Statistica Sinica Preprint No: SS-2025-0330	
Title	Optimal Treatment Allocations Accounting for Population Differences
Manuscript ID	SS-2025-0330
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0330
Complete List of Authors	Wei Zhang, Zhiwei Zhang and Aiyi Liu
Corresponding Authors	Wei Zhang
E-mails	zhangwei@amss.ac.cn
Notice: Accepted author version.	

Optimal Treatment Allocations Accounting for Population Differences

Wei Zhang[1], Zhiwei Zhang[2,*], and Aiyi Liu[3]

[1]State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

[2]Biostatistics Innovation Group, Gilead Sciences, Foster City, California, USA

[3]Biostatistics and Bioinformatics Branch, Division of Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

*zhiwei.zhang6@gilead.com

Abstract

The treatment allocation mechanism in a randomized clinical trial can be optimized by maximizing the nonparametric efficiency bound for a specific measure of treatment effect. Optimal treatment allocations which may or may not depend on baseline covariates have been derived for a variety of effect measures focusing on the trial population, the patient population represented by the trial participants. Frequently, clinical trial data are used to estimate treatment effects in a target population that is related to but different from the trial population. This article provides optimal treatment allocations that account for the impact of such population differences. We consider three cases with different data configurations: transportation, generalization, and post-stratification. Our results indicate that, for general effect measures, optimal treatment allocations may depend on the covariate distribution in the target population but not on the configuration of data or information that

*ORCID ID of Wei Zhang: https://orcid.org/0000-0002-4315-8380

describes the target covariate distribution. For estimating average treatment effects, there is a unique covariate-dependent allocation that achieves maximal efficiency regardless of the target covariate distribution and the associated data configuration.

Key words: covariate adjustment; covariate-dependent randomization; generalizability; optimal design; propensity score; transportability

1 Introduction

The treatment allocation mechanism in a randomized clinical trial can be optimized for statistical efficiency. A well-known result in this domain is the Neyman allocation (Neyman, 1934), which makes the size of each treatment group proportional to the standard deviation of the outcome of interest in the same treatment group. The Neyman allocation is optimal, in the sense of minimum asymptotic variance, for estimating the average treatment effect (ATE) with the sample mean difference. The sample mean difference is commonly used but not fully efficient in the presence of baseline covariates associated with treatment outcomes. Motivated by semiparametric theory (Bickel et al., 1993; Tsiatis, 2006), more efficient estimators have been developed to incorporate information from baseline covariates (e.g., Tsiatis et al., 2008; Zhang et al., 2008; Moore and van der Laan, 2009; Rosenblum and van der Laan, 2010; Tian et al., 2012; Zhang and Ma, 2019; Ye et al., 2023; Bannick et al., 2025). Some of these estimators leverage modern machine learning and ensemble learning methods (e.g., Hastie et al., 2009; Polley et al., 2011) and have a better chance (than those based on parametric models) to attain or approach the nonparametric efficiency bound for treatment effect estimation. In light of these recent advances in efficient estimation of treatment effects, it makes sense to

optimize treatment allocation for an efficient estimator attaining the nonparametric efficiency bound. Optimal treatment allocations that maximize the nonparametric efficiency bound have been derived by Zhang et al. (2023), who considered both the traditional covariate-independent randomization (CIR) design and a covariate-dependent randomization (CDR) design that resembles observational studies except that the propensity score is specified and known by the investigator. These efforts to improve statistical efficiency generally aim at a treatment effect measure, such as the ATE, in the trial population (i.e., the patient population represented by the trial participants).

Clinical trial data are frequently used to estimate treatment effects in a target patient population related to but different from the trial population. This is done for various reasons, and we give two examples here. First, the trial cohort is often a convenience sample and may not be truly representative of the patient population for treatment indication. If the latter population is better described by another data source, say a disease registry, it is of interest to combine trial data with registry information, which may be individual patient data or summary statistics, to estimate the effect of the experimental treatment in the patient population represented by the disease registry (Cole and Stuart, 2010; Stuart et al., 2011; Dahabreh et al., 2019; Cappiello et al., 2021). Second, calibrating the treatment effect observed in one trial to another trial population can be useful for making an indirect comparison of two treatments that have not been compared in a randomized fashion. For example, when an experimental treatment is evaluated against an active control treatment in a randomized non-inferiority trial, investigators and regulators often wonder how the experimental treatment would compare with a placebo control in the same setting. If the active control has been compared with placebo in a previous randomized trial, this question

can be answered by calibrating the active control effect from the previous trial to the non-inferiority trial population (Zhang, 2009; Nie et al., 2013; Zhang et al., 2016).

The increasing use of clinical trial data to estimate treatment effects in non-trial populations raises important questions about clinical trial design, particularly treatment allocation. For example, do the optimal CIR and CDR allocations for the trial population remain optimal for a different target population? If so, under what conditions? If not, which aspects of the target population impact optimal allocations and in what manner? These questions are highly relevant to consider in designing a clinical trial and, to our knowledge, have not been considered in the current statistical literature.

This article addresses the above questions by providing optimal treatment allocations for estimating treatment effects in a target population that overlaps with but differs from the trial population in terms of baseline covariates. We consider three cases with different data configurations. In the first case (transportation), individual covariate data are available in a target cohort (a random sample from the target population) which does not include any trial participants (e.g., Zhang et al., 2016; Rudolph and van der Laan, 2017; Cappiello et al., 2021). The second case (generalization) is similar to the first one except that the target cohort contains the trial cohort as a subset (Stuart et al., 2011; Dahabreh et al., 2019). In the third case (post-stratification), the target population is the result of reweighting the trial population with a set of known weights assigned to a finite number of strata such as demographic subgroups (Breslow and Day, 1987; Cole and Stuart, 2010). Post-stratification is often conducted to approximate a "real" target population (e.g., HIV-infected adults in the United States) for which individual-level data are unavailable and only summary statistics are available (Cole and Stuart, 2010). In each case, we derive the optimal CIR and CDR

designs, the latter being generally more efficient because CDR includes CIR as a special case. Interestingly, the optimal CIR and CDR designs can be characterized using general expressions that simultaneously apply to all three cases, indicating that optimal treatment allocations may depend on the covariate distribution in the target population but not on the configuration of data or information that describes the target covariate distribution. For estimating the ATE in the target population, the optimal CDR design does not even depend on the target covariate distribution and is identical to the optimal CDR design for estimating the ATE in the trial population (Zhang et al., 2023). Thus, for ATE estimation, the optimal CDR design has a desirable invariance property not shared by optimal CIR designs, in addition to an efficiency advantage.

The rest of the article is organized as follows. In the next section, we review existing results and present new ones on optimal treatment allocations. In Section 3, we report numerical results from simulation studies and a real example. The article ends with a discussion in Section 4.

