Statistica Sinica Preprint No: SS-2025-0495	
Title	Sharp Bounds and Covariate Adjustment for Ordinal Outcomes in Stratified Randomized Experiments
Manuscript ID	SS-2025-0495
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0495
Complete List of Authors	Guangyuran Liao and Yuehan Yang
Corresponding Authors	Yuehan Yang
E-mails	yyh@cufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

SHARP BOUNDS AND COVARIATE ADJUSTMENT FOR ORDINAL OUTCOMES IN STRATIFIED RANDOMIZED EXPERIMENTS

Guangyuran Liao, Yuehan Yang*

Central University of Finance and Economics

Abstract: Assessing causal effects of interventions on ordinal outcomes is essential in economics, biological experiments, and behavioral studies. In the potential outcomes framework, average treatment effects are defined by comparing outcomes under treatment and control conditions, but ordinal outcomes are often difficult to interpret. To address this, we consider the relative treatment effect, defined as the difference between the probability that the treatment is beneficial and the probability that it is harmful. However, it depends on the association between potential outcomes, which is not identifiable from the observed data. We derive sharp bounds on the relative treatment effect within a stratified randomization framework, given the marginal distributions of the potential outcomes. Moreover, we allow for the special case of a single stratum, and generalize the approach to handle scenarios with binary and multi-category potential outcomes. We enhance estimation precision through covariate adjustment and further narrow the bounds, and also consider cases with independent potential outcomes. We construct nonparametric bootstrap confidence intervals for sharp bounds and evaluate the proposed estimator via numerical simulations and an empirical analysis.

Key words and phrases: Causal inference, covariate adjustment, ordinal outcomes, partial identification, stratification, sharp bound.

1. Introduction

Ordinal outcomes are frequently recorded in clinical trials and other empirical studies where responses admit a natural ordering but lack a meaningful numeric scale. For example, in clinical studies of cirrhosis, outcomes are often classified into six ordered categories from "no effect" to "death" (D'Amico et al., 2020); similar ordinal endpoints also arise in social and

*Corresponding author. E-mail: yyh@cufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

educational research (Scott et al., 1997; Praet and Desoete, 2014; Hoff, 2009). In many such studies, treatment is assigned using covariate-adaptive randomization schemes, with stratified block randomization as a widely adopted special case, to improve covariate balance across arms (Rosenberger and Sverdlov, 2008; Ma et al., 2020). Under these designs, randomization justifies design-based causal inference in the Neyman-Rubin potential outcomes framework (Splawa-Neyman et al., 1990; Rubin, 1974), and motivates covariate-adjusted analyses that exploit the randomization structure to improve efficiency while maintaining validity (Lin, 2013).

In randomized experiments, particularly with stratified randomized designs, stratification is a common approach to improve the precision of causal estimates. Following Fisher (1992), stratification partitions experimental units into groups based on prognostic variables and then randomizes treatment within each group. This design improves covariate balance across treatment groups and enhances efficiency when stratification variables are correlated with the outcome (Miratrix et al., 2013; Imai, 2008; Liu and Yang, 2020). In practice, especially with many covariates, residual imbalances may remain after stratification. Covariate (regression) adjustment at the analysis stage is a prevalent strategy to address these residual imbalances and further increase precision and robustness (Morgan and Rubin, 2015; Yue et al., 2019; Bloniarz et al., 2016). These considerations are particularly important with ordinal outcomes, where precision is often critical for interpretation.

For ordinal outcomes, the average treatment effect, while widely used under the stable unit treatment value assumption (Rubin, 1980), may fail to respect the inherent ordering of categories or may misrepresent differences between them, since distances between adjacent categories need not reflect differences between their distributions. A useful alternative is the relative treatment effect proposed by Agresti and Kateri (2017) and Lu et al. (2020). Although this parameter is well defined for any outcome type and is especially relevant for ordinal data, it depends on the association between potential outcomes, which is not identifiable from observed data. To address such non-identification, partial identification and sensitivity analysis formalize how bounds can be sharpened under minimal assumptions

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

(Manski, 2003; Richardson et al., 2014). Related developments include explicit expressions for bounds on other causal parameters (Lu et al., 2018), Bayesian approaches that place priors on the joint distribution of potential outcomes (Chiba, 2018; Volfovsky et al., 2015), and methods that dispense with proportional odds assumptions for ordinal outcomes (Díaz et al., 2016).

These studies provide valuable insights into the derivation of causal parameter bounds with ordinal outcomes. Therefore, we propose novel strategies under a stratified randomized framework, leveraging the inherent advantages of stratification to refine bounds and enhance estimation accuracy. Our approach integrates covariate adjustment methods with stratified designs to systematically narrow the bound ranges and construct more reliable confidence intervals for treatment effects. These methodological advancements offer a robust framework for analyzing relative treatment effects, particularly in contexts involving ordinal outcomes, where achieving precision and interpretability is essential.

Our first contribution is the derivation of a closed-form expression for the sharp bounds on the relative treatment effect within a stratified randomized framework, achieved without additional parametric outcome models or distributional assumptions. Our framework includes the single-stratum case, including Lu et al. (2018) as a special case, and it unifies binary and multi-category outcomes. Furthermore, Ding and Dasgupta (2016) discusses binary outcomes under completely randomized experiments from a Bayesian perspective. Wu et al. (2024) quantify individual risk through the fraction of negative effects (FNA) under binary outcomes and derive precise bounds that surpass the Fréchet-Hoeffding bounds. Compared to existing literature that utilizes partial identification strategies, our approach tightens the identified set and improves the precision of bound estimators for ordinal outcomes by leveraging stratification to balance covariates. Stratified randomization first stratifies experimental units into strata based on stratification variables and then assigns treatment using block randomization within each stratum. Recent surveys indicate that approximately 70% of randomized trials have utilized stratified randomization (Lin et al., 2015; Ciolino et al., 2019). Additionally, earlier works such as Huang et al. (2017) and Balke and Pearl (1997)

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

relied on numerical approaches to compute bounds, whereas we provide explicit analytical solutions for the sharp upper and lower bounds for the relative treatment effect. Zhang et al. (2024) defined and identified the probability of necessity under ordinal outcomes with the assumption of monotonic effects and partial identification strategies. Our approach can be adapted to such parameters. To further enrich this analysis, we also consider the case of (conditionally) independent potential outcomes given covariates and strata, a commonly used benchmark in the study of treatment effects (Agresti, 2010; Cheng, 2009; Rubin, 1978; Agresti and Kateri, 2017).

Our second contribution addresses the efficiency of treatment effect estimation through covariate adjustment. Unadjusted bounds on treatment effects are often overly conservative. To improve efficiency, we estimate conditional relative treatment effects given covariates within the stratified randomization framework, averaging these conditional estimates over the empirical distribution of covariates. This approach aligns with prior methods that leverage covariates to refine bounds (Lee, 2009; Jiang and Ding, 2018). Under common conditions, we demonstrate that covariate adjustment narrows the bounds, yielding more reliable and precise estimates of the treatment effects. Furthermore, we construct nonparametric bootstrap confidence intervals for these treatment effects, enhancing the practical utility of our proposed methods.

The article proceeds as follows. Section 2 presents the framework and notation for ordinal results in randomized stratified experiments. Section 3 gives a theoretical derivation of the bounds estimator for relative effects. Section 4 shows how the construction of probability matrix is a precondition for obtaining bounds on relative treatment effects. Section 5 proposes covariate adjustment to improve the efficiency of the estimation and validates them with propositions and examples. Section 6 provides an extensive simulation study. Section 7 contains an application to a pilot dataset of cardia cancer patients. Section 8 concludes the article with a discussion. Proofs are given in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

2. Causal inference with ordinal outcomes in stratified randomized experiments

In this section, we follow the framework and notation introduced in Liu and Yang (2020) for randomized experiments. Moreover, we introduce notation for the design stage, incorporating stratification.

Define $\{Y_i(1), Y_i(0)\}$ as the potential outcomes of unit $i = 1, \ldots, n$ under treatment and control respectively. Consider ordinal outcomes labeled as $0, \ldots, J-1$, where 0 and $J-1$ represent the worst and best categories respectively. Let Z_i denote an indicator for whether the ith unit receives the treatment ($Z_i = 1$) or the control ($Z_i = 0$). Throughout this context, we use '1' to denote treatment and '0' to denote control. We analyze a stratified randomized experiment involving n units consisting of M strata. Within each stratum, independent completely randomized experiments are conducted, allowing varying relative sizes of treatment and control groups. Let $n_m = \sum_{i=1}^{n} I(B_i = m)$ denote the number of units in block m, where $I(\cdot)$ is the indicator function and $B_i \in \{1, \ldots, M\}$ is the stratification label for unit i. Let $n_{m[1]} = \sum_{i:B_i=m} Z_i$ denote the number of individuals receiving the treatment in block m. Since each unit is assigned to either the treatment or control group but not both, we cannot observe both potential outcomes simultaneously, and the observed outcome for unit i is thus given by:

$$Y_i^{\text{obs}} = Z_i Y_i(1) + (1 - Z_i)\, Y_i(0).$$

In this case, for all $k, l = 0, ..., J-1$, the probability that the potential outcome is k under treatment and l under control within block m is denoted by

$$p_{kl}^{[m]} = \Pr\left\{Y_i\left(1\right) = k, Y_i\left(0\right) = l | B_i = m\right\}.$$

The label $[m]$ denotes stratum m. We then construct the probability matrix for stratum m as $L_{[m]} = \{p_{kl}^{[m]}\}_{k,l=0,\ldots,J-1} \in \mathbb{R}^{J\times J}$, which describes the joint distribution of the potential

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

outcomes. By definition, the following constraints must hold:

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

denoting the row and column sums of $L_{[m]}$ by $p_{k+}^{[m]} = \Pr\{Y_i(1) = k | B_i = m\} = \sum_{l=0}^{J-1} p_{kl}^{[m]}$ and $p_{+l}^{[m]} = \Pr\{Y_i(0) = l | B_i = m\} = \sum_{k=0}^{J-1} p_{kl}^{[m]}$. The arrangement of joint probabilities $p_{kl}^{[m]}$ corresponds to a distinct matrix $L_{[m]}$, and the feasible set formed by these matrices is denoted by

$$\mathcal{F}_{[m]} = \left\{ L_{[m]} = \left\{ p_{kl}^{[m]} \right\}_{J\times J} : p_{kl}^{[m]} \geqslant 0, \sum_{l=0}^{J-1} p_{kl}^{[m]} = p_{k+}^{[m]}, \sum_{k=0}^{J-1} p_{kl}^{[m]} = p_{+l}^{[m]} \right\},$$

such that matrix $L_{[m]} \in \mathcal{F}_{[m]}$ is compatible with given marginal probabilities $p_{k+}^{[m]}$ and $p_{+l}^{[m]}$. By definition, the set $\mathcal{F}_{[m]}$ is defined by constraints that are closed and bounded. Therefore, in a finite-dimensional Euclidean space, $\mathcal{F}_{[m]}$ is compact. Due to the specificity of ordinal outcomes combined with stratified randomized experiments, the relative treatment effect in stratum m under joint distribution $L_{[m]}$ is defined as:

$$\begin{aligned}\gamma\left(L_{[m]}\right) &= \Pr\{Y_i(1) > Y_i(0) \,|B_i = m\} - \Pr\{Y_i(1) < Y_i(0) \,|B_i = m\} \\ &= \sum_{k=0}^{J-1}\sum_{l=0}^{J-1} \operatorname{sgn}(k-l)\, p_{kl}^{[m]},\end{aligned}$$

where $\operatorname{sgn}(\cdot)$ is the sign function. For simplicity, we omit the dependence on $L_{[m]}$ and denote the stratified relative treatment effect as $\gamma_{[m]}$. The population relative treatment effect is

$$\gamma = \sum_{m=1}^{M}\sum_{k=0}^{J-1}\sum_{l=0}^{J-1} \pi_{[m]} \operatorname{sgn}(k-l)\, p_{kl}^{[m]} = \sum_{m=1}^{M} \pi_{[m]}\gamma_{[m]},$$

where $\pi_{[m]} = \Pr\{B_i = m\}$ is the proportion of units in stratum m. From the definition, $\gamma_{[m]}$ represents the difference between the sum of the elements below the main diagonal of the probability matrix $L_{[m]}$ and the sum of the elements above it. When considering stratified randomized experiments, there is a distinction between the cases of $J = 2$ and $J \geqslant 3$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

When $J = 2$, it is established that the treatment effect $\gamma_{[m]}$ is related to the traditional average treatment effect $\tau_{[m]}$. For stratum m, define the stratum-specific average treatment effect as $\tau_{[m]} = \left(n_{[m]}\right)^{-1} \sum_{i:B_i=m} \left\{Y_i(1) - Y_i(0)\right\}$, the population average treatment effect is $\tau = \sum_{m=1}^{M} \pi_{[m]}\tau_{[m]}$, and Liu and Yang (2020) thoroughly discusses the theoretical properties of $\tau_{[m]}$. Proposition 1 is presented below.

