Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Rejoinder: Learning Optimal Distributionally Robust Individualized Treatment Rules

Weibin Mo, Zhengling Qi & Yufeng Liu

To cite this article: Weibin Mo, Zhengling Qi & Yufeng Liu (2021) Rejoinder: Learning Optimal Distributionally Robust Individualized Treatment Rules, Journal of the American Statistical Association, 116:534, 699-707, DOI: 10.1080/01621459.2020.1866581

To link to this article: https://doi.org/10.1080/01621459.2020.1866581

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 534, 699–707: Theory and Methods Special Issue on Precision Medicine and Individualized Policy Discovery
https://doi.org/10.1080/01621459.2020.1866581

Rejoinder: Learning Optimal Distributionally Robust Individualized Treatment Rules

Weibin Mo*[a], Zhengling Qi*[b], and Yufeng Liu[c]

[a]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; [b]Department of Decision Sciences, George Washington University, Washington, DC; [c]Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Science, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC

We thank the opportunity offered by editors for this discussion and the discussants for their insightful comments and thoughtful contributions. We also want to congratulate Kallus (2020) for his inspiring work in improving the efficiency of policy learning by retargeting. Motivated from the discussion in Dukes and Vansteelandt (2020), we first point out interesting connections and distinctions between our work and Kallus (2020) in Section 1. In particular, the assumptions and sources of variation for consideration in these two articles lead to different research problems with different scopes and focuses. In Section 2, following the discussions in Li, Li, and Luedtke (2020); Liang and Zhao (2020), we also consider the efficient policy evaluation problem when we have some data from the testing distribution available at the training stage. We show that under the assumption that the sample sizes from training and testing are growing in the same order, efficient value function estimates can deliver competitive performance. We further show some connections of these estimates with existing literature. However, when the growth of testing sample size available for training is in a slower order, efficient value function estimates may not perform well anymore. In contrast, the requirement of the testing sample size for DRITR is not as strong as that of efficient policy evaluation using the combined data. Finally, we highlight the general applicability and usefulness of DRITR in Section 3.

1. Efficiency and Robustness

The discussion in Dukes and Vansteelandt (2020) highlighted the importance of leveraging relevant data when inferring which treatment to assign. In particular, the covariate weight functions considered in DRITR and retargeted policy learning can imply different ways of utilizing relevant data during training, and different target populations during testing. In this section, we further clarify the differences and connections between DRITR and retargeted policy learning.

DRITR and retargeted policy learning can be distinct from each other in terms of the following two main perspectives:

(I) The assumptions used in these two articles are different. If the true *conditional treatment effect (CTE)* function $C(\boldsymbol{x})$ induces a globally optimal ITR $\boldsymbol{x} \mapsto \text{sign}[C(\boldsymbol{x})]$ that belongs to the class $\mathcal{D}$, then policy learning over $\mathcal{D}$ is not sensitive to covariate reweighting/retargeting, as were discussed in Mo, Qi, and Liu (2020, Remark 1) and Kallus (2020, Lemma 2.1). In particular, Kallus (2020) referred this case as $\mathcal{D}$ being *correctly specified*, and focused on this case to obtain the *efficient* covariate weighting function. In contrast, Mo, Qi, and Liu (2020) studied the learning problem over a restricted ITR class $\mathcal{D}$ that is *misspecified* for the globally optimal ITR, and optimized the worst-case reweighted value function as a *robust* objective;

(II) The sources of variation considered in these two articles are also different. The optimality criteria for an efficient covariate weight function in Kallus (2020) focuses on reducing the *conditional-on-covariate* variance of the weighted outcome, that is, the variance explained by $(A, Y)|\boldsymbol{X}$ as in their Equation (6). In contrast, Mo, Qi, and Liu (2020) considered the robust criteria due to *covariate* variations. In the formulation of DRITR, the effect of $(A, Y)|\boldsymbol{X}$ is absorbed into the CTE function $C(\boldsymbol{X}) = \mathbb{E}[Y(1) - Y(-1)|\boldsymbol{X}]$ as a conditional mean function in $\boldsymbol{X}$. The DR-value function in Mo, Qi, and Liu (2020, eq. (4)) robustifies the underlying covariate distribution for evaluation of $C(\boldsymbol{X})$.