2 Optimal Treatment Allocations

2.1 Preliminaries

Suppose a randomized trial is to be conducted to compare an experimental treatment with a control treatment. For a generic patient in the trial, let $Y(a)$ denote the potential outcome for the experimental ($a = 1$) or control ($a = 0$) treatment, W a vector of baseline covariates that may be associated with one or both potential outcomes, $A \in \{0, 1\}$ the randomly

assigned treatment, and $Y = Y(A) = AY(1) + (1 - A)Y(0)$ the actual observed outcome. For treatment allocation, we consider both CIR and CDR, defined as

$$\mathrm{pr}\{A = 1|W, Y(0), Y(1)\} = \begin{cases} \mathrm{pr}(A = 1) = \pi \in (0, 1) & \text{(CIR)} \\ \mathrm{pr}(A = 1|W) = p(W) \in (0, 1) & \text{(CDR)} \end{cases},$$

where $p(W)$ is known as the propensity score for treatment assignment. CDR may look like an observational study but is fundamentally a randomized trial in that the treatment assignment mechanism is entirely controlled by the investigator. From a causal inference perspective, CIR and CDR provide the same level of statistical rigor as they both allow causal estimands to be identified and estimated without making strong assumptions (Zhang et al., 2023). The trial cohort is regarded as a random sample from some patient population, which we call the trial population. Accordingly, the observed trial data are conceptualized as independent copies of (W, A, Y) and denoted by (W_i, A_i, Y_i) $(i = 1, \ldots, n)$.

Our general objective is to optimize treatment allocation, as determined by π or p, by maximizing the nonparametric efficiency bound for estimating a specified effect measure. Consider, for example, $\Delta = \mathrm{E}\{Y(1) - Y(0)\}$, the ATE in the trial population. According to Zhang et al. (2023), for estimating Δ using the trial data, the optimal CIR design is given by

$$\pi_{\mathrm{opt}} = \frac{[\mathrm{E}\{v_1(W)\}]^{1/2}}{[\mathrm{E}\{v_1(W)\}]^{1/2} + [\mathrm{E}\{v_0(W)\}]^{1/2}},$$

and the optimal CDR design by

$$p_{\mathrm{opt}}(W) = \frac{v_1(W)^{1/2}}{v_1(W)^{1/2} + v_0(W)^{1/2}}, \tag{1}$$

where $v_a(W) = \mathrm{var}\{Y(a)|W\}$, $a = 0, 1$. The optimal CDR design can be seen as applying the Neyman allocation to each sub-population defined by W.

This article is mainly concerned with effect measures of the form $\Delta^* = \int \delta(w)\mathrm{d}F^*(w)$, where $\delta(w) = m(1,w) - m(0,w)$, $m(a,w) = \mathrm{E}\{Y(a)|W = w\}$, and F^* is the covariate distribution in some target population. (Additional effect measures are considered in Section 2.5.) Under a certain exchangeability assumption (e.g., Zhang et al., 2016; Dahabreh et al., 2019; Colnet et al., 2024), Δ^* is the ATE in the target population. More generally, Δ^* can be interpreted as a calibrated or adjusted treatment effect that accounts for population differences. Let F denote the distribution of W in the trial population. Throughout, we make the following assumption:

Assumption 1. F and F^* dominate each other.

Let r denote the density of F^* with respect to F, which is also known as the density ratio of F^* to F with respect to a common dominating measure. Assumption 1 implies that r exists and that $\mathrm{pr}\{0 < r(W) < \infty\} = 1$. Assumption 1 ensures that Δ^* is identifiable if F^* is known or identifiable. Information about F^* may come in different forms, as indicated in Section 1. In the next three subsections, we elaborate on the three cases mentioned in Section 1. Any new notations we use to describe a case are only applicable within that context.

Remark 1. Assumption 1 essentially states that $\mathcal{W} = \mathcal{W}^*$, where $\mathcal{W}$ (rsp. $\mathcal{W}^*$) denotes the support of F (rsp. F^*). Technically, we only need $\mathcal{W} \supseteq \mathcal{W}^*$ for identifiability. If $\mathcal{W} \supsetneq \mathcal{W}^*$, trial participants with $W_i \in \mathcal{W} \setminus \mathcal{W}^*$ are not informative of Δ^* without making strong assumptions, and we therefore exclude such patients from consideration, effectively restricting F to $\mathcal{W}^*$. As an example, consider estimating the ATE in a subpopulation

$\mathcal{W}^* \subset \mathcal{W}$. For this purpose, the optimal value of π is easily seen to be

$$\frac{[\mathrm{E}\{v_1(W)|W \in \mathcal{W}^*\}]^{1/2}}{[\mathrm{E}\{v_1(W)|W \in \mathcal{W}^*\}]^{1/2} + [\mathrm{E}\{v_0(W)|W \in \mathcal{W}^*\}]^{1/2}},$$

and the optimal propensity score is (1) for $W \in \mathcal{W}^*$. In the event $\mathcal{W} \not\supseteq \mathcal{W}^*$, we restrict both distributions to the common support $\mathcal{W} \cap \mathcal{W}^*$, and redefine (F, F^*) as the restricted distributions so as to satisfy Assumption 1.

2.2 The Case of Transportation

In this case, F^* is described by observed covariate values (denoted by W_j^*, $j = 1, \ldots, n^*$) in a target cohort, a random sample from the target population that shares no patients with the trial cohort. We assume that the W_j^*'s are independent of each other and of $\{(W_i, A_i, Y_i), i = 1, \ldots, n\}$, and identically distributed according to F^*. For a generic patient in the trial or target cohort, let Z be a source indicator ($Z = 1$ if in the trial; 0 otherwise), and let $X = (W, A, Y)$ (if $Z = 1$) or W^* (if $Z = 0$). Then the combined data, $\{(W_i, A_i, Y_i), i = 1, \ldots, n\} \cup \{W_j^*, j = 1, \ldots, n^*\}$, can be considered independent copies of $O = (Z, X)$.

Under CIR, Zhang et al. (2016) derived the efficient influence function for estimating Δ^* as

$$\psi_{\mathrm{CIR}}^{\mathrm{tr}}(O) = \frac{(1-Z)\{\delta(W^*) - \Delta^*\}}{1-\gamma} + \frac{ZAr(W)\{Y - m(1,W)\}}{\gamma\pi} - \frac{Z(1-A)r(W)\{Y - m(0,W)\}}{\gamma(1-\pi)},$$

where $\gamma = \mathrm{pr}(Z = 1)$ and the superscript tr represents transportation. This result has motivated doubly robust, locally efficient estimators of Δ^* based on working parametric models for (m, r) (Zhang et al., 2016). The nuisance functions (m, r) can also be estimated using nonparametric machine learning methods, leading to nonparametric estimators of Δ^*

that are consistent, asymptotically linear, and efficient under mild conditions (Cappiello et al., 2021).

Lemma 1. *In the case of transportation, the efficient influence function for estimating* Δ^* *under CDR is*

$$\psi^{\mathrm{tr}}_{\mathrm{CDR}}(O) = \frac{(1-Z)\{\delta(W^*)-\Delta^*\}}{1-\gamma} + \frac{ZAr(W)\{Y-m(1,W)\}}{\gamma p(W)} - \frac{Z(1-A)r(W)\{Y-m(0,W)\}}{\gamma\{1-p(W)\}}.$$

Thus, $\psi^{\mathrm{tr}}_{\mathrm{CDR}}(O)$ includes $\psi^{\mathrm{tr}}_{\mathrm{CIR}}(O)$ as a special case with $p(W) \equiv \pi$. The existing estimators of Δ^*, developed for CIR, can be extended to CDR while maintaining their essential properties; details are given in Section D.1 of Supplementary Materials.