Proposition 1. *When $J = 2$, the stratified relative treatment effect $\gamma_{[m]}$ and the stratified average treatment effect $\tau_{[m]}$ satisfy $\gamma_{[m]} = \tau_{[m]}$, and their respective population treatment effects satisfy $\gamma = \tau$.*

When the potential outcomes are binary, Proposition 1 establishes equivalence between $\gamma_{[m]}$ and $\tau_{[m]}$ (see Supplementary S1). Since $\gamma_{[m]}$ depends solely on the marginal distributions of the potential outcomes, it can be directly estimated from the observed data when $J = 2$ via $\tau_{[m]}$. However, when $J \geqslant 3$, $\gamma_{[m]}$ becomes inestimable, as it relies on the joint distribution of the treated and control potential outcomes. Thus, we define the sharp upper and lower bounds, $\gamma^U_{[m]}$ and $\gamma^L_{[m]}$, as follows. For the joint probability matrix $L_{[m]} = \{p^{[m]}_{kl}\}_{k,l=0,\ldots,J-1}$, where $m = 1, \ldots, M$, the bounds are expressed as:

$$\begin{aligned}
\gamma^U_{[m]} &= \max_{L_{[m]}} \sum_{k=0}^{J-1} \sum_{l=0}^{J-1} \operatorname{sgn}(k-l)\, p^{[m]}_{kl}, \\
\gamma^L_{[m]} &= \min_{L_{[m]}} \sum_{k=0}^{J-1} \sum_{l=0}^{J-1} \operatorname{sgn}(k-l)\, p^{[m]}_{kl}, \\
&\text{subject to} \sum_{l'=0}^{J-1} p^{[m]}_{kl'} = p^{[m]}_{k+}; \sum_{k'=0}^{J-1} p^{[m]}_{k'l} = p^{[m]}_{+l}; p^{[m]}_{kl} \geqslant 0.
\end{aligned}$$

By definition, $\gamma_{[m]}$ is a function of the joint probabilities $p^{[m]}_{kl}$, equivalently a function of the matrix $L_{[m]} = \{p^{[m]}_{kl}\}$. By optimizing over all joint distribution of potential outcomes $L_{[m]}$, we can obtain the sharp upper bound $\gamma^U_{[m]}$ and lower bound $\gamma^L_{[m]}$ by maximizing and minimizing $\gamma_{[m]}$ under the constraints of row sums, column sums and non-negative probabilities.

To derive an explicit expression for the bounds of $\gamma_{[m]}$, we obtain a quantity $\delta^{[m]}_{k'l'}$ by applying the Fréchet-Hoeffding bounds (Fréchet, 1935; Rüschendorf, 1981), which yields an

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

upper bound on $\gamma_{[m]}$, as shown in the following Proposition 2. Moreover, Huang et al. (2017) obtained numerical bounds for other different causal parameters under ordinal outcomes in linear programming problems.

Proposition 2. *For each fixed* $k' = 1, \ldots, J-1$ *and* $l' = 1, \ldots, J-k'$, *define* $\delta_{k'l'}^{[m]}$ *as*

$$\delta_{k'l'}^{[m]} = \sum_{k=k'}^{J-1} p_{k+}^{[m]} + \sum_{k=k'+l'}^{J-1} p_{k+}^{[m]} + \sum_{l=0}^{k'-2} p_{+l}^{[m]} - \sum_{l=k'+l'-1}^{J-1} p_{+l}^{[m]},$$

where the summation equals zero when the range is empty. Furthermore, for all $m = 1, \ldots, M$, $\gamma_{[m]} \leqslant \delta_{k'l'}^{[m]}$.

The proof of Proposition 2 is provided in the Supplementary S1. Based on Fréchet-Hoeffding bounds, although the joint probability $p_{kl}^{[m]}$ is not identifiable from observed data, we obtain a loose upper bound $\delta_{k'l'}^{[m]}$ for $\gamma_{[m]}$ through marginal information $p_{k+}^{[m]}$ and $p_{+l}^{[m]}$. For illustration, we provide a straightforward probability calculation. Note that we defined the treatment effect as

$$\gamma_{[m]} = \Pr\left\{Y_i\left(1\right) > Y_i\left(0\right) | B_i = m\right\} - \Pr\left\{Y_i\left(1\right) < Y_i\left(0\right) | B_i = m\right\}.$$

For the first term on the right hand of the above equation, we have

$$\begin{aligned}
&\Pr\{Y_i(1) > Y_i(0) | B_i = m\} \\
&= \Pr\{Y_i(1) > Y_i(0), Y_i(0) = 0 | B_i = m\} + \Pr\{Y_i(1) > Y_i(0), Y_i(0) \geqslant 1 | B_i = m\} \\
&\leqslant \Pr\{Y_i(1) \geqslant 1, Y_i(0) = 0 | B_i = m\} + \Pr\{Y_i(1) \geqslant 2, Y_i(0) \geqslant 1 | B_i = m\} \\
&\leqslant \Pr\{Y_i(1) \geqslant 1 | B_i = m\} + \Pr\{Y_i(1) \geqslant 2 | B_i = m\} - \Pr\{Y_i(1) \geqslant 1, Y_i(0) \geqslant 1 | B_i = m\}.
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Similarly, for $\Pr\{Y_i(1) < Y_i(0)|B_i = m\}$, we have

$$\begin{aligned}\Pr\{Y_i(1) < Y_i(0)|B_i = m\} &\geqslant \Pr\{Y_i(1) < Y_i(0), Y_i(1) = 0|B_i = m\} \\ &= \Pr\{Y_i(1) = 0, Y_i(0) \geqslant 1|B_i = m\} \\ &= \Pr\{Y_i(0) \geqslant 1|B_i = m\} - \Pr\{Y_i(1) \geqslant 1, Y_i(0) \geqslant 1|B_i = m\}.\end{aligned}$$

Combining the two inequalities above, the treatment effect $\gamma_{[m]}$ has the following upper bound

$$\begin{aligned}\gamma_{[m]} &\leqslant \Pr\{Y_i(1) \geqslant 1|B_i = m\} + \Pr\{Y_i(1) \geqslant 2|B_i = m\} - \Pr\{Y_i(0) \geqslant 1|B_i = m\} \\ &= \sum_{k=1}^{J-1} p_{k+}^{[m]} + \sum_{k=2}^{J-1} p_{k+}^{[m]} - \sum_{l=1}^{J-1} p_{+l}^{[m]} \\ &= \delta_{11}^{[m]},\end{aligned}$$

where the last equality holds follows from the definition of $\delta_{k'l'}^{[m]}$. As is evident from the calculations above, $\delta_{11}^{[m]}$ serves as a loose upper bound for $\gamma_{[m]}$. In the next section, we derive the sharp bounds for $\gamma_{[m]}$ building on Proposition 2.

3. Theoretical guarantees of sharp bounds

In this section, we provide theoretical guarantees for the sharp bounds on the relative treatment effect for ordinal outcomes with stratification. We construct an optimal probability matrix $L_{[m]}^*$ in the feasible set $\mathcal{F}_{[m]}$ and show the sharpness of $\gamma(L_{[m]}^*)$ (the construction are given in Section 4). We derive explicit sharp bounds for $\gamma_{[m]}$ without additional parametric or distributional assumptions. To broaden the scope, we further investigate the case where the potential outcomes are independent, offering additional insights into this commonly studied scenario. We provide the following results.

Theorem 1. *Consider stratified randomized experiments with ordinal outcomes having J*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

categories, $J \geqslant 3$. For any feasible joint probability matrix $L_{[m]} \in \mathcal{F}_{[m]}$, we have

$$\gamma(L_{[m]}) \leqslant \gamma^U_{[m]} := \max_{L_{[m]} \in \mathcal{F}_{[m]}} \gamma(L_{[m]}) = \min_{1 \leqslant k' \leqslant J-1} \min_{1 \leqslant l' \leqslant J-k'} \delta^{[m]}_{k'l'}.$$

Moreover, the maximum is attained: there exists at least one optimal solution $L^*_{[m]} \in \mathcal{F}_{[m]}$ with $\gamma(L^*_{[m]}) = \gamma^U_{[m]}$. At the population level, with stratum weights $\pi_{[m]}$, the sharp upper bound is

$$\gamma^U = \sum_{m=1}^{M} \pi_{[m]} \gamma^U_{[m]}.$$

The proof of Theorem 1 is provided in the Supplementary S4. The theorem presents three points. First, $\delta^{[m]}_{k'l'}$ serves as an upper bound for $\gamma_{[m]}$, ensuring that $\gamma_{[m]} \leqslant \delta^{[m]}_{k'l'}$. Second, compactness of $\mathcal{F}_{[m]}$ implies that the sharp upper bound $\gamma^U_{[m]}$ is attained by some $L^*_{[m]}$ (with the construction given in Section 4). Third, because γ is a stratum-weighted average and the feasible sets factor by stratum, the population sharp upper bound equals $\sum_m \pi_{[m]} \gamma^U_{[m]}$.

Following Theorem 1, we derive the sharp lower bound on the treatment effect $\gamma_{[m]}$ by introducing an auxiliary quantity, denoted as $\xi^{[m]}_{k'l'}$. Using arguments analogous to those in the proof of Theorem 1, we obtain the following theorem for the lower bound $\gamma^L_{[m]}$.

Theorem 2. *Consider ordinal outcomes with $J \geqslant 3$. For any feasible probability matrix $L_{[m]} \in \mathcal{F}_{[m]}$, we have*

$$\gamma(L_{[m]}) \geqslant \gamma^L_{[m]} := \min_{L_{[m]} \in \mathcal{F}_{[m]}} \gamma(L_{[m]}) = \max_{1 \leqslant k' \leqslant J-1} \max_{1 \leqslant l' \leqslant J-k'} \xi^{[m]}_{k'l'}.$$

For all $k' = 1, \ldots, J-1$,$l' = 1, \ldots, J-k'$, and $m = 1, \ldots, M$,

$$\xi^{[m]}_{k'l'} = \sum_{k=k'+l'-1}^{J-1} p^{[m]}_{k+} - \sum_{k=0}^{k'-2} p^{[m]}_{k+} - \sum_{l=k'}^{J-1} p^{[m]}_{+l} - \sum_{l=k'+l'}^{J-1} p^{[m]}_{+l}.$$

If any summation range is empty, the corresponding sum is zero. Moreover, the minimum is attained: there exists at least one optimal matrix $L'_{[m]} \in \mathcal{F}_{[m]}$ with $\gamma\left(L'_{[m]}\right) = \gamma^L_{[m]}$. The lower sharp bound of the population treatment effect is $\gamma^L = \sum_{m=1}^{M} \pi_{[m]} \gamma^L_{[m]}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Remark 1. *By Theorems 1 and 2, since $\mathcal{F}_{[m]}$ is convex and compact and $\gamma(\cdot)$ is linear in $L_{[m]}$, the image set $\{\gamma(L_{[m]}) : L_{[m]} \in \mathcal{F}_{[m]}\}$ equals the closed interval $[\gamma^L_{[m]}, \gamma^U_{[m]}]$, and both endpoints are attained at extreme points of $\mathcal{F}_{[m]}$. In particular,*

$$\gamma^U_{[m]} = \max_{L_{[m]} \in \mathcal{F}_{[m]}} \gamma(L_{[m]}) = \min_{1 \leqslant k' \leqslant J-1} \min_{1 \leqslant l' \leqslant J-k'} \delta^{[m]}_{k'l'}, \quad \gamma^L_{[m]} = \min_{L_{[m]} \in \mathcal{F}_{[m]}} \gamma(L_{[m]}) = \max_{1 \leqslant k' \leqslant J-1} \max_{1 \leqslant l' \leqslant J-k'} \xi^{[m]}_{k'l'}.$$

The constructions in Propositions 3 and 4 yield feasible matrices that attain $\gamma^U_{[m]}$ and $\gamma^L_{[m]}$, respectively. At the population level, the identified set is $\left[\sum_{m=1}^{M} \pi_{[m]}\gamma^L_{[m]}, \sum_{m=1}^{M} \pi_{[m]}\gamma^U_{[m]}\right]$. When $J = 2$, the interval degenerates to the point $\gamma_{[m]} = \tau_{[m]}$.

