and the parameter τ itself. We can draw several conclusions from the simulation results. First, both intervals achieve nominal coverage rates for the bounds (τ_L, τ_U), expect for some "edge cases" (i.e., when $\tau \approx \tau_U \approx 1$). Second, both intervals (inevitably) over-covers τ. Third, for all cases Horowitz and Manski (2000)'s bootstrapped interval performs equally well compared to Jiang and Ding (2018)'s interval, if not better. For example, for Case 16 where

$$\boldsymbol{p}_1 = (0.484, 0.134, 0.382), \quad \boldsymbol{p}_0 = (0.504, 0.295, 0.201), \quad \tau_L = 0.504, \quad \tau = \tau_U = 1,$$

Horowitz and Manski (2000)'s and Jiang and Ding (2018)'s intervals achieve 0.957 and 0.896 coverage rates for (τ_L, τ_U) respectively, and 0.978 and 0.914 for τ respectively.



(a) Coverage rates for the bounds (τ_L, τ_U).

(b) Coverage rates for the parameter τ.

Figure 1: Additional simulation results. In each subfigure, the horizontal axis denotes the simulation case labels, and the vertical axis denotes the coverage rates for the 95% Horowitz and Manski (2000) interval (denoted as "H-M interval," round dot) and the 95% Jiang and Ding (2018) interval (denoted as "J-D interval," triangular dot).

References

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

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72:1845–1857.

Jiang, Z. and Ding, P. (2018). Using missing types to improve partial identification with missing binary outcomes. *Ann. App. Stat.*, in press. arXiv preprint:1610.01198.

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.*, 36:423–439.