The optimal CIR and CDR designs for transportation can be derived by minimizing $\mathrm{var}\{\psi^{\mathrm{tr}}_{\mathrm{CIR}}(O)\}$ and $\mathrm{var}\{\psi^{\mathrm{tr}}_{\mathrm{CDR}}(O)\}$ with respect to π and p, respectively.

Theorem 1. *Under Assumption 1,* $\mathrm{var}\{\psi^{\mathrm{tr}}_{\mathrm{CIR}}(O)\}$ *is minimized uniquely by setting* π *equal to*

$$\pi^*_{\mathrm{opt}} = \frac{[\mathrm{E}\{r(W)^2 v_1(W)\}]^{1/2}}{[\mathrm{E}\{r(W)^2 v_1(W)\}]^{1/2} + [\mathrm{E}\{r(W)^2 v_0(W)\}]^{1/2}}, \tag{2}$$

and $\mathrm{var}\{\psi^{\mathrm{tr}}_{\mathrm{CDR}}(O)\}$ *is minimized uniquely by setting* p *equal to* p_{opt} *defined by* (1).

Note that the optimal CDR design does not depend on F^* whereas the optimal CIR design does. If $F^* = F$, then $r(W) \equiv 1$ and $\pi^*_{\mathrm{opt}} = \pi_{\mathrm{opt}}$; otherwise, π^*_{opt} may differ from π_{opt}.

2.3 The Case of Generalization

In this case, F^* is described by observed covariate values (denoted by W^*_j, $j = 1, \ldots, N$) in a target cohort, a random sample from the target population that contains the trial cohort as a subset. Note that the trial cohort need not be a random sub-sample of the target cohort.

For a generic patient in the target cohort, let Z be an indicator for trial participation ($Z = 1$ if in the trial; 0 otherwise), and let $X = (W, A, Y) = (W^*, A, Y)$ (if $Z = 1$) or W^* (if $Z = 0$). In the present notation, $W^* \sim F^*$ and $(W^*|Z = 1) \sim W \sim F$. The available data, $\{(W_i, A_i, Y_i), i = 1, \ldots, n\} \cup \{W_j^* : Z_j = 0\}$, can be considered independent copies of $O = (Z, X)$.

Under CIR, Dahabreh et al. (2019) derived the efficient influence function for estimating Δ^* as

$$\psi_{\text{CIR}}^{\text{gen}}(O) = \delta(W^*) - \Delta^* + \frac{ZA\{Y - m(1, W^*)\}}{e(W^*)\pi} - \frac{Z(1-A)\{Y - m(0, W^*)\}}{e(W^*)(1-\pi)},$$

where the superscript gen represents generalization and $e(W^*) = \text{pr}(Z = 1|W^*)$ is the propensity score for trial participation. Assumption 1 implies that $\text{pr}\{0 < e(W^*) < 1\} = 1$. This result has been used to construct doubly robust, locally efficient estimators of Δ^* based on working parametric models for (m, e) (Dahabreh et al., 2019).

Lemma 2. *In the case of generalization, the efficient influence function for estimating* Δ^* *under CDR is given by*

$$\psi_{\text{CDR}}^{gen}(O) = \delta(W^*) - \Delta^* + \frac{ZA\{Y - m(1, W^*)\}}{e(W^*)p(W^*)} - \frac{Z(1-A)\{Y - m(0, W^*)\}}{e(W^*)\{1 - p(W^*)\}}.$$

Section D.2 of Supplementary Materials describes estimators of Δ^* based on these efficient influence functions, where the nuisance functions (m, e) may be estimated using parametric models or nonparametric machine learning methods.

Minimizing $\text{var}\{\psi_{\text{CIR}}^{\text{gen}}(O)\}$ and $\text{var}\{\psi_{\text{CDR}}^{\text{gen}}(O)\}$ with respect to π and p, respectively, yields the optimal CIR and CDR designs for generalization. For comparability (with the other cases), the optimal CIR design will be expressed using the general notation $r(W)$, which in the present setting is inversely proportional to $e(W)$: $r(W) = \text{pr}(Z = 1)/e(W)$.

Theorem 2. *Under Assumption 1,* $\text{var}\{\psi^{gen}_{\text{CIR}}(O)\}$ *is minimized uniquely by setting* π *equal to* π^*_{opt} *defined by* (2)*, and* $\text{var}\{\psi^{gen}_{\text{CDR}}(O)\}$ *is minimized uniquely by setting* p *equal to* p_{opt} *defined by* (1)*.*

Thus, despite the use of a different sampling mechanism, the optimal treatment allocations for generalization are identical to those for transportation.

2.4 The Case of Post-Stratification

In this case, F^* is not described by individual patient data but by a known stratification with known weights. Let $\mathcal{W}$ be partitioned as $\cup_{k=1}^K \mathcal{W}_k$ with $\mathcal{W}_k \cap \mathcal{W}_l = \emptyset$ for $k \neq l$, and let $(\tau_1^*, \ldots, \tau_K^*)$ be a set of known weights such that $\sum_{k=1}^K \tau_k^* = 1$ and $\tau_k^* > 0$, $k = 1, \ldots, K$. For example, the $\mathcal{W}_k$'s may be demographic subgroups based on age, gender and/or race, and the τ_k^*'s may be obtained from a census or a disease registry much larger than the trial cohort (Cole and Stuart, 2010). Based on this information, F^* may be defined as follows:

$$F^*(\mathcal{B}) = \sum_{k=1}^K \tau_k^* \,\text{pr}(W \in \mathcal{B} | W \in \mathcal{W}_k) = \sum_{k=1}^K \frac{\tau_k^* F(\mathcal{B} \cap \mathcal{W}_k)}{F(\mathcal{W}_k)}$$

for any measurable subset $\mathcal{B}$ of $\mathcal{W}$. It follows that $\Delta^* = \int \delta(w) \text{d}F^*(w) = \sum_{k=1}^K \tau_k^* \Delta_k$, where $\Delta_k = \text{E}\{Y(1) - Y(0) | W \in \mathcal{W}_k\}$, $k = 1, \ldots, K$. In words, Δ^* is a weighted average of subgroup-specific ATE's in the trial population.

Because the τ_k^*'s are known constants, the efficient influence function for estimating Δ^* is a weighted average of those for estimating the Δ_k's. Let $O = (W, A, Y)$, let $I(\cdot)$ be the indicator function, and let $\tau_k = \text{pr}(W \in \mathcal{W}_k)$, $k = 1, \ldots, K$.

Lemma 3. *In the case of post-stratification, the efficient influence function for estimating*

Δ^* *is*

$$\psi^{\text{ps}}_{\text{CIR}}(O) = \sum_{k=1}^{K} \frac{I(W \in \mathcal{W}_k)\tau_k^*}{\tau_k} \left[\delta(W) - \Delta_k + \frac{A\{Y - m(1,W)\}}{\pi} - \frac{(1-A)\{Y - m(0,W)\}}{1-\pi}\right]$$

under CIR, and

$$\psi^{\text{ps}}_{\text{CDR}}(O) = \sum_{k=1}^{K} \frac{I(W \in \mathcal{W}_k)\tau_k^*}{\tau_k} \left[\delta(W) - \Delta_k + \frac{A\{Y - m(1,W)\}}{p(W)} - \frac{(1-A)\{Y - m(0,W)\}}{1-p(W)}\right]$$

under CDR.