The proof of Theorem 2 is provided in the Supplementary S5. Based on the theoretical results, we illustrate the relationship between $\gamma^U_{[m]}$, $\gamma^L_{[m]}$, and potential outcomes through the following example. For example, when $J = 3$, we verify that

$$\delta^{[m]}_{11} = \sum_{k=1}^{2} p^{[m]}_{k+} + \sum_{k=2}^{2} p^{[m]}_{k+} - \sum_{l=1}^{2} p^{[m]}_{+l} = p^{[m]}_{1+} + 2p^{[m]}_{2+} - p^{[m]}_{+1} - p^{[m]}_{+2}.$$

Similarly,

$$\delta^{[m]}_{12} = p^{[m]}_{2+} + p^{[m]}_{1+} - p^{[m]}_{+2}, \quad \text{and} \quad \delta^{[m]}_{21} = p^{[m]}_{2+} + p^{[m]}_{+0} - p^{[m]}_{+2}.$$

According to Theorem 1, the sharp upper bound on $\gamma_{[m]}$ is given by

$$\gamma^U_{[m]} = \min\left(p^{[m]}_{1+} + 2p^{[m]}_{2+} - p^{[m]}_{+1} - p^{[m]}_{+2}, p^{[m]}_{2+} + p^{[m]}_{1+} - p^{[m]}_{+2}, p^{[m]}_{2+} + p^{[m]}_{+0} - p^{[m]}_{+2}\right),$$

and the sharp upper bound on γ is

$$\gamma^U = \sum_{m=1}^{M} \pi_{[m]}\gamma^U_{[m]}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

The sharp lower bound is

$$\gamma_{[m]}^{L} = \max\left(p_{1+}^{[m]} + p_{2+}^{[m]} - p_{+1}^{[m]} - 2p_{+2}^{[m]}, p_{2+}^{[m]} - p_{+1}^{[m]} - p_{+2}^{[m]}, p_{2+}^{[m]} - p_{0+}^{[m]} - p_{+2}^{[m]}\right).$$

The sharp lower bound on γ is $\gamma^{L} = \sum_{m=1}^{M} \pi_{[m]}\gamma_{[m]}^{L}$. This process can be extended to derive results for other values of J. This example shows that the sharpness of the upper bound $\gamma_{[m]}^{U}$ is attained by concentrating the $p_{kl}^{[m]}$ in the lower-left region of matrix $L_{[m]}^{*}$, corresponding to positive association between the potential outcomes $Y_i(1)$ and $Y_i(0)$. The sharpness of the lower bound $\gamma_{[m]}^{L}$ is attained by concentrating the $p_{kl}^{[m]}$ in the upper-right part of the matrix $L_{[m]}'$, similarly corresponding to the negative correlation between $Y_i(1)$ and $Y_i(0)$. To broaden the scope of this study, we further investigate the case where the potential outcomes are independent, and next state the corresponding result.

Corollary 1. *If the potential outcomes are independent in stratified randomized experiments, that is, $p_{kl}^{[m]} = p_{k+}^{[m]}p_{+l}^{[m]}$, the stratified treatment effect is point identified as*

$$\gamma_{[m]}^{ind} = \sum_{k>l}\sum p_{k+}^{[m]}p_{+l}^{[m]} - \sum_{l>k}\sum p_{k+}^{[m]}p_{+l}^{[m]} = \sum_{k=0}^{J-1}\sum_{l=0}^{J-1} \mathrm{sgn}(k-l)p_{k+}^{[m]}p_{+l}^{[m]},$$

and the population relative treatment effect γ^{ind} under independent potential outcomes equals

$$\gamma^{ind} = \sum_{m=1}^{M} \pi_{[m]}\gamma_{[m]}^{ind}.$$

Furthermore, we have $\gamma^{L} \leqslant \gamma^{ind} \leqslant \gamma^{U}$.

Remark 2. *Corollary 1 distinguishes two uses of independence. If independence of potential outcomes is imposed as the maintained identifying assumption on the joint distribution, then $p_{kl}^{[m]} = p_{k+}^{[m]}p_{+l}^{[m]}$, so that*

$$\gamma_{[m]} = \sum_{k=0}^{J-1}\sum_{l=0}^{J-1} \mathrm{sgn}(k-l)p_{k+}^{[m]}p_{+l}^{[m]} = \gamma_{[m]}^{\mathrm{ind}},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

and after averaging over strata $\gamma = \gamma^{\text{ind}}$*. Under this assumption, the relative treatment effect is point identified and the sharp identified set degenerates to* $\{\gamma^{\text{ind}}\}$*.*

Without imposing independence on the joint distribution, however, $p_{kl}^{[m]} = p_{k+}^{[m]} p_{+l}^{[m]}$ *corresponds only to one particular feasible coupling among all joint distributions compatible with the marginal probabilities, and* γ^{ind} *is the value of* γ *evaluated at this single coupling. Corollary 1 then gives only the inclusion* $\gamma^L \leq \gamma^{\text{ind}} \leq \gamma^U$*, and we do not claim* $\gamma \geq \gamma^{\text{ind}}$*. The sharp identified interval under arbitrary dependence remains* $[\gamma^L, \gamma^U]$*.*

We additionally report $[\gamma^{\text{ind}}, \gamma^U]$ in Sections 6 and 7 as an auxiliary independence-based reference interval, following Cheng (2009); Agresti and Kateri (2017). The motivation is interpretive rather than identification-theoretic: the sharp lower endpoint γ^L is attained only under strong negative dependence between the two potential outcomes of the same unit, which is often regarded as implausible, and related work restricts attention to nonnegative dependence structures (Ding et al., 2019; Frandsen and Lefgren, 2021). The independence-based value γ^{ind} thus provides a transparent reference point determined entirely by the marginals. We emphasize, however, that $[\gamma^{\text{ind}}, \gamma^U]$ is reported only as a reference interval; it is not the sharp identified set, and no restriction beyond the marginals is imposed on the joint distribution in our main theory or main analysis. In the next section, we detail the construction of joint probability matrices that attain the sharp bounds.

4. Construction of the optimal probability matrix

This section presents a detailed examination of the upper and lower bounds for the treatment effect $\gamma_{[m]}$. The sharpness of the bounds $\gamma_{[m]}^U$ and $\gamma_{[m]}^L$ in Theorems 1 and 2 is attained by the specific allocation of probabilities $p_{kl}^{[m]}$ in the optimal joint probability matrices $L_{[m]}^*$ and $L_{[m]}'$, respectively. Accordingly, Proposition 3 gives an explicit construction of a lower triangular joint probability matrix, and Proposition 4 describes three allocation cases for the cell probabilities $p_{kl}^{[m]}$ that attain the sharp bounds. We now state Proposition 3.

Proposition 3. *Given two nonnegative vectors* $\boldsymbol{c}_1 = \left(c_{10}, \ldots, c_{1(J-1)}\right)^{\mathrm{T}}$, $\boldsymbol{c}_2 = \left(c_{20}, \ldots, c_{2(J-1)}\right)^{\mathrm{T}}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

For $s = 0, \ldots, J-1$, *define the tail sums* $C_1^t(s) := \sum_{r=s}^{J-1} c_{1r}$, $C_2^t(s) := \sum_{r=s}^{J-1} c_{2r}$, *and the prefix sums* $C_1^p(s) := \sum_{r=0}^{s} c_{1r}$, $C_2^p(s) := \sum_{r=0}^{s} c_{2r}$. *(1) If the tail capacity condition* $C_1^t(s) \geqslant C_2^t(s)$ *holds for all* $s = 0, \ldots, J-1$, *then there exists a* $J \times J$ *nonnegative lower triangular matrix* $\mathbf{A}_J = (a_{kl})_{0 \leqslant k,l \leqslant J-1}$ *such that*

$$\sum_{l'=0}^{J-1} a_{kl'} \leqslant c_{1k}, \quad \sum_{k'=0}^{J-1} a_{k'l} = c_{2l} \quad (k, l = 0, \ldots, J-1).$$

Moreover, if $C_1^t(0) = C_2^t(0)$, *then* $\sum_{l'=0}^{J-1} a_{kl'} = c_{1k}$.
(2) If the prefix capacity condition $C_1^p(s) \leqslant C_2^p(s)$ *holds for all* $s = 0, \ldots, J-1$, *then there exists a* $J \times J$ *nonnegative lower triangular matrix* $\mathbf{B}_J = (b_{kl})_{0 \leqslant k,l \leqslant J-1}$ *such that*

$$\sum_{l'=0}^{J-1} b_{kl'} = c_{1k}, \quad \sum_{k'=0}^{J-1} b_{k'l} \leqslant c_{2l} \quad (k, l = 0, \ldots, J-1).$$

Moreover, if $C_1^p(J-1) = C_2^p(J-1)$, *then* $\sum_{k'=0}^{J-1} b_{k'l} = c_{2l}$.

Proofs are provided in the Supplementary S2. Proposition 3 (Strassen, 1965) establishes conditions for constructing a lower triangular matrix and describes the quantitative relationships that hold among its elements. To connect the treatment effect to the structure of the probability matrix, we define the lexicographically first 2-tuple that achieves the minimum value of $\delta_{k'l'}^{[m]}$:

$$(k_0, l_0) \in \operatorname*{argmin}_{(k',l') \in \Omega} \delta_{k'l'}^{[m]}, \text{ with the pair chosen by the lexicographic order on } \Omega,$$

where $\Omega = \{(1,1), \ldots, (1, J-1); (2,1), \ldots, (2, J-2); \ldots; (J-1, 1)\}$ is the ordered set of 2-tuples (k', l') for $k' = 1, \ldots, J-1$ and $l' = 1, \ldots, J-k'$. By identifying the tuple (k_0, l_0) that minimizes $\delta_{k'l'}^{[m]}$, we summarize the cumulative probabilities associated with these tuples

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

by defining:

$$\lambda_0^{[m]} = \sum_{k=k_0}^{k_0+l_0-1} p_{k+}^{[m]} - \sum_{l=k_0-1}^{k_0+l_0-2} p_{+l}^{[m]}.$$

We now present the following proposition, which outlines the conditions and inequalities across three cases, based on the values of k_0 and l_0.

Proposition 4. *When $k_0 > 1$, let $\Omega_1 = \{1, ..., k_0 - 1\}$. For $k \in \Omega_1 \backslash \{k_0 - 1\}$, we define $q_{k+}^{[m]}$ as follows: $q_{k+}^{[m]} = p_{k+}^{[m]}$ and $q_{(k_0-1)+}^{[m]} = p_{(k_0-1)+}^{[m]} + \min(0, \lambda_0^{[m]})$. Then, the following inequality holds:*

$$\sum_{l=n-1}^{k_0-2} p_{+l}^{[m]} \leqslant \sum_{k=n}^{k_0-1} q_{k+}^{[m]}, \quad n = 1, ..., k_0 - 1.$$

When $l_0 > 1$, let $\Omega_2 = \{k_0 + 1, \ldots, k_0 + l_0 - 1\}$. For $k \in \Omega_2 \backslash \{k_0 + l_0 - 1\}$, we define $q_{k+}^{[m]} = p_{k+}^{[m]}$ and $q_{(k_0+l_0-1)+}^{[m]} = p_{(k_0+l_0-1)+}^{[m]} - \max(0, \lambda_0^{[m]})$. Then, we have the following inequality:

$$\sum_{l=k_0+n-1}^{k_0+l_0-2} p_{+l}^{[m]} \leqslant \sum_{k=k_0+n}^{k_0+l_0-1} q_{k+}^{[m]}, \quad n = 1, ..., l_0 - 1.$$

When $k_0 + l_0 < J$, let $\Omega_3 = \{k_0 + l_0 - 1, \ldots, J - 2\}$. For $l \in \Omega_3 \backslash \{k_0 + l_0 - 1\}$, we define: $q_{+l}^{[m]} = p_{+l}^{[m]}$ and $q_{+(k_0+l_0-1)}^{[m]} = p_{+(k_0+l_0-1)}^{[m]} - \max(0, \lambda_0)$. We then obtain the following inequality,

$$\sum_{k=k_0+l_0}^{k_0+n-1} p_{k+}^{[m]} \leqslant \sum_{l=k_0+l_0-1}^{k_0+n-2} q_{+l}^{[m]}, \quad n = l_0 + 1, ..., J - k_0.$$

The proof of Proposition 4 is provided in the Supplementary S3. This proposition covers three cases that encompass all possible values of k_0 and l_0, detailing the associated marginal distributions of the potential outcomes under treatment and control. We first define (k_0, l_0) and the minimum value $\lambda_0^{[m]}$ derived from them, then establish the comparative sizes of $p_{kl}^{[m]}$ and $q_{kl}^{[m]}$ by examining the three distinct cases for k_0 and l_0. The procedure is presented in Algorithm 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Algorithm 1 The construction process of Proposition 4.