Efficient estimation of Δ^* can be approached by first estimating each Δ_k efficiently. The latter can be achieved by applying an efficient method for ATE estimation to the subgroup of trial participants with $W_i \in \mathcal{W}_k$. General approaches to efficient estimation of the ATE under both CIR and CDR are available in the causal inference literature (e.g., van der Laan and Robins, 2003; van der Laan and Rose, 2011; Zhang et al., 2023). Explicit formulas are provided in Section D.3 of Supplementary Materials.

The next result provides the optimal CIR and CDR designs for post-stratification. In the present setting, we have $r(W) = \sum_{k=1}^{K} I(W \in \mathcal{W}_k)\tau_k^*/\tau_k$.

Theorem 3. *Under Assumption 1,* $\text{var}\{\psi^{\text{ps}}_{\text{CIR}}(O)\}$ *is minimized uniquely by setting* $\pi = \pi^*_{\text{opt}}$ *defined by* (2)*, and* $\text{var}\{\psi^{\text{ps}}_{\text{CDR}}(O)\}$ *is minimized uniquely by setting* $p = p_{\text{opt}}$ *defined by* (1)*.*

In the three cases considered so far, the optimal CIR design depends on F^* but not on the configuration of data or information describing F^*, and the optimal CDR design does not even depend on F^*.

2.5 Extension to Other Effect Measures

The results in Sections 2.2–2.4 can be extended to other effect measures of the form $\Delta^*(g) = g(\mu_1^*) - g(\mu_0^*)$, where $\mu_a^* = \int m(a,w)\mathrm{d}F^*(w)$, $a = 0, 1$, and g is a specified differentiable and increasing function. If g is the identity function, then $\Delta^*(g) = \Delta^*$. Other examples of g include the log and logit functions. Under a suitable mean exchangeability assumption (e.g., Zhang et al., 2016; Dahabreh et al., 2019; Colnet et al., 2024), μ_a^* represents the mean potential outcome for treatment $a \in \{0, 1\}$ in the target population, and $\Delta^*(g)$ is a contrast of μ_1^* with μ_0^*. We assume that Assumption 1 holds and write g' for the derivative function of g.

The same arguments made in the proofs of Lemmas 1-3 can be used to derive efficient influence functions for estimating $\Delta^*(g)$ under both CIR and CDR in the three cases considered (transportation, generalization, and post-stratification). For transportation, the efficient influence function under CDR is

$$\psi_{\mathrm{CDR}}^{\mathrm{tr}}(O;g) = g'(\mu_1^*)\left[\frac{(1-Z)\{m(1,W^*)-\mu_1^*\}}{1-\gamma} + \frac{ZAr(W)\{Y-m(1,W)\}}{\gamma p(W)}\right] - g'(\mu_0^*)\left[\frac{(1-Z)\{m(0,W^*)-\mu_0^*\}}{1-\gamma} + \frac{Z(1-A)r(W)\{Y-m(0,W)\}}{\gamma\{1-p(W)\}}\right],$$

where $\gamma = \mathrm{pr}(Z = 1)$ is the probability of trial membership. The corresponding efficient influence function under CIR, denoted by $\psi_{\mathrm{CIR}}^{\mathrm{tr}}(O;g)$, is obtained by replacing $p(W)$ with the constant π. For generalization, the efficient influence function under CDR is

$$\psi_{\mathrm{CDR}}^{\mathrm{gen}}(O;g) = g'(\mu_1^*)\left[\{m(1,W^*)-\mu_1^*\} + \frac{ZA\{Y-m(1,W^*)\}}{e(W^*)p(W^*)}\right] - g'(\mu_0^*)\left[\{m(0,W^*)-\mu_0^*\} + \frac{Z(1-A)\{Y-m(0,W^*)\}}{e(W^*)\{1-p(W^*)\}}\right].$$

Replacing $p(W^*)$ with π yields the corresponding efficient influence function under CIR,

denoted by $\psi^{\text{gen}}_{\text{CIR}}(O;g)$. For post-stratification, the efficient influence function under CDR is

$$\psi^{\text{ps}}_{\text{CDR}}(O;g) = g'(\mu_1^*)\sum_{k=1}^{K}\frac{I(W\in\mathcal{W}_k)\tau_k^*}{\tau_k}\left[\{m(1,W)-\mu_{1k}\}+\frac{A\{Y-m(1,W)\}}{p(W)}\right]$$
$$- g'(\mu_0^*)\sum_{k=1}^{K}\frac{I(W\in\mathcal{W}_k)\tau_k^*}{\tau_k}\left[\{m(0,W)-\mu_{0k}\}+\frac{(1-A)\{Y-m(0,W)\}}{1-p(W)}\right],$$

where $\mu_{1k} = \text{E}\{Y(1)|W\in\mathcal{W}_k\}$ and $\mu_{0k} = \text{E}\{Y(0)|W\in\mathcal{W}_k\}$. Replacing $p(W)$ with π yields the efficient influence function under CIR, denoted by $\psi^{\text{ps}}_{\text{CIR}}(O;g)$. In all three cases, an efficient estimator of $\Delta^*(g) = g(\mu_1^*) - g(\mu_0^*)$ can be obtained by substituting efficient estimators of μ_1^* and μ_0^* (described in Sections 2.2–2.4 and the references therein).

Optimal CIR designs for estimating $\Delta^*(g)$ are obtained by minimizing $\text{var}\{\psi^{\text{tr}}_{\text{CIR}}(O;g)\}$, $\text{var}\{\psi^{\text{gen}}_{\text{CIR}}(O;g)\}$ and $\text{var}\{\psi^{\text{ps}}_{\text{CIR}}(O;g)\}$ with respect to π. In each case, the optimal CIR allocation takes the form

$$\pi^*_{\text{opt}}(g) = \frac{g'(\mu_1^*)[\text{E}\{r(W)^2v_1(W)\}]^{1/2}}{g'(\mu_1^*)[\text{E}\{r(W)^2v_1(W)\}]^{1/2}+g'(\mu_0^*)[\text{E}\{r(W)^2v_0(W)\}]^{1/2}}.$$

For CDR, the optimal propensity score is given by

$$p^*_{\text{opt}}(W;g) = \{g'(\mu_1^*)v_1(W)^{1/2}\}/\{g'(\mu_1^*)v_1(W)^{1/2}+g'(\mu_0^*)v_0(W)^{1/2}\}$$

in all three cases. These results follow from straightforward modifications of the proofs of Theorems 1–3; details are omitted. Note that, when g is nonlinear, the optimal CDR design depends on F^* through (μ_0^*,μ_1^*).

2.6 Implementation

The optimal CDR designs considered in this article are generally easier to implement than their CIR counterparts. All of these designs require information about the conditional variance functions (v_1,v_0) or at least the ratio v_1/v_0, and the optimal CIR designs also require

information about the density ratio r, which involves the covariate distributions in the trial and target populations. It is unlikely that both covariate distributions are well understood at the design stage, which makes it challenging to implement these optimal CIR designs in practice.

The optimal CDR designs for ATE estimation can be implemented using the general strategies proposed by Zhang et al. (2023) and Zhang et al. (2025). The functions (v_1, v_0) can be estimated from preliminary data (if available) as in Zhang et al. (2023) or from accruing trial data under the adaptive design developed by Zhang et al. (2025). In general, the amount of relevant information available should be a major consideration in defining the scope of design optimization. For example, if the amount of preliminary data is small, the (initial) optimization could be restricted to a coarsened version of W such as a subset of covariates or a discretization (Zhang et al., 2023). As more information becomes available, the scope of design optimization could be broadened successively under the adaptive design of Zhang et al. (2025).

These general strategies remain applicable if the estimand is $\Delta^*(g)$ with a nonlinear function g. For such estimands, design optimization requires preliminary information about (μ_1^*, μ_0^*) in addition to (v_1, v_0); this is true for both CDR and CIR designs. It may be difficult to estimate (μ_1^*, μ_0^*) in a data-driven manner when the trial is still being designed and the target population may or may not be well characterized. However, scientific background information may be available to suggest a range of plausible values for (μ_1^*, μ_0^*), which could be used to approximate the optimal CDR design. In this case, it is important to assess the sensitivity of the optimal CDR design to the values of (μ_1^*, μ_0^*) within the range of plausible values.

When the unknown quantities in an optimal design are replaced by empirical estimates, the resulting "optimized" design is generally sub-optimal when considered as a fixed design, and its actual efficiency depends on the quality of the underlying estimates. Under a broader asymptotic framework that accommodates a sequence of data-driven designs, it seems reasonable to expect an optimized design based on consistent estimates to be asymptotically efficient and equivalent to the true optimal design in a suitable sense. Formalizing this conjecture is beyond the scope of this article.

3 Numerical Results

3.1 Simulation

Imagine we are designing a randomized trial in a specific patient population with two important baseline covariates. Specifically, $W = (W_1, W_2)'$ with $W_1 \sim N(0, 0.75^2; -2, 2)$ and $W_2 \sim B(0.3)$, independent of each other. Here and in the sequel, $N(\mu, \sigma^2; l, u) \sim (X | l \leq X \leq u)$ with $X \sim N(\mu, \sigma^2)$ and $B(q)$ denotes the Bernoulli distribution with success probability q. Let us consider a continuous outcome first. Given W, the potential outcomes are distributed as follows:

$$Y(0) \sim N\big(W_1 + W_2 + 0.1W_1W_2 - 0.2W_1^2, \exp(-2 + W_1 + 2W_2)\big),$$

$$Y(1) \sim N\big(1 + W_2 - 0.2W_1W_2 + 0.1W_1^2, \exp(1 - W_1 - 2W_2)\big).$$

The possible dependence of $Y(0)$ and $Y(1)$ given W is irrelevant and left unspecified.

The immediate objective of the trial is to estimate the ATE in the trial population, $\Delta \approx 1.158$. The trial data may also be used for transportation, generalization, or post-stratification. In the case of transportation, the target population is defined by $W_1^* \sim$

$N(0.5, 1; -2, 2)$ and $W_2^* \sim B(0.6)$, independent of each other, and the estimand is $\Delta_{\text{tr}}^* \approx 0.814$. In the case of generalization, the target population is a 1:1 mixture of the trial population and the target population for transportation. To be precise, we may write $F_{\text{gen}}^* = 0.5F + 0.5F_{\text{tr}}^*$, where F_{gen}^* and F_{tr}^* are the target covariate distributions for generalization and transportation, respectively. Note that, under F_{gen}^*, W_1^* and W_2^* are no longer independent of each other. The estimand in this case is $\Delta_{\text{gen}}^* \approx 0.986$. In the case of post-stratification, we partition $\mathcal{W} = [-2, 2] \times \{0, 1\}$ into four strata: $\mathcal{W}_1 = [-2, 0) \times \{0\}$, $\mathcal{W}_2 = [0, 2] \times \{0\}$, $\mathcal{W}_3 = [-2, 0) \times \{1\}$, and $\mathcal{W}_4 = [0, 2] \times \{1\}$. The associated weights are $(\tau_1^*, \ldots, \tau_4^*) = (0.1, 0.2, 0.3, 0.4)$, and the resulting estimand value is $\Delta_{\text{ps}}^* \approx 1.024$.

As noted in Section 2, the four estimands share the same optimal CDR design, which is depicted in Figure 1. The four estimands do correspond to different optimal CIR designs, which are shown in Table 1. We also consider 1:1 CIR as a comparator. For each of the six designs, we set $n = 500$ for the trial cohort and $n^* = 500$ for transportation, and sample N from a negative binomial distribution (with the goal of enrolling n patients into the trial) in the case of generalization. Each design is simulated 2,000 times. The resulting data are analyzed to estimate the four estimands using the debiased machine learning approach (Chernozhukov et al., 2018), where the nuisance functions are estimated using a super learner ensemble that combines three algorithms: generalized linear regression with an interaction between the two covariates, generalized additive regression, and regression tree. The super learner is based on a five-fold cross-validation, which is nested within a five-fold cross-fitting for debiased machine learning. Additional implementation details are provided in Section D of the Supplementary Materials.

Table 1 compares the six designs in terms of efficiency with the 1:1 CIR design as the

reference. The relative efficiency of an optimal design is calculated as the variance ratio of the 1:1 CIR design to the design in question. For each estimand, the corresponding optimal CIR design is indeed more efficient than the other CIR designs. The relative efficiency results of the optimal CIR designs vary widely across the four estimands. In contrast, the optimal CDR design attains the highest level of efficiency for all four estimands, as predicted by theory.

Table S1 of the Supplementary Materials reports additional simulation results: empirical bias, standard deviation, median standard error, and coverage proportion (for 95% Wald confidence intervals). Standard errors are based on estimated efficient influence functions (with cross-fitted nuisance function estimates). The results in Table S1 show that point estimators are virtually unbiased and that confidence intervals achieve close-to-nominal coverage in most cases. In a few exceptional cases (e.g., estimating Δ^*_{tr} and Δ^*_{gen} under π_{opt}), there are some signs of slight under-coverage, which can be resolved by doubling the sample size (results not shown).

Next, we consider estimating log-mean-ratios (LMRs) for a count outcome following Poisson regression models (conditional on $W = (W_1, W_2)'$):

$$Y(0) \sim \text{Poisson}\big(\exp(-1.5 - 2W_1 - 0.5W_2 + 0.05W_1^2)\big),$$

$$Y(1) \sim \text{Poisson}\big(\exp(-3 + 1.5W_1 + W_2 - 0.1W_1W_2)\big).$$

In the trial population, we have $W_1 \sim N(0, 1; -2, 2)$ and $W_2 \sim B(0.3)$, independent of each other. The LMR in the trial population is defined as $\Delta(\log) = \log(\mu_1) - \log(\mu_0)$, where $\mu_a = E\{Y(a)\}$ for $a = 1, 0$; the value of $\Delta(\log)$ is approximately -1.546. For a different target population, the LMR estimand is defined as $\Delta^*(\log) = \log(\mu_1^*) - \log(\mu_0^*)$, where $\mu_a^* = \int m(a, w)\mathrm{d}F^*(w)$. In the case of transportation, the target population is defined

by $W_1^* \sim N(1, 1; -2, 2)$ and $W_2^* \sim B(0.8)$, independent of each other, and the estimand is $\Delta_{\text{tr}}^*(\log) \approx 1.435$. In the case of generalization, we define $F_{\text{gen}}^* = 0.5F + 0.5F_{\text{tr}}^*$, leading to $\Delta_{\text{gen}}^*(\log) \approx -0.150$. For post-stratification, we partition $\mathcal{W} = [-2, 2] \times \{0, 1\}$ into four strata: $\mathcal{W}_1 = [-2, -0.25) \times \{0\}$, $\mathcal{W}_2 = [-0.25, 2] \times \{0\}$, $\mathcal{W}_3 = [-2, -0.25) \times \{1\}$, and $\mathcal{W}_4 = [-0.25, 2] \times \{1\}$. The corresponding weights are given by $(\tau_1^*, \ldots, \tau_4^*) = (0.1, 0.2, 0.3, 0.4)$, which implies $\Delta_{\text{ps}}^*(\log) \approx -0.587$.