Input: Marginal probabilities $p_{k+}^{[m]}$ and $p_{+l}^{[m]}$ in stratum m, number of categories J, the ordered set Ω.

Output: Pair (k_0, l_0), adjustment quantity $\lambda_0^{[m]}$, updated marginal probabilities $q_{k+}^{[m]}$ and $q_{+l}^{[m]}$.

1: **Initialize:** Given $p_{k+}^{[m]}$ and $p_{+l}^{[m]}$, set $q_{k+}^{[m]} \leftarrow p_{k+}^{[m]}$ and $q_{+l}^{[m]} \leftarrow p_{+l}^{[m]}$ for all k, l.

2: **Compute $\delta_{k'l'}^{[m]}$ for each $(k', l') \in \Omega$:**

$$\delta_{k'l'}^{[m]} = \sum_{k=k'}^{J-1} p_{k+}^{[m]} + \sum_{k=k'+l'}^{J-1} p_{k+}^{[m]} + \sum_{l=0}^{k'-2} p_{+l}^{[m]} - \sum_{l=k'+l'-1}^{J-1} p_{+l}^{[m]},$$

with empty sums interpreted as zero.

3: **Select pair (k_0, l_0) and compute update adjustment:**

$$(k_0, l_0) = \arg\min_{(k',l')\in\Omega} \delta_{k'l'}^{[m]}, \quad \lambda_0^{[m]} = \sum_{k=k_0}^{k_0+l_0-1} p_{k+}^{[m]} - \sum_{l=k_0-1}^{k_0+l_0-2} p_{+l}^{[m]}.$$

4: **Modify marginal probabilities:**

$$\text{if } k_0 > 1, \text{ update } q_{(k_0-1)+}^{[m]} = p_{(k_0-1)+}^{[m]} + \min(0, \lambda_0^{[m]}).$$
$$\text{if } l_0 > 1, \text{ update } q_{(k_0+l_0-1)+}^{[m]} = p_{(k_0+l_0-1)+}^{[m]} - \max(0, \lambda_0^{[m]}).$$
$$\text{if } k_0 + l_0 < J, \text{ update } q_{+(k_0+l_0-1)}^{[m]} = p_{+(k_0+l_0-1)}^{[m]} - \max(0, \lambda_0).$$

5: **Return:** (k_0, l_0), $\lambda_0^{[m]}$, $q_{k+}^{[m]}$, $q_{+l}^{[m]}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Using the capacity inequalities in Propositions 3 and 4, we can present a optimal matrix $L^*_{[m]} \in \mathcal{F}_{[m]}$ for stratum m that satisfies $\gamma\left(L^*_{[m]}\right) = \gamma^U_{[m]}$. Although the maximizing construction of $L^*_{[m]}$ is not unique, this non-uniqueness does not affect sharpness or identification. In particular, by Theorems 1 and 2, the sharp bounds γ^U and γ^L depend only on the marginals $\{p^{[m]}_{k+}, p^{[m]}_{+l}\}$, not on the particular realization of $L^*_{[m]}$.

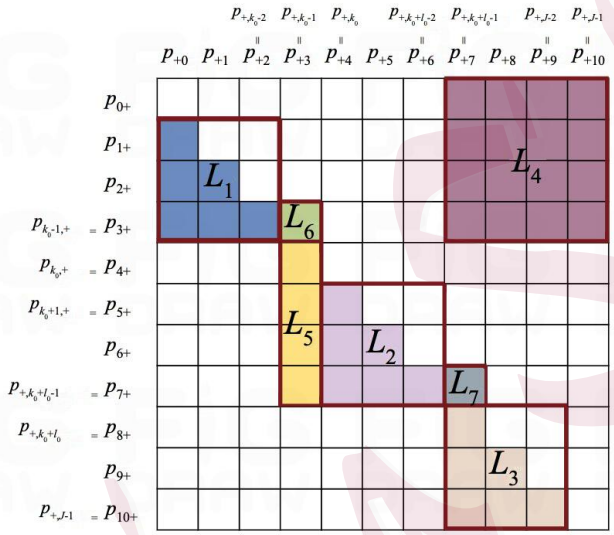


Figure 1: Matrix visualization for $J = 11$, $k_0 = 4$, $l_0 = 4$, and $M = 1$. The matrix is partitioned into seven blocks determined by (k_0, l_0); unshaded entries are zero. Row markers on the left indicate $p^{[m]}_{k+}$ and column markers on the top indicate $p^{[m]}_{+l}$ for $L^*_{[m]}$.

To illustrate the structure, write $L^*_{[m]} = \{L^{(b)}_{[m]}\}_{b=1}^{7}$ for the seven blocks induced by the cutpoints $k_0 - 1$ and $k_0 + l_0 - 1$. To visualize the matrix construction, we set $J = 11$, $k_0 = 4$, $l_0 = 4$ as an example, as shown in Figure 1.

We initialize $L^*_{[m]}$ with $p^{[m]}_{kl} = 0$ for all $k, l \in \{0, \ldots, J-1\}$, and throughout interpret empty sums as 0. From the first case of Proposition 4, for $n = 1, \ldots, k_0 - 1$,

$$\left(q^{[m]}_{1+}, q^{[m]}_{2+}, ..., q^{[m]}_{(k_0-1)+}\right) \text{ and } \left(p^{[m]}_{+0}, p^{[m]}_{+1}, ..., p^{[m]}_{+(k_0-2)}\right),$$

and then due to

$$\sum_{l=n-1}^{k_0-2} p^{[m]}_{+l} \leqslant \sum_{k=n}^{k_0-1} q^{[m]}_{k+},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

which satisfies the conditions of the first case in Proposition 3. We then update the sub-matrix $L^1_{[m]} = \{p^{[m]}_{kl}\}_{1\leqslant k\leqslant k_0-1, 0\leqslant l\leqslant k_0-2}$ with non-negative entries, ensuring that $L^1_{[m]}$ remains lower-triangular and satisfies

$$\sum_{l=0}^{k_0-2} p^{[m]}_{kl} \leqslant q^{[m]}_{k+} \text{ and } \sum_{k=1}^{k_0-1} p^{[m]}_{kl} = p^{[m]}_{+l}.$$

Following the same steps for sub-matrix $L^1_{[m]}$, from in the second and third cases of Proposition 4, we obtain the vectors

$$\left(q^{[m]}_{(k_0+1)+}, ..., q^{[m]}_{(k_0+l_0-1)+}\right) \text{ and } \left(p^{[m]}_{+k_0}, .., p^{[m]}_{+(k_0+l_0-2)}\right),$$

and

$$\left(q^{[m]}_{+(k_0+l_0-1)}, ..., q^{[m]}_{+(J-2)}\right) \text{ and } \left(p^{[m]}_{(k_0+l_0)+}, .., p^{[m]}_{(J-1)+}\right).$$

Using the first and second cases in Proposition 3, we apply them to $L^2_{[m]}$ and $L^3_{[m]}$, respectively. Subsequently, we update $L^2_{[m]} = \{p^{[m]}_{kl}\}_{k_0+1\leqslant k\leqslant k_0+l_0-1, k_0\leqslant l\leqslant k_0+l_0-2}$ and $L^3_{[m]} = \{p^{[m]}_{kl}\}_{k_0+l_0\leqslant k\leqslant J-1, k_0+l_0-1\leqslant l\leqslant J-2}$ with non-negative terms, ensuring that they remain lower-triangular and satisfying:

$$\sum_{l=k_0}^{k_0+l_0-2} p^{[m]}_{kl} \leqslant q^{[m]}_{k+} \text{ and } \sum_{k=k_0+1}^{k_0+l_0-1} p^{[m]}_{kl} = p^{[m]}_{+l},$$

and

$$\sum_{l=k_0+l_0-1}^{J-2} p^{[m]}_{kl} = p^{[m]}_{k+}, \sum_{k=k_0+l_0}^{J-1} p^{[m]}_{kl} \leqslant q^{[m]}_{+l}.$$

In sub-matrix $L^4_{[m]}$, for $k = 0, ..., k_0 - 1, l = k_0 + l_0 - 1, ..., J - 1$, we set

$$p^{[m]}_{kl} = \left(p^{[m]}_{k+} - \sum_{l=0}^{k_0+l_0-2} p^{[m]}_{kl}\right)\left(p^{[m]}_{+l} - \sum_{k=k_0}^{J-1} p^{[m]}_{kl}\right).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Similarly, for the remaining sub-matrices, we have, for $L^5_{[m]}$:

$$p^{[m]}_{k(k_0-1)} = p^{[m]}_{k+} - \sum_{l=k_0}^{J-1} p^{[m]}_{kl},$$

where $k = k_0, ..., k_0 + l_0 - 1$. For $L^6_{[m]}$ and $L^7_{[m]}$:

$$p^{[m]}_{(k_0-1)(k_0-1)} = p^{[m]}_{+(k_0-1)} - \sum_{k=k_0}^{k_0+l_0-1} p^{[m]}_{k(k_0-1)},$$

and

$$p^{[m]}_{(k_0+l_0-1)(k_0+l_0-1)} = \max\left(0, \lambda^{[m]}_0\right),$$

respectively. To summarize, we construct the lower triangular matrices $L^1_{[m]}, L^2_{[m]}$, and $L^3_{[m]}$ as outlined in Proposition 4. Matrices $L^4_{[m]}, L^5_{[m]}, L^6_{[m]}$ and $L^7_{[m]}$ are constructed based on the values of k_0 and l_0. In the following sections, we incorporate covariate adjustment to refine the estimates, accounting for additional sources of variation within stratified randomized experiments.

5. Covariate adjustment

Covariate adjustment improves the precision of treatment effect estimators in stratified randomized experiments (Lin, 2013; Liu and Yang, 2020; Freedman, 2008; Lei and Ding, 2021). We view stratification as a design choice that partitions units into homogeneous blocks and fixes within-block allocation, while covariate adjustment is an analysis strategy that conditions on baseline covariates to reduce residual imbalance and variance.