In the present setting, there is not a single optimal CDR design but one for each estimand. The corresponding propensity scores are plotted in Figure 2. The optimal CIR designs are shown in Table 2. The same estimation methods used for a continuous outcome are adopted here with obvious modifications (e.g., link function). In all other aspects, this simulation study is identical to the previous one with a continuous outcome. Table 2 reports relative efficiency results with 1:1 CIR as the reference. The results for CIR designs follow the same pattern as those in Table 1. The results for CDR designs are consistent with theoretical predictions and illustrate the lack of a universally optimal CDR design for LMR estimation. Taken together, the results in Table 2 indicate that optimal CDR designs are generally more competitive than their CIR counterparts. Table S2 of the Supplementary Materials reports additional simulation results including bias and coverage, which are similar to those in Table S1.

3.2 Example

We now illustrate our findings by applying them retrospectively to a completed trial. The BENCHMRK trial is a randomized placebo-controlled study of raltegravir, an HIV-1 inte-

grase inhibitor, for treating human immunodeficiency virus type 1 (HIV-1) in treatment-experienced patients (Steigbigel et al., 2008; Cooper et al., 2008). This trial randomized 699 HIV-1 patients in a 2:1 ratio to receive raltegravir or placebo, each combined with optimized background therapy. The outcome measure of interest to us is the virologic response rate at week 48 of treatment, defined as the proportion of patients with viral load below 50 copies/ml, with discontinuation counted as failure. The observed virologic response rate is 60.4% in the raltegravir group and 30.8% in the placebo group, for a difference of 29.6% (95% confidence interval: 22.2–37.0%), which is significantly greater than 0 (one-sided p value < 0.0001).

We first calculate optimal treatment allocations for estimating the ATE of raltegravir in the BENCHMRK trial population. The covariate vector W consists of race (black or not), age, body mass index, log-transformed viral load and CD4 cell count, as well as genotypic and phenotypic sensitivity scores (defined as the number of concurrently used antiretroviral drugs to which a subject's HIV is fully susceptible, based on genotypic or phenotypic resistance testing). In this calculation, we assume that W is distributed as observed in the BENCHMRK trial (excluding 37 patients with missing covariate data) and that, given W, the potential outcomes follow a logistic regression model:

$$\text{logit}[\text{pr}\{Y(a) = 1 \mid W\}] = \beta_0 + \beta_1 a + \beta_2^T W + \beta_3^T (aW), \quad a = 0, 1. \tag{3}$$

The parameter values in this model are equated to their estimates from fitting the observed-data model $\text{logit}\{\text{pr}(Y = 1 \mid A, W)\} = \beta_0 + \beta_1 A + \beta_2^T W + \beta_3^T (AW)$, to the actual trial data. This calculation yields $\pi_{\text{opt}} \approx 0.540 < 2/3$.

Next, we consider transporting the BENCHMRK trial results to another population.

The target population is defined by the SAILING trial, a randomized non-inferiority trial comparing dolutegravir, a newer HIV-1 integrase inhibitor, with raltegravir for treating HIV-1 in treatment-experienced patients (Cahn et al., 2013). Such a transportation analysis is useful for making an indirect comparison of dolutegravir with placebo (Zhang et al., 2016). This calculation assumes the same outcome model (3) as well as a log-linear model for the density ratio: $\log\{r(w)\} = \alpha_0 + \alpha_1^T w$, with parameter values estimated from a logistic regression analysis of trial membership conditional on covariates (Zhang et al., 2016). The optimal value of π is found to be $\pi_{\text{opt}}^{\text{tr}} \approx 0.481$.

A transportation analysis typically requires access to individual patient data in both trials. Without access to individual patient data from SAILING, one could use the published SAILING results (Cahn et al., 2013), together with individual patient data from BENCHMRK, to perform a post-stratification analysis that partially adjusts for population differences. For illustration, we consider post-stratification on the phenotypic sensitivity score, a known effect modifier (Cooper et al., 2008, Figure 1) which appears differentially distributed between the two trials (Zhang et al., 2016, Table 3). The weights of the strata are taken to be the observed proportions in SAILING. Under the same assumptions, the optimal value of π for post-stratification is found to be $\pi_{\text{opt}}^{\text{ps}} \approx 0.528$.

Table 3 reports an efficiency comparison of the optimal CDR design, the three optimal CIR designs described above, and the original 2:1 CIR design, for estimating ATEs in different populations. For a given estimand, the relative efficiency of each optimal design is calculated theoretically as its information bound divided by the information bound for the original design. The results in Table 3 are consistent with theoretical predictions as well as the simulation results. In particular, the optimal CDR design is most efficient in all three

settings.

To illustrate the results in Section 2.5, we also consider the log-odds-ratio (LOR) as an alternative effect measure, corresponding to $\Delta^*(g)$ with g being the logit function. Optimal CIR and CDR allocations for LOR estimation are calculated for the BENCHMRK trial as well as the transportation and post-stratification settings described earlier. Note that, with $g = \text{logit}$, each setting has a different optimal CDR allocation. Table 4 shows the π-values of the three optimal CIR designs as well as the relative efficiency results of the six optimal designs (3 CIR; 3 CDR) with the original 2:1 CIR design as the reference. As expected, the optimal CDR designs are generally more efficient than their CIR counterparts.

4 Discussion

The theoretical and numerical results presented in Sections 2 and 3 provide direct answers to the practical questions raised in Section 1 as well as new insights on the comparison of optimal CDR and CIR designs. It is well known that, for any given effect measure, the optimal CDR design is generally more efficient than the optimal CIR design because CDR includes CIR as a special case (Zhang et al., 2023). The results in the present article highlight two additional advantages of optimal CDR designs for clinical trials that may be used to estimate treatment effects in a target population that differs from the trial population. First, optimal CDR designs are generally easier to implement than their CIR counterparts, which require information about the density ratio r to be available before or during the trial. Second, optimal CDR designs are generally less sensitive than their CIR counterparts to covariate differences between the trial and target populations. Indeed, for ATE estimation,

there is a single CDR design that is optimal for any target population. For the nonlinear effect measures considered in Section 2.5, the optimal CDR design depends on the target population through (μ_0^*, μ_1^*), but the dependence is relatively simple and does not involve r, in contrast to the optimal CIR design. These observations support the use of optimal CDR designs in clinical trial practice.

While optimal CDR designs offer clear advantages in terms of statistical efficiency and invariance to population differences, their implementation may face practical, ethical, or regulatory challenges. Highly imbalanced treatment probabilities for certain covariate values may be viewed unfavorably by investigators, clinicians, or oversight committees, particularly if treatment assignment becomes strongly predictable. One possible practical solution to this problem is to truncate an optimal propensity score into the interval $[\epsilon, 1-\epsilon]$, where ϵ is a specified positive value. Additional operational constraints—such as delayed covariate ascertainment, real-time randomization infrastructure, or regulatory conservatism—may limit the extent to which optimal CDR designs can be adopted in confirmatory trials.

A key assumption in this work is Assumption 1, which is essentially a common-support assumption (i.e., $\mathcal{W} = \mathcal{W}^*$). As indicated in Remark 1, possible violations of this assumption can be fixed by restricting both F and F^* to the common support $\mathcal{W} \cap \mathcal{W}^*$, assuming that the intersection is not empty. This easy fix implies a change of estimand if the original supports are such that $\mathcal{W} \not\supseteq \mathcal{W}^*$. In this situation, if one insists on estimating a treatment effect in the original $\mathcal{W}^*$, one would have to make some additional assumptions (e.g., modeling assumptions) in order to achieve identification. Such assumptions would lead to a different statistical model, and the associated estimation and design problems would require a separate investigation.

The current literature on optimal treatment allocations largely focuses on single-time-point outcomes without censoring or repeated assessments. Many clinical trials, particularly in oncology or cardiovascular settings, involve time-to-event outcomes, recurrent events, or longitudinal endpoints. Extensions to such settings would require incorporating additional nuisance parameters (e.g., hazard functions, censoring mechanisms) and may lead to different forms of optimal allocation rules. Another topic for future research is to extend the available results for two-arm trials to multi-arm trials, platform trials, and trials with adaptive treatment strategies.

Acknowledgements

The research of Wei Zhang was supported by the National Key R&D Program of China [grant number 2022YFA1004800]. The research of Aiyi Liu was supported by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

References

Bannick, M.S., Shao, J., Liu, J., Du, Y., Yi, Y., & Ye, T. (2025). A general form of covariate adjustment in randomized clinical trials. *Biometrika*, 112, asaf029.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. & Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

Breslow, N. & Day, N. (1987) *Statistical methods in cancer research, volume II: The design and analysis of cohort studies* World Health Organization.

Cahn, P., Pozniak, A. L., Mingrone, H., Shuldyakov, A., Brites, C., Andrade-Villanueva, J. F. & et al. for the SAILING Study Team (2013). Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: Week 48 results from the randomised, double-blind, non-inferiority SAILING study. *Lancet*, 382, 700–708.

Cappiello, L., Zhang, Z., Shen, C., Butala, N.M., Cui, X. & Yeh, R.W. (2021) Adjusting for population differences using machine learning methods. *Journal of the Royal Statistical Society, Series C*, 70, 750–769.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21: C1–C68.

Cole, S.R. & Stuart, E.A. (2010) Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *American Journal of Epidemiology*, 172, 107–115.

Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G. & et al. (2024) Causal

inference methods for combining randomized trials and observational studies: A review. *Statistical Science*, 39, 165–191.

Cooper, D.A., Steigbigel, R.T., Gatell, J.M., Rockstroh, J.K., Katlama, C. & et al. for the BENCHMRK Study Teams. (2008). Subgroup and resistant analyses of Raltegravir for resistant HIV-1 infection. *New England Journal of Medicine* 359, 355–365.

Dahabreh, I.J., Robertson, S.E., Tchetgen Tchetgen, E.J., Stuart, E.A. & Hernan, M.A. (2019) Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75, 685–694.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, Springer-Verlag.

Ingall, T.J., O'Fallon, W.M., Asplund, K., Goldfrank, L.R., Hertzberg, V.S., Louis, T.A. & et al. (2004) Findings from the reanalysis of the NINDS tissue plasminogen activator for acute ischemic stroke treatment trial. *Stroke*, 35, 2418–2424.

Moore, K.L. & van der Laan, M.J. (2009) Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statistics in Medicine*, 28, 39–64.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558–625.

Nie, L., Zhang, Z., Rubin, D.B. & Chu, J. (2013) Calibration of treatment effect size through

propensity score ratio reweighting, with application to clinical trials. *Annals of Applied Statistics*, 7, 1796–1813.

NINDS rt-PA Stroke Study Group. (1995) Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine*, 333, 1581–1587.

Polley, E.C., Rose, S. & van der Laan, M.J. (2011) Super learning. In *Targeted Learning*, pages 43–66. New York, Springer.

Rosenblum, M. & van der Laan, M.J. (2010) Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *International Journal of Biostatistics*, 6, article 13.

Rudolph, K.E. & van der Laan, M.J. (2017) Robust estimation of encouragement design intervention effects transported across sites. *Journal of the Royal Statistical Society, Series B*, 79, 1509–1525.

Steigbigel, R.T., Cooper, D.A., Kumar, P.N., Eron, J.E., Schechter, M., Markowitz, M. & et al. for the BENCHMRK Study Teams (2008) Raltegravir with optimized background therapy for resistant HIV-1 infection. *New England Journal of Medicine* 359, 339–354.

Stuart, E.A., Cole, S.R., Bradshaw, C.P. & Leaf, P.J. (2011) The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society, Series A*, 174, 369–386.

Tian, L., Cai, T., Zhao, L. & Wei, L.J. (2012) On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial. *Biostatistics*, 13, 256–273.

Tsiatis, A.A. (2006) *Semiparametric Theory and Missing Data*. New York, Springer.

Tsiatis, A.A., Davidian, M., Zhang, M. & Lu, X. (2008) Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in Medicine*, 27, 4658–4677.

van der Laan, M.J. & Robins, J.M. (2003) *Unified Methods for Censored Longitudinal Data and Causality*. New York, Springer-Verlag.

van der Laan, M.J. & Rose, S. (2011) *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, Springer.

Ye, T., Shao, J., Yi, Y. & Zhao, Q. (2023) Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 118, 2370–2382.

Zhang, M., Tsiatis, A.A. & Davidian, M. (2008) Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, 64, 707–715.

Zhang, W., Zhang, Z. & Liu, A. (2023) Optimizing treatment allocation in randomized clinical trials by leveraging baseline covariates. *Biometrics*, 79, 2815–2829.

Zhang, W., Zhang, Z. & Liu, A. (2025) An adaptive design for optimizing treatment assignment in randomized clinical trials. http://arxiv.org/abs/2509.00429.

Zhang, Z. (2009) Covariate-adjusted putative placebo analysis in active-controlled clinical trials. *Statistics in Biopharmaceutical Research*, 1, 279–290.

Zhang, Z. & Ma, S. (2019) Machine learning methods for leveraging baseline covariate information to improve the efficiency of clinical trials. *Statistics in Medicine*, 38, 1703–1714.

Zhang, Z., Nie, L., Soon, G. & Hu, Z. (2016) New methods for treatment effect calibration, with applications to non-inferiority trials. *Biometrics*, 72, 20–29.

Supporting Information

The Supplementary Material includes proofs and details omitted in the main text.