To estimate the treatment effect γ^U in Theorem 1 and γ^{ind} in Corollary 1, we begin by estimating the marginal probabilities $p^{[m]}_{k+}$ and $p^{[m]}_{+l}$ using the observed data $\{Z_i, B_i, X_i, Y_i^{\text{obs}}\}_{i=1}^N$. For unadjusted plug-in estimates of γ^U and γ^{ind}, we replace $p^{[m]}_{k+}$ and $p^{[m]}_{+l}$ with their sample

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

analogues:

$$\hat{p}_{k+}^{[m]} = \left(\sum_{i:B_i=m} Z_i \right)^{-1} \sum_{i:B_i=m} Z_i I_{\{Y_i^{\text{obs}}=k\}}, \quad \hat{p}_{+l}^{[m]} = \left\{ \sum_{i:B_i=m} (1-Z_i) \right\}^{-1} \sum_{i:B_i=m} (1-Z_i) I_{\{Y_i^{\text{obs}}=l\}},$$

which require $\sum_{i:B_i=m} Z_i > 0$ and $\sum_{i:B_i=m}(1-Z_i) > 0$; otherwise one may pool strata or use stabilized estimators. To estimate the covariate-adjusted bounds, we initially estimate the marginal probabilities of the potential outcomes for unit i, conditional on the given covariates X_i, $p_{k+}^{[m]}(X_i)$ and $p_{+l}^{[m]}(X_i)$:

$$\hat{p}_{k+}^{[m]}(X_i) = \Pr\left\{ Y_i^{\text{obs}} = k \mid Z_i = 1, B_i = m, X_i \right\},$$
$$\hat{p}_{+l}^{[m]}(X_i) = \Pr\left\{ Y_i^{\text{obs}} = l \mid Z_i = 0, B_i = m, X_i \right\}.$$

Following Imbens and Rubin (2015), sample analogues remain applicable for low-dimensional and discrete covariates. For high-dimensional or continuous covariates, the proportional odds model is a commonly employed approach for ordinal outcomes; other flexible prediction methods may also be used to obtain $\hat{p}_{k+}^{[m]}(X_i)$ and $\hat{p}_{+l}^{[m]}(X_i)$. Incorporating covariates allows for further refinement of the bounds on the relative treatment effect γ (Lee, 2009; Lu et al., 2018). We define the bounds of γ conditional on the covariates X_i as follows. First, define the stratified conditional contrast

$$\gamma_{[m]}(X_i) = \Pr\left\{ Y_i(1) > Y_i(0) \mid B_i = m, X_i \right\} - \Pr\left\{ Y_i(1) < Y_i(0) \mid B_i = m, X_i \right\}.$$

Then, using the conditional marginals, we obtain the adjusted stratified lower bound $\gamma_{[m]}^{L}(X_i) = \max_{1 \leqslant k' \leqslant J-1} \max_{1 \leqslant l' \leqslant J-k'} \xi_{k'l'}^{[m]}(X_i)$, the upper bound $\gamma_{[m]}^{U}(X_i) = \min_{1 \leqslant k' \leqslant J-1} \min_{1 \leqslant l' \leqslant J-k'} \delta_{k'l'}^{[m]}(X_i)$, and under independent potential outcomes

$$\gamma_{[m]}^{\text{ind}}(X_i) = \sum_{k=0}^{J-1} \sum_{l=0}^{J-1} \text{sgn}(k-l) p_{k+}^{[m]}(X_i) p_{+l}^{[m]}(X_i),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

where $\delta_{k'l'}^{[m]}(X_i)$ and $\xi_{k'l'}^{[m]}(X_i)$ are defined by $p_{k+}^{[m]}(X_i)$ and $p_{+l}^{[m]}(X_i)$. By averaging these conditional quantities over the covariate distribution within each stratum and then weighting by the stratum proportions, we obtain the adjusted bounds at the population level:

$$\tilde{\gamma}_{[m]}^{U} = E\Big\{\gamma_{[m]}^{U}(X_i) \mid B_i = m\Big\},\ \ \tilde{\gamma}_{[m]}^{L} = E\Big\{\gamma_{[m]}^{L}(X_i) \mid B_i = m\Big\},\ \ \tilde{\gamma}_{[m]}^{\text{ind}} = E\Big\{\gamma_{[m]}^{\text{ind}}(X_i) \mid B_i = m\Big\},$$

$$\tilde{\gamma}^{U} = \sum_{m=1}^{M} \pi_{[m]}\tilde{\gamma}_{[m]}^{U}, \qquad \tilde{\gamma}^{L} = \sum_{m=1}^{M} \pi_{[m]}\tilde{\gamma}_{[m]}^{L}, \qquad \tilde{\gamma}^{\text{ind}} = \sum_{m=1}^{M} \pi_{[m]}\tilde{\gamma}_{[m]}^{\text{ind}}.$$

Then, we obtain the adjusted bounds by the following Theorem.

Theorem 3. *In a randomized block experiment with ordinal outcomes, the covariate-adjusted sharp bounds of the relative treatment effect γ are*

$$\tilde{\gamma}^{L} = \sum_{m=1}^{M} \pi_{[m]} E\left\{\gamma_{[m]}^{L}(X_i) | B_i = m\right\}, \tilde{\gamma}^{U} = \sum_{m=1}^{M} \pi_{[m]} E\left\{\gamma_{[m]}^{U}(X_i) | B_i = m\right\}.$$

Furthermore, the adjusted bounds are tighter than the unadjusted bounds, i.e.,

$$\gamma^{L} \leqslant \tilde{\gamma}^{L} \leqslant \tilde{\gamma}^{U} \leqslant \gamma^{U}.$$

Under independent potential outcomes, the covariate-adjusted independence benchmark is

$$\tilde{\gamma}^{\text{ind}} = \sum_{m=1}^{M} \pi_{[m]} E\left\{\gamma_{[m]}^{\text{ind}}(X_i) | B_i = m\right\},$$

and the difference between the adjusted and unadjusted values satisfies

$$\tilde{\gamma}_{[m]}^{\text{ind}} - \gamma_{[m]}^{\text{ind}} = \sum_{k=0}^{J-1}\sum_{l=0}^{J-1} \text{sgn}(k-l)\text{Cov}\big\{p_{k+}^{[m]}(X_i), p_{+l}^{[m]}(X_i)\big| B_i = m\big\},$$

so that $\tilde{\gamma}^{\text{ind}} - \gamma^{\text{ind}} = \sum_{m=1}^{M} \pi_{[m]}\big(\tilde{\gamma}_{[m]}^{\text{ind}} - \gamma_{[m]}^{\text{ind}}\big)$.

Proof of Theorem 3 is provided in the Supplementary S5. To demonstrate that covariate adjustment can shorten the bounds, consider any population F with two equally sized

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

subpopulations and $J = 3$, defined by a binary covariate $X \in \{X_1, X_2\}$. Assume that the potential outcomes conditional on $X = X_1$ and $X = X_2$ follow distributions $F(X_1)$ and $F(X_2)$, respectively. Let the marginal probability vectors be

$$\mathbf{p}_{\text{row}}(x) = \big(p_{0+}(x), p_{1+}(x), p_{2+}(x)\big)^{\mathrm{T}}, \qquad \mathbf{p}_{\text{col}}(x) = \big(p_{+0}(x), p_{+1}(x), p_{+2}(x)\big)^{\mathrm{T}},$$

and compute the corresponding bounds for γ. Without covariate information, using Theorems 1 and 2 and Corollary 1, the unadjusted values are $\gamma^L = -5/250$, $\gamma^U = 125/250$ and $\gamma^{\text{ind}} = 62/250$. After adjusting for X, the bounds become sharper: $\tilde{\gamma}^L = -5/250$, $\tilde{\gamma}^U = 95/250$, and $\tilde{\gamma}^{\text{ind}} = 56/250$. These values are consistent with the theoretical ordering $\gamma^L \leqslant \tilde{\gamma}^L \leqslant \tilde{\gamma}^U \leqslant \gamma^U$. The joint distribution and marginal distributions are as follows:

$$F = \begin{pmatrix} 4/50 & 3/50 & 3/50 \\ 9/50 & 4/50 & 7/50 \\ 11/50 & 4/50 & 5/50 \end{pmatrix}, \quad \mathbf{p}_{\text{row}} = \begin{pmatrix} 1/5 \\ 2/5 \\ 2/5 \end{pmatrix}, \quad \mathbf{p}_{\text{col}} = \begin{pmatrix} 24/50 \\ 11/50 \\ 15/50 \end{pmatrix}.$$

$$F(X_1) = \begin{pmatrix} 1/25 & 2/25 & 2/25 \\ 6/25 & 3/25 & 6/25 \\ 2/25 & 1/25 & 2/25 \end{pmatrix}, \quad \mathbf{p}_{\text{row}}(X_1) = \begin{pmatrix} 1/5 \\ 3/5 \\ 1/5 \end{pmatrix}, \quad \mathbf{p}_{\text{col}}(X_1) = \begin{pmatrix} 9/25 \\ 6/25 \\ 2/5 \end{pmatrix}.$$

$$F(X_2) = \begin{pmatrix} 3/25 & 1/25 & 1/25 \\ 3/25 & 1/25 & 1/25 \\ 9/25 & 3/25 & 3/25 \end{pmatrix}, \quad \mathbf{p}_{\text{row}}(X_2) = \begin{pmatrix} 1/5 \\ 1/5 \\ 3/5 \end{pmatrix}, \quad \mathbf{p}_{\text{col}}(X_2) = \begin{pmatrix} 3/5 \\ 1/5 \\ 1/5 \end{pmatrix}.$$

Where $\mathbf{p}_{\text{row}}$ and $\mathbf{p}_{\text{col}}$ denote the population marginal probability vectors obtained by summing the joint distribution over rows and columns, respectively. This example imposes no special restrictions on the joint distribution of potential outcomes. The choice under

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

X_2 is only for illustration, and the same qualitative point can be demonstrated using other feasible joint distributions or covariate structures. In what follows, we report $[\gamma^{\text{ind}}, \gamma^U]$ as an auxiliary independence-based reference interval in both the simulation study and the empirical application, following the interpretive motivation described in Remark 2, and focus on γ^{ind} as the reference lower endpoint.

After obtaining plug-in estimates of the marginal probabilities, we estimate the interval $[\gamma^{\text{ind}}, \gamma^U]$. We construct a joint $(1-\alpha)$ confidence interval for $[\gamma^{\text{ind}}, \gamma^U]$ using a stratified bootstrap that resamples within each stratum m separately and preserves the within-stratum treatment allocation. Let $(\hat{\gamma}^{\text{ind}}, \hat{\gamma}^U)$ be the plug-in estimators from the original sample, and $(\hat{\gamma}^{\text{ind}}_{(b)}, \hat{\gamma}^U_{(b)})$ the corresponding estimators from the b-th bootstrap sample, $b = 1, \ldots, B$. Define the two-sided maximal deviation

$$a_{(b)} = \max\left\{\hat{\gamma}^{\text{ind}} - \hat{\gamma}^{\text{ind}}_{(b)},\ \hat{\gamma}^U_{(b)} - \hat{\gamma}^U\right\},$$

let z_α be the $(1-\alpha)$ empirical quantile of $\{a_{(b)}\}_{b=1}^B$, and form the simultaneous confidence interval

$$\left[\ \hat{\gamma}^{\text{ind}} - z_\alpha,\ \hat{\gamma}^U + z_\alpha\ \right].$$

This percentile-of-max construction accommodates the non-smooth $\max(\cdot)$ operator and yields an interval that is at least conservative under regularity conditions; see Horowitz and Manski (2000). In the next section, we compare the four estimators in numerical simulations.

6. Simulation study

In this section, we evaluate how stratification and covariate adjustment affect the estimation of the relative treatment effect γ. Following Agresti (2010); Cheng (2009); Agresti and Kateri (2017) and the discussion in Section 3, we report $[\gamma^{\text{ind}}, \gamma^U]$ as an auxiliary reference interval, where γ^{ind} serves as the benchmark lower endpoint and γ^U remains the unrestricted sharp upper bound. We construct the nonparametric confidence interval introduced in Section 5. We consider four analysis strategies: (i) an unadjusted estimator without stratification, (ii)

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

a covariate-adjusted estimator without stratification, (iii) a stratified estimator (introduced in Section 2), and (iv) a stratified, covariate-adjusted estimator (introduced in Section 5). For each strategy, we compute the plug-in estimator $\hat{\gamma}^{\text{ind}}$ and compare its performance.

Our data-generating process is similar to those of Liu and Yang (2020) and Li et al. (2011). We simulate potential outcomes using the cumulative logistic regression model:

$$\begin{aligned}
\text{logit}\ \Pr\{Y_i(1) \leqslant j | X_i = x, B_i = m\} &= \alpha_{1,m,j} - \beta_{1,m}^{\mathrm{T}} x, \\
\text{logit}\ \Pr\{Y_i(0) \leqslant j | X_i = x, B_i = m\} &= \alpha_{0,m,j} - \beta_{0,m}^{\mathrm{T}} x,
\end{aligned}$$

where $m \in \{1, \ldots, M\}$ and $j = 0, \ldots, J-2$.

The covariate vector X_i is drawn from a multivariate normal distribution with mean zero and covariance Σ, where $\Sigma_{hh'} = \rho^{|h-h'|}$. We generate $Y_i(1)$ and $Y_i(0)$ conditional on (X_i, B_i). The number of strata M, the total number of units in each stratum, and the regression coefficients $\beta_{z,m}$ and thresholds $\alpha_{z,m,j}$ (with $\alpha_{z,m,0} \leq \cdots \leq \alpha_{z,m,J-2}$) are varied according to the following two scenarios.

Scenario 1. To visualize Theorem 1, we numerically verify that the sharp upper bound γ^U is attained by the optimal matrix L^*. We consider four joint matrices with the same fixed marginals: (i) a randomly generated feasible matrix L_1, (ii) a matrix inducing negative association between $Y_i(1)$ and $Y_i(0)$, denoted L_2, (iii) a matrix inducing strong positive association, denoted L_3, and (iv) the constructed optimizer L^*. We compute $\gamma(L_1)$, $\gamma(L_2)$, and $\gamma(L_3)$ under their respective joint distributions and compare them to $\gamma(L^*) = \gamma^U$. We then report a box plot of $\hat{\gamma}^U - \hat{\gamma}^{\text{ind}}$ and distribution plots (e.g., histograms or kernel densities) for all four matrices with $J = 30$.

Scenario 2. There are many small strata. We set $M = 10$ and $J = 3$. The correlation parameter $\rho \in \{0.6, 0.8\}$ with the 10-dimensional X_i covariate vector. We consider three stratum sizes $n_m \in \{10, 15, 20\}$. The regression coefficients $\beta_{0,m}$ and $\beta_{1,m}$ are generated once

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

per stratum and then held fixed:

$$\beta_{1,m} \sim t(3), \qquad \beta_{0,m} \sim t(3) + \beta_{1,m},$$

Set the baseline thresholds $\alpha_{m,j}(1)$ and $\alpha_{m,j}(0)$ by applying a logit transformation to the normalized cumulative probability and define

$$\alpha_{1,m,j} = \alpha_{m,j}(1) + u_m + \varepsilon_m(1), \quad \alpha_{0,m,j} = \alpha_{m,j}(0) + u_m + \varepsilon_m(0),$$

where $u_m \sim N(0, \sigma_u^2)$ and $\varepsilon_m(z) \sim N(0, \sigma_\varepsilon^2)$. This setting allows us to evaluate the impact of sample size on covariate adjustment in a multi-stratum context with a moderate number of ordinal categories.

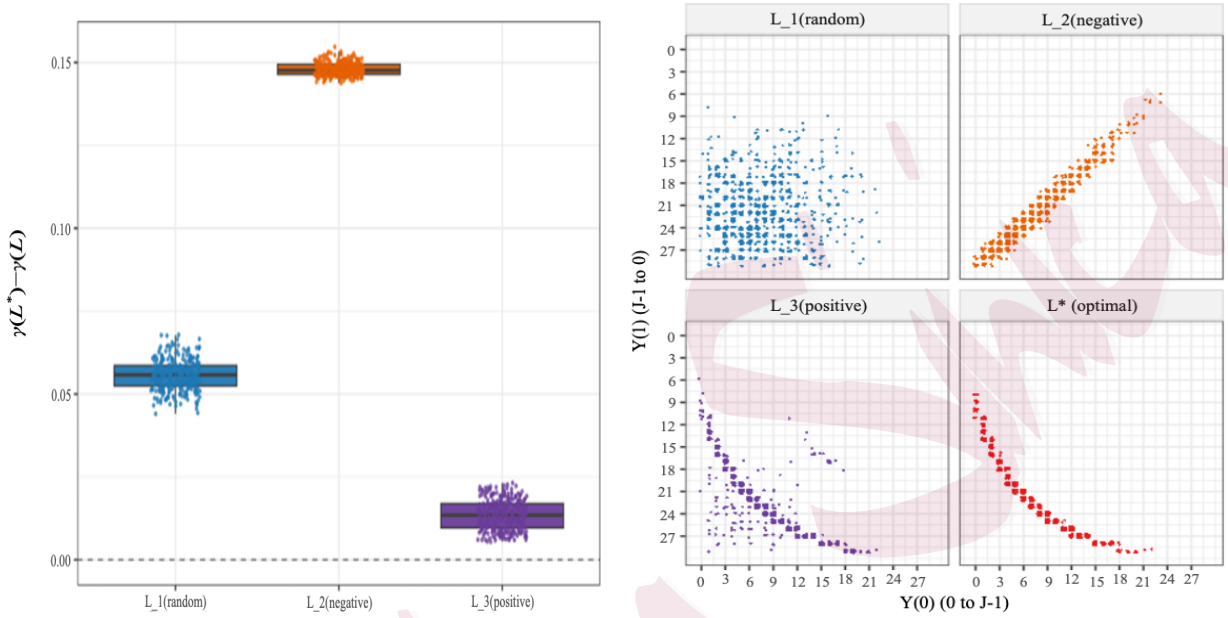
Scenario 3. There are two large strata. We set $M = 2$ and $J = 5$, with $\rho \in \{0.3, 0.45\}$. As in Scenario 2, we consider $n_m \in \{250, 320, 400\}$ under the 6-dimensional X_i covariate vector and generate $\beta_{0,m}, \beta_{1,m}$, $\alpha_{1,m,j}$ and $\alpha_{0,m,j}$ in the same way as in Scenario 2, independently across the two strata. The two strata are independent; within each stratum, treatment and control follow the same data-generating model.

Both covariates and potential outcomes are generated 1000 times to simulate stratified randomized experiments. For each simulation, $n_{m[1]} = n_m/2$ units are assigned to the treatment group, with the remaining units assigned to the control group. Due to the data setting, we compare four methods: the unadjusted estimator without stratification $\hat{\gamma}_{\text{unadj}}$, the covariate-adjusted estimator without stratification $\hat{\gamma}_{\text{adj}}$, the stratified estimator $\hat{\gamma}_{\text{stra}}$, and the stratified, covariate-adjusted estimator $\hat{\gamma}_{\text{stra_adj}}$.

The simulation results for the scenarios are presented in Figures 2 - 4, and Tables 1 - 2. Figure 2 illustrates the shape differences between the optimal matrix L^* and the other three matrices, along with a comparison of their respective $\gamma(L)$ values. Figure 3 shows the violin plots of the bound length $[\hat{\gamma}^{\text{ind}}, \hat{\gamma}^U]$ for the four strategies, while Figure 4 illustrates the root mean square error (RMSE) for each method. Tables 1 and 2 further evaluate the performance of the four methods by comparing their absolute bias, standard deviation,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

RMSE, standard errors, and the average length of bound intervals. These metrics offer a detailed comparison of the methods' precision and reliability. The main findings from these simulations are summarized as follows.



(a) Box plot of $\gamma(L^*)-\gamma(L)$, matrix construction from left to right: L_1 (random), L_2 (negative) , L_3 (positive)

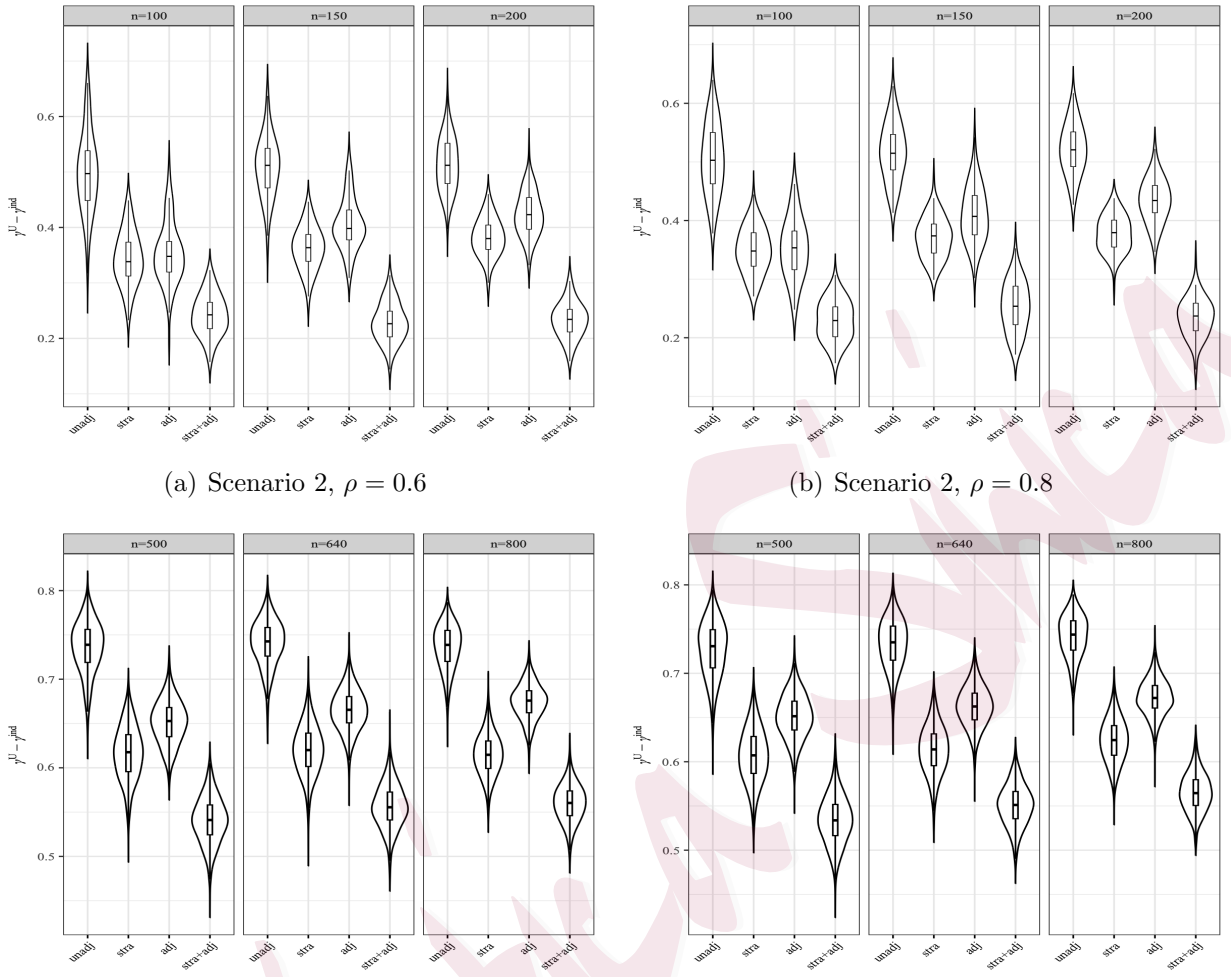
(b) Matrix distribution plots

Figure 2: Comparison results of different matrices in Scenario 1.

The left panel in Figure 2 uses the sharp upper bound $\gamma(L^*)$ under the optimal matrix L^* as the baseline value, calculating $\gamma(L)$ for three different joint-distribution constructions. The figure shows that $\gamma(L^*)$ consistently yields the largest values across all settings. The largest discrepancy occurs under the negatively correlated construction L_2 (negative), while the gap is smaller under the positively correlated construction L_3 (positive). This demonstrates that the sharp upper bound $\gamma(L^*)$ corresponds to positive correlation in potential outcomes (as illustrated in the numerical example in Section 3). The right panel also clearly shows that the smallest difference occurs between the positively correlated structure L_3 (positive) and the optimal matrix L^*, due to their similarity.