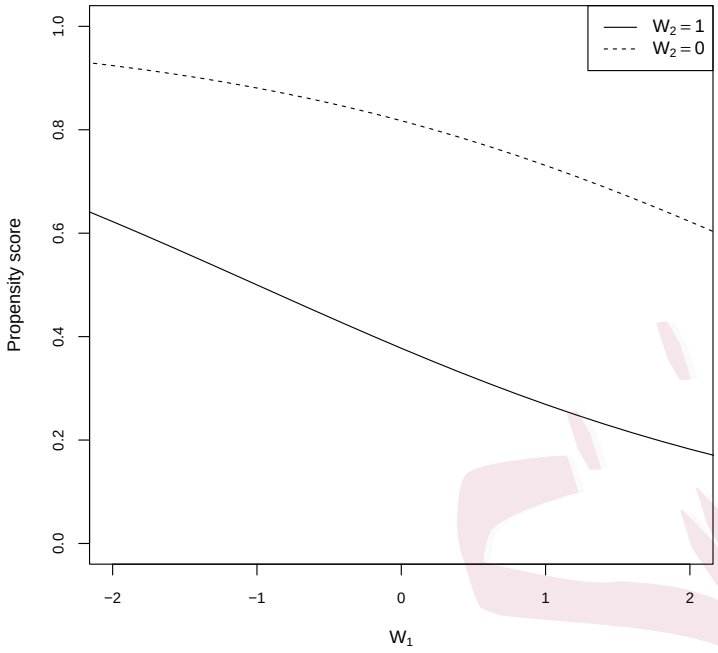

Figure 1: The optimal CDR design for ATE estimation in the simulation study with a continuous outcome.

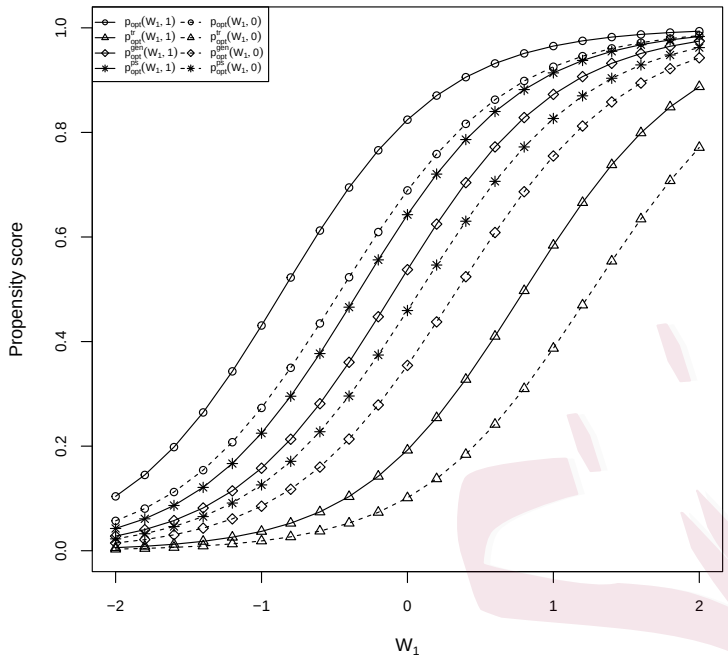

Figure 2: Optimal CDR designs for LMR estimation in the simulation study with a count outcome.

Table 1: Relative efficiency results for ATE estimation in the simulation study with a continuous outcome, with 1:1 allocation as the reference.

Allocation	Δ	Δ^*_{tr}	Δ^*_{gen}	Δ^*_{ps}
$\pi = 0.5$	1	1	1	1
$\pi_{\text{opt}} \approx 0.693$	1.131	0.657	0.784	0.706
$\pi^{\text{tr}}_{\text{opt}} \approx 0.275$	0.624	1.144	0.844	0.916
$\pi^{\text{gen}}_{\text{opt}} \approx 0.440$	0.941	1.043	1.006	1.052
$\pi^{\text{ps}}_{\text{opt}} \approx 0.389$	0.822	1.094	0.963	1.068
$p_{\text{opt}}(W)$	1.307	1.277	1.261	1.145

Δ: ATE in the trial population; Δ^*_{tr}: estimand for transportation; Δ^*_{gen}: estimand for generalization; Δ^*_{ps}: estimand for post-stratification.

Table 2: Relative efficiency results for LMR estimation in the simulation study with a count outcome, with 1:1 allocation as the reference.

Allocation	$\Delta(\log)$	$\Delta^*_{\text{tr}}(\log)$	$\Delta^*_{\text{gen}}(\log)$	$\Delta^*_{\text{ps}}(\log)$
$\pi = 0.5$	1	1	1	1
$\pi_{\text{opt}} \approx 0.684$	1.060	1.008	1.216	1.070
$\pi^{\text{tr}}_{\text{opt}} \approx 0.584$	0.993	1.070	1.104	1.018
$\pi^{\text{gen}}_{\text{opt}} \approx 0.725$	1.052	1.013	1.229	0.995
$\pi^{\text{ps}}_{\text{opt}} \approx 0.633$	1.047	1.039	1.162	1.087
$p_{\text{opt}}(W)$	1.293	0.933	1.618	1.389
$p^{\text{tr}}_{\text{opt}}(W)$	0.876	1.337	1.424	0.947
$p^{\text{gen}}_{\text{opt}}(W)$	1.172	1.199	1.675	1.352
$p^{\text{ps}}_{\text{opt}}(W)$	1.168	1.093	1.611	1.403

$\Delta(\log)$: LMR in the trial population; $\Delta^*_{\text{tr}}(\log)$: estimand for transportation; $\Delta^*_{\text{gen}}(\log)$: estimand for generalization; $\Delta^*_{\text{ps}}(\log)$: estimand for post-stratification.

Table 3: Relative efficiency results for ATE estimation in the HIV example, with the original 2:1 allocation as the reference.

Allocation	Δ	Δ^*_{tr}	Δ^*_{ps}
$\pi = 2/3$	1	1	1
$\pi_{\text{opt}} \approx 0.540$	1.071	1.139	1.084
$\pi^{\text{tr}}_{\text{opt}} \approx 0.481$	1.057	1.154	1.075
$\pi^{\text{ps}}_{\text{opt}} \approx 0.528$	1.071	1.144	1.085
$p_{\text{opt}}(W)$	1.101	1.172	1.109

Δ: ATE in the trial population; Δ^*_{tr}: estimand for transportation; Δ^*_{ps}: estimand for post-stratification.

Table 4: Relative efficiency results for LOR estimation in the HIV example, with the original 2:1 allocation as the reference.

Allocation	$\Delta(\text{logit})$	$\Delta^*_{\text{tr}}(\text{logit})$	$\Delta^*_{\text{ps}}(\text{logit})$
$\pi = 2/3$	1	1	1
$\pi_{\text{opt}} \approx 0.512$	1.104	1.098	1.064
$\pi^{\text{tr}}_{\text{opt}} \approx 0.519$	1.104	1.098	1.065
$\pi^{\text{ps}}_{\text{opt}} \approx 0.543$	1.100	1.096	1.068
$p_{\text{opt}}(W)$	1.135	1.097	1.085
$p^{\text{tr}}_{\text{opt}}(W)$	1.117	1.115	1.090
$p^{\text{ps}}_{\text{opt}}(W)$	1.127	1.113	1.092

$\Delta(\text{logit})$: LOR in the trial population; $\Delta^*_{\text{tr}}(\text{logit})$: estimand for transportation; $\Delta^*_{\text{ps}}(\text{logit})$: estimand for post-stratification.