The stratified, covariate-adjusted estimator $\hat{\gamma}_{\text{stra_adj}}$ consistently demonstrates the lowest RMSE across different scenarios, indicating its superior accuracy in estimating treatment

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495



(a) Scenario 2, $\rho = 0.6$

(b) Scenario 2, $\rho = 0.8$

(c) Scenario 3, $\rho = 0.3$

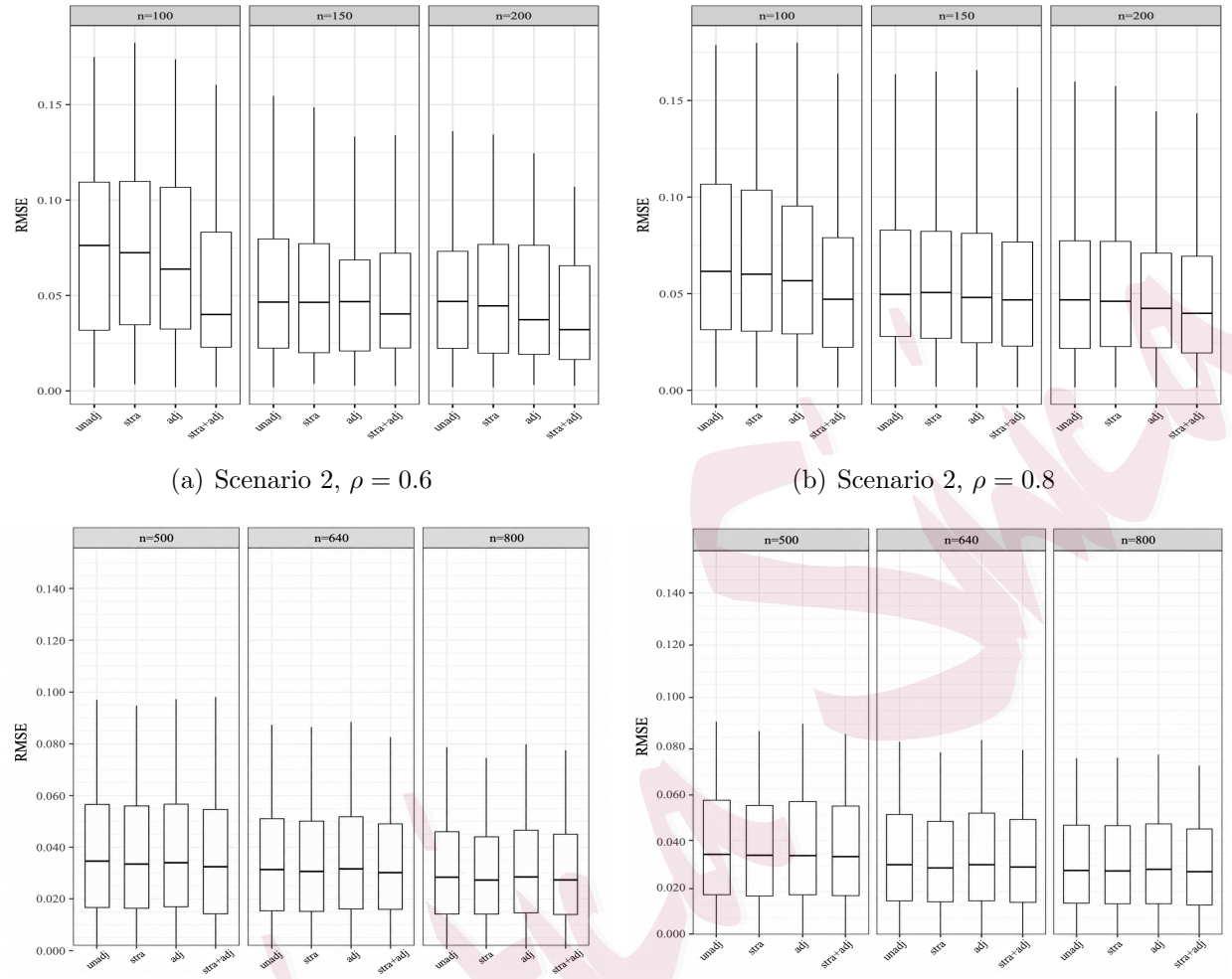
(d) Scenario 3, $\rho = 0.45$

Figure 3: Violin plots of $\hat{\gamma}^U - \hat{\gamma}^{\text{ind}}$ for four effect estimators. In each subplot, the violin plots correspond to the methods, from left to right, $\hat{\gamma}_{\text{unadj}}$, $\hat{\gamma}_{\text{stra}}$, $\hat{\gamma}_{\text{adj}}$, and $\hat{\gamma}_{\text{stra_adj}}$.

effects compared to other methods. Specifically, in Scenario 2, $\hat{\gamma}_{\text{stra_adj}}$ reduces the RMSE of $\hat{\gamma}_{\text{unadj}}$ by 18.6% on average; and shortens the length of the bound interval (i.e., $\hat{\gamma}^U - \hat{\gamma}^{\text{ind}}$) of $\hat{\gamma}_{\text{unadj}}$ by 54.1% on average. In addition, $\hat{\gamma}_{\text{adj}}$ and $\hat{\gamma}_{\text{stra}}$, respectively shorten the length of the bound interval of $\hat{\gamma}_{\text{unadj}}$ by 22.4% and 29.2% on average, demonstrating that stratification and covariate adjustment facilitate narrower bounds and enhance estimation efficiency.

In Scenario 3, we set the number of strata and categories differently from Scenario 2. $\hat{\gamma}_{\text{stra_adj}}$ continues to perform best across different values of ρ. Despite the limited number of strata, $\hat{\gamma}_{\text{stra}}$ still reduces the average bound length of $\hat{\gamma}_{\text{unadj}}$ by 16.2%, showing the importance

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495



(a) Scenario 2, $\rho = 0.6$

(b) Scenario 2, $\rho = 0.8$

(c) Scenario 3, $\rho = 0.3$

(d) Scenario 3, $\rho = 0.45$

Figure 4: Box plots of the RMSE of $\hat{\gamma}$. In each subplot, the box plots correspond to the methods, from left to right, $\hat{\gamma}_{\text{unadj}}$, $\hat{\gamma}_{\text{stra}}$, $\hat{\gamma}_{\text{adj}}$, and $\hat{\gamma}_{\text{stra_adj}}$.

of stratification in reducing estimation imbalances. $\hat{\gamma}_{\text{adj}}$ shortens the average bound length of $\hat{\gamma}_{\text{unadj}}$ by 9.8%. Across the two simulation scenarios, covariate adjustment improves the finite-sample efficiency of the relative treatment effect estimator in addition to shortening the bounds. Compared with the unadjusted estimator $\hat{\gamma}_{\text{unadj}}$, the covariate-adjusted estimator $\hat{\gamma}_{\text{adj}}$ reduces the SD by an average of 14.3%, while the estimator $\hat{\gamma}_{\text{stra_adj}}$ reduces the SD by an average of 25.2%. These results show that, in the simulation settings considered here, covariate adjustment not only sharpens the bounds but also leads to a meaningful reduction in sampling variability, with the combined strategy providing the largest efficiency gain.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

In Scenario 2, where $M = 10$ strata with small stratum sizes $n_m \in \{10, 15, 20\}$ and covariate dimension of 10 are considered, stratified randomization and covariate adjustment alone reduce the SD by approximately 1.3% and 9.4%, respectively, with the limited gains reflecting finite-sample constraints imposed by small stratum sizes and relatively high covariate dimension. In Scenario 3, where stratum sizes are larger and coefficients are generated independently, the reductions increase to approximately 12.1% and 19.3%. Across both scenarios, the combined application achieves an average SD reduction of approximately 25.2%, showing that jointly applying both methods consistently delivers the greatest efficiency gain. Overall, covariate adjustment improves the precision of the relative treatment effect, findings that are consistent with our theoretical expectations. Furthermore, appropriate stratification is essential for eliminating imbalances among covariates.

Table 1: Comparison results of relative treatment effect estimators in Scenario 2.

ρ	n_m	Methods	Bias	SD	RMSE	SE	Interval Length	CP
0.6	10	$\hat{\gamma}_{\text{unadj}}$	-0.000	0.094	0.094	0.085	0.501	0.940
		$\hat{\gamma}_{\text{stra}}$	-0.000	0.093	0.094	0.083	0.343	0.958
		$\hat{\gamma}_{\text{adj}}$	0.000	0.085	0.085	0.083	0.355	0.944
		$\hat{\gamma}_{\text{stra_adj}}$	0.003	0.071	0.071	0.069	0.242	0.951
	15	$\hat{\gamma}_{\text{unadj}}$	0.001	0.078	0.078	0.069	0.510	0.964
		$\hat{\gamma}_{\text{stra}}$	0.002	0.077	0.077	0.068	0.362	0.942
		$\hat{\gamma}_{\text{adj}}$	0.003	0.069	0.069	0.064	0.407	0.956
		$\hat{\gamma}_{\text{stra_adj}}$	0.004	0.060	0.060	0.060	0.229	0.949
	20	$\hat{\gamma}_{\text{unadj}}$	-0.001	0.061	0.061	0.060	0.515	0.941
		$\hat{\gamma}_{\text{stra}}$	-0.001	0.060	0.060	0.059	0.382	0.957
		$\hat{\gamma}_{\text{adj}}$	-0.001	0.056	0.056	0.056	0.426	0.945
		$\hat{\gamma}_{\text{stra_adj}}$	-0.000	0.049	0.049	0.058	0.233	0.950
0.8	10	$\hat{\gamma}_{\text{unadj}}$	-0.004	0.096	0.096	0.083	0.507	0.940
		$\hat{\gamma}_{\text{stra}}$	-0.004	0.095	0.095	0.082	0.348	0.944
		$\hat{\gamma}_{\text{adj}}$	-0.002	0.085	0.085	0.083	0.354	0.955
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.071	0.071	0.071	0.229	0.949
	15	$\hat{\gamma}_{\text{unadj}}$	-0.001	0.075	0.075	0.069	0.518	0.962
		$\hat{\gamma}_{\text{stra}}$	-0.001	0.074	0.074	0.068	0.370	0.956
		$\hat{\gamma}_{\text{adj}}$	0.001	0.069	0.069	0.066	0.412	0.945
		$\hat{\gamma}_{\text{stra_adj}}$	0.004	0.068	0.068	0.065	0.253	0.951
	20	$\hat{\gamma}_{\text{unadj}}$	0.000	0.065	0.065	0.062	0.523	0.941
		$\hat{\gamma}_{\text{stra}}$	0.000	0.064	0.064	0.061	0.377	0.946
		$\hat{\gamma}_{\text{adj}}$	0.000	0.060	0.060	0.058	0.436	0.954
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.059	0.059	0.060	0.235	0.950

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Table 2: Comparison results of relative treatment effect estimators in Scenario 3.

ρ	n_m	Methods	Bias	SD	RMSE	SE	Interval Length	CP
0.3	250	$\hat{\gamma}_{\text{unadj}}$	0.000	0.054	0.054	0.047	0.735	0.965
		$\hat{\gamma}_{\text{stra}}$	0.000	0.048	0.048	0.046	0.616	0.959
		$\hat{\gamma}_{\text{adj}}$	0.001	0.044	0.044	0.047	0.651	0.942
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.037	0.037	0.048	0.540	0.951
	320	$\hat{\gamma}_{\text{unadj}}$	0.003	0.048	0.049	0.042	0.740	0.940
		$\hat{\gamma}_{\text{stra}}$	0.003	0.042	0.042	0.040	0.619	0.958
		$\hat{\gamma}_{\text{adj}}$	0.003	0.039	0.039	0.044	0.665	0.943
		$\hat{\gamma}_{\text{stra_adj}}$	0.002	0.033	0.033	0.042	0.556	0.949
	400	$\hat{\gamma}_{\text{unadj}}$	0.000	0.041	0.041	0.037	0.735	0.964
		$\hat{\gamma}_{\text{stra}}$	0.000	0.036	0.036	0.036	0.614	0.941
		$\hat{\gamma}_{\text{adj}}$	0.001	0.033	0.033	0.035	0.674	0.958
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.028	0.028	0.035	0.559	0.950
0.45	250	$\hat{\gamma}_{\text{unadj}}$	0.000	0.056	0.056	0.054	0.726	0.962
		$\hat{\gamma}_{\text{stra}}$	0.000	0.049	0.049	0.051	0.606	0.943
		$\hat{\gamma}_{\text{adj}}$	0.001	0.045	0.045	0.054	0.651	0.956
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.038	0.038	0.050	0.533	0.950
	320	$\hat{\gamma}_{\text{unadj}}$	0.002	0.048	0.048	0.054	0.732	0.940
		$\hat{\gamma}_{\text{stra}}$	0.001	0.042	0.042	0.050	0.613	0.957
		$\hat{\gamma}_{\text{adj}}$	0.003	0.038	0.038	0.053	0.662	0.944
		$\hat{\gamma}_{\text{stra_adj}}$	0.002	0.032	0.032	0.048	0.550	0.951
	400	$\hat{\gamma}_{\text{unadj}}$	-0.000	0.042	0.042	0.036	0.741	0.963
		$\hat{\gamma}_{\text{stra}}$	-0.000	0.037	0.037	0.035	0.624	0.956
		$\hat{\gamma}_{\text{adj}}$	0.001	0.034	0.034	0.036	0.672	0.943
		$\hat{\gamma}_{\text{stra_adj}}$	0.000	0.029	0.029	0.035	0.565	0.949

7. Analysis of cardia cancer patient data

To illustrate our theory and methods, we analyze data from a study conducted at the Karolinska Institute in Stockholm, Sweden, as described by Rubin (2008). The study aims to evaluate whether being diagnosed at hospitals with differing patient volumes affects survival time for cardia cancer patients. Hospital capacity is used as an indicator to distinguish between high volume and low volume hospitals, with capacity defined by the number of patients treated in recent years. The dataset includes 158 cardia cancer patients diagnosed between 1988 and 1995. Of these, 79 were diagnosed in high volume hospitals, defined as those treating more than 10 patients during this period, while the remaining 79 were diagnosed in low volume hospitals.

Some argue that receiving treatment at high volume hospitals yields better outcomes. However, such comparisons may lack rigorous adjustment and can be influenced by several complex factors. The dataset includes information on hospital capacity, patient character-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

istics, and survival outcomes. The outcome variable is the survival time after diagnosis, categorized into three ordered groups: $\leqslant$1 year, 2–4 years, and $\geqslant$5 years. Among patients diagnosed in high volume hospitals, 51 survived $\leqslant$1 year, 18 survived 2–4 years, and 10 survived $\geqslant$5 years. In low volume hospitals, the corresponding numbers are 50, 21, and 8, respectively. Pretreatment covariates, including age, gender, and rural residence (among others), are used for covariate adjustment, and in addition we stratify by rural residence, which is an important confounder in this example.

We first fit separate proportional odds models for the high volume (treatment) and low volume (control) groups to estimate fitted probabilities for each individual under treatment and control. Subsequently, we apply four estimators to compute 95% confidence intervals for the bound endpoints and the corresponding point estimates. To enhance objectivity and reduce potential confounding by age, patients aged 85 and above, as well as those aged 20 and below, are excluded from the analysis. The results of this study are presented in Table 3.

Table 3: Comparison of patient diagnostic results

Methods	$\hat{\gamma}^{\text{ind}}$	$\hat{\gamma}_U$	CI	CIlength
$\hat{\gamma}_{\text{unadj}}$	-0.002	0.354	[-0.140, 0.492]	0.633
$\hat{\gamma}_{\text{stra}}$	0.023	0.352	[-0.112, 0.488]	0.600
$\hat{\gamma}_{\text{adj}}$	0.065	0.342	[-0.075, 0.483]	0.558
$\hat{\gamma}_{\text{stra_adj}}$	0.123	0.339	[-0.037, 0.500]	0.537

As shown in Table 3, it is evident that $\hat{\gamma}_{\text{stra}}$, $\hat{\gamma}_{\text{adj}}$, and $\hat{\gamma}_{\text{stra_adj}}$ all improve the lower bound point estimate of the unadjusted estimator, $\hat{\gamma}_{\text{unadj}}$. The degree of improvement becomes progressively more pronounced, and the average lengths of the confidence intervals for the bounds also tend to decrease. Among the four estimators, $\hat{\gamma}_{\text{stra_adj}}$ performs the best, improving the lower bound point estimate of $\hat{\gamma}_{\text{unadj}}$ and reducing the interval length by 15.1%. Meanwhile, $\hat{\gamma}_{\text{stra}}$ and $\hat{\gamma}_{\text{adj}}$ shorten the interval lengths of $\hat{\gamma}_{\text{unadj}}$ by 5.2% and 11.8% respectively.

Although the point estimate for the lower bound is positive, its confidence interval includes negative values. Therefore, the study findings indicate no causal relationship between

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

receiving treatment at a high-volume hospital and survival time. Considering the actual background, the following factors are taken into account: (1) High-volume hospitals may not always be the preferred choice due to factors such as cost, travel distance, or the lack of support from family and friends, making smaller hospitals more accessible. (2) Local government policies regarding resource allocation might selectively close small hospitals that do not significantly impact patient survival, potentially influencing observed outcomes. This conclusion aligns with the views of Rubin (2008); Agresti and Kateri (2017).

8. Discussion

In this paper, we introduce a framework for analyzing stratified randomized experiments with ordinal potential outcomes. This framework builds on foundational theories of randomized experiments and extends them to accommodate the unique challenges posed by ordinal data. Recognizing that the traditional average treatment effect cannot adequately capture the nuances of ordinal data, we adopt relative treatment effects as a more suitable measure. This approach addresses the limitations of traditional methods in portraying differences across ordinal categories. However, we acknowledge the challenge of unidentifiable relative treatment effects due to over-categorization. To mitigate this issue, we focus on establishing bounds for relative treatment effects, deriving sharp expressions for their upper and lower limits in stratified randomized experiments. These bounds offer a robust basis for inference, even in scenarios where treatment effects remain partially unidentifiable.

To enhance the efficiency and precision of estimation, we propose the integration of covariate adjustment techniques and provide the corresponding theoretical derivations. We further evaluate the proposed methods empirically, comparing the performance of four estimation approaches in terms of their accuracy and reliability. Our findings underscore the utility of covariate adjustment and stratification in improving the precision of bounds on relative treatment effects.

While our framework establishes a foundation for analyzing ordinal outcomes in stratified randomized experiments, several avenues merit future exploration. One potential direction is

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

to address the challenges of incorporating missing covariates into the randomization framework, which has been shown to improve the robustness of inference methods in randomized and observational studies (Miao et al., 2023; Tu et al., 2024; Zhao and Ding, 2024). Another avenue involves investigating the impact of different covariate adjustment strategies on the bounds derived in our framework, and it is also interesting to examine the impact of block-treatment interactions on treatment effects. Such studies could provide practical insights for optimizing estimation precision in various experimental designs. Additionally, while the current focus is on bounds for relative treatment effects, exploring alternative metrics or hybrid approaches could provide a more nuanced and comprehensive perspective on ordinal treatment effects.

Supplementary Material

The online Supplementary Material contains proofs for Proposition 1-4, Theorems 1-3, and Corollary 1.

Acknowledgments

We thank the Associate Editor and the reviewers for their valuable comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 12371281); the Emerging Interdisciplinary Project, the Fundamental Research Funds, and the Disciplinary Funds in Central University of Finance and Economics.

References

Agresti, A. (2010). *Analysis of ordinal categorical data* (Second ed.). John Wiley & Sons.

Agresti, A. and M. Kateri (2017). Ordinal probability effect measures for group comparisons in multinomial cumulative link models. *Biometrics 73*(1), 214–219.

Balke, A. and J. Pearl (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association 92*(439), 1171–1176.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390.

Cheng, J. (2009). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics 65*(1), 96–103.

Chiba, Y. (2018). Bayesian inference of causal effects for an ordinal outcome in randomized trials. *Journal of Causal Inference 6*(2), Article 20170019.

Ciolino, J. D., H. L. Palac, A. Yang, M. Vaca, and H. M. Belli (2019). Ideal vs. real: a systematic review on handling covariates in randomized controlled trials. *BMC Medical Research Methodology 19*(1), Article 136.

Díaz, I., E. Colantuoni, and M. Rosenblum (2016). Enhanced precision in the analysis of randomized trials with ordinal outcomes. *Biometrics 72*(2), 422–431.

Ding, P. and T. Dasgupta (2016). A potential tale of two-by-two tables from completely randomized experiments. *Journal of the American Statistical Association 111*(513), 157–168.

Ding, P., A. Feller, and L. Miratrix (2019). Decomposing treatment effect variation. *Journal of the American Statistical Association 114*(525), 304–317.

D'Amico, G., J. G. Abraldes, P. Rebora, M. G. Valsecchi, and G. Garcia-Tsao (2020). Ordinal outcomes are superior to binary outcomes for designing and evaluating clinical trials in compensated cirrhosis. *Hepatology 72*(3), 1029–1042.

Fisher, R. A. (1992). The arrangement of field experiments. In S. Kotz and N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution*, pp. 82–91. New York: Springer.

Frandsen, B. R. and L. J. Lefgren (2021). Partial identification of the distribution of treatment effects with an application to the Knowledge is Power Program (KIPP). *Quantitative Economics 12*(1), 143–171.

Fréchet, M. (1935). Généralisation du théorème des probabilités totales. *Fundamenta Mathematicae 25*(1), 379–387.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics 40*(2), 180–193.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer.

Horowitz, J. L. and C. F. Manski (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association 95*(449), 77–84.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Huang, E. J., E. X. Fang, D. F. Hanley, and M. Rosenblum (2017). Inequality in treatment benefits: Can we determine if a new treatment benefits the many or the few? *Biostatistics 18*(2), 308–324.

Imai, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine 27*(24), 4857–4873.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jiang, Z. and P. Ding (2018). Using missing types to improve partial identification with application to a study of HIV prevalence in Malawi. *The Annals of Applied Statistics 12*(3), 1831–1852.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*(3), 1071–1102.

Lei, L. and P. Ding (2021). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika 108*(4), 815–828.

Li, B., H. F. Lingsma, E. W. Steyerberg, and E. Lesaffre (2011). Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Medical Research Methodology 11*, Article 77.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics 7*(1), 295–318.

Lin, Y., M. Zhu, and Z. Su (2015). The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials 45*, 21–25.

Liu, H. and Y. Yang (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika 107*(4), 935–948.

Lu, J., P. Ding, and T. Dasgupta (2018). Treatment effects on ordinal outcomes: Causal estimands and sharp bounds. *Journal of Educational and Behavioral Statistics 43*(5), 540–567.

Lu, J., Y. Zhang, and P. Ding (2020). Sharp bounds on the relative treatment effect for ordinal outcomes. *Biometrics 76*(2), 664–669.

Ma, W., Y. Qin, Y. Li, and F. Hu (2020). Statistical inference for covariate-adaptive randomization procedures. *Journal of the American Statistical Association 115*(531), 1488–1497.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.

Miao, W., W. Hu, E. L. Ogburn, and X.-H. Zhou (2023). Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association 118*(543), 1953–1967.

Miratrix, L. W., J. S. Sekhon, and B. Yu (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology 75*(2), 369–396.

Morgan, K. L. and D. B. Rubin (2015). Rerandomization to balance tiers of covariates. *Journal of the American Statistical Association 110*(512), 1412–1421.

Praet, M. and A. Desoete (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., M. G. Hudgens, P. B. Gilbert, and J. P. Fine (2014). Nonparametric bounds and sensitivity analysis of treatment effects. *Statistical Science 29*(4), 596–618.

Rosenberger, W. F. and O. Sverdlov (2008). Handling covariates in the design of clinical trials. *Statistical Science 23*(3), 404–419.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics 6*(1), 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association 75*(371), 591–593.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics 2*(3), 808–840.

Rüschendorf, L. (1981). Sharpness of Fréchet-bounds. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 57*(2), 293–302.

Scott, S. C., M. S. Goldberg, and N. E. Mayo (1997). Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology 50*(1), 45–55.

Splawa-Neyman, J., D. M. Dabrowska, and T. P. Speed (1990). On the application of probability theory to agricultural

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0495

experiments. essay on principles. section 9. *Statistical Science 5*(4), 465–472.

Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics 36*(2), 423–439.

Tu, F., W. Ma, and H. Liu (2024). A unified framework for covariate adjustment under stratified randomisation. *Stat 13*(4), Article e70016.

Volfovsky, A., E. M. Airoldi, and D. B. Rubin (2015). Causal inference for ordinal outcomes. *arXiv preprint arXiv:1501.01234*.

Wu, P., P. Ding, Z. Geng, and Y. Liu (2024). Quantifying individual risk for binary outcome. *arXiv preprint arXiv:2402.10537*.

Yue, L., G. Li, H. Lian, and X. Wan (2019). Regression adjustment for treatment effect with multicollinearity in high dimensions. *Computational Statistics & Data Analysis 134*, 17–35.

Zhang, C., Z. Geng, W. Li, and P. Ding (2024). Identifying and bounding the probability of necessity for causes of effects with ordinal outcomes. *arXiv preprint arXiv:2411.01234*.

Zhao, A. and P. Ding (2024). To adjust or not to adjust? estimating the average treatment effect in randomized experiments with missing covariates. *Journal of the American Statistical Association 119*(545), 450–460.

Guangyuran Liao

School of Statistics and Mathematics, Central University of Finance and Economics

E-mail: lgyr@email.cufe.edu.cn

Yuehan Yang

School of Statistics and Mathematics, Central University of Finance and Economics

E-mail: yyh@cufe.edu.cn