Two distinct assumptions on the ITR class mentioned above can explain different goals of these two. When assuming a correctly specified ITR class, Kallus (2020) can leverage the retargeting invariance property for efficiency improvement. In contrast, when allowing the misspecified ITR class, Mo, Qi, and Liu (2020) focused on the worst case of covariate changes to carry out the robust policy optimization for generalizability. The phenomenon that different model assumptions result in different goals can be remotely analogous to semiparametric inference (Robins, Rotnitzky, and van der Laan 2000). Specifically, when nuisance models are correct, the semiparametric *efficient* estimate can be obtained. When either one of but not both nuisance models are correct, a doubly *robust* estimate remains consistent. However, even for semiparametric inference, the goals of efficiency and robustness may not coexist for a specific estimate. For example, a generic construction of multiply robust estimate for factorized likelihood models is generally not semiparametric

CONTACT Yufeng Liu yfliu@email.unc.edu Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Science, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC 27599.
*Co-first authors for the article.

efficient under any model assumptions (Molina et al. 2017). It depends on the main focuses of applications to choose either a semiparametric efficient estimate or a multiply robust estimate. Analogously, the use of retargeted policy learning or DRITR also depends on the goal for *efficiency* or *robustness*, subject to the practitioners' optimistic or pessimistic beliefs on the working class of ITRs to learn from. In particular, DRITR is *robust* to potential covariate changes under the misspecified ITR class assumption.

The distinctions on source of variation mentioned in (II) characterize two different types of research questions. When questions are related to the variations of $(A, Y)|\boldsymbol{X}$, such as limited overlap and heteroscedasticity, retargeting weights in Kallus (2020) can provide an optimal way to control conditional-on-covariate variances (Crump et al. 2006, 2009). However, such optimal weights cannot control the variances of covariates themselves. Therefore, retargeted policy learning may generally work well for the case that conditional-on-covariate variances are the estimation bottleneck, while covariate variations can be ignored. Such an example can be found in Athey and Wager (2020, sec. 4). Besides ignoring variances from covariates, the violation of retargeting invariance, that is, misspecifying the ITR class, can also contribute biases due to reweighting. Kallus (2020, sec. 5.2) also pointed out that the levels of ITR class misspecification and covariate changes need to be assumed mild when applying retargeting policy learning.

In contrast to the scope of retargeted policy learning, DRITR has an explicit focus on covariate changes. This was motivated from the fact that the challenges of generalizing causal estimands are mainly due to covariate changes, and reweighting can correctly target the testing population of interest (Stuart et al. 2011). Without prior information on the testing population at the training stage, DRITR took the worst-case reweighting scheme to guarantee robust performance. A similar strategy was also leveraged by Zhao et al. (2019). However, given that reweighting in this case mainly aims for covariate-change correction, DRITR is not intended for handling variations from treatments and outcomes. Therefore, our formulation utilizes a nonparametric estimate of the CTE function to remove the variations from treatments and outcomes.

One potential research question is whether robust reweighting in DRITR can also handle the limited overlap and heteroscedasticity problems considered in retargeted policy learning. Unfortunately, we suspect that robust reweighting and efficient retargeting may have opposite effects. In particular, we notice that the optimal retargeting weight function from Kallus (2020) is inverse-proportionate to $\sum_a \frac{\sigma^2(\boldsymbol{x},a)}{\pi_A(a|\boldsymbol{x})}$, where $\pi_A(a|\boldsymbol{x})$ is the propensity score function. In contrast, for robust reweighting, Qi et al. (2019) studied a modified version of DRITR that focused on the dual formulation. Although they did not explicitly link to the robust reweighting, the resulting robust weight function in Qi et al. (2019) is increasing in the variance function $\sigma^2(\boldsymbol{x}, a) := \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = a)$. Consequently, the robust weight function may not be compatible with the retargeting weight function. This suggests that DRITR and retargeted policy learning may need to be utilized in different scenarios. It may be interesting to study a combined version of retargeted policy learning and DRITR that can enjoy both efficiency and robustness.

To conclude this section, we make a comparison between DRITR and retargeted policy learning in Table 1.

2. Efficient Policy Evaluation Under Specific Covariate Changes

DRITR aims for performance guarantee in presence of general covariate changes. It assumes no access to any information from the testing distribution at the training stage. When a small set of calibrating data is available from the testing distribution, such information is only used for choosing a DR-constant to determine the final DRITR. However, the problem of combining the training and calibrating data during training as discussed in Li, Li, and Luedtke (2020) and Liang and Zhao (2020) is also worthwhile to study. In this section, we focus on efficient policy evaluation with training and calibrating data from a specific testing distribution. It should be highlighted that the true covariate density ratio of testing with respect to training is not readily available, and can only be inferred from the observed training and calibrating data.

Consider two possible types of pooled datasets $\mathcal{O}^{(1)} = \{\boldsymbol{O}_i = (\boldsymbol{X}_i, A_i, Y_i, S_i)\}_{i=1}^n$ and $\mathcal{O}^{(2)} = \{\boldsymbol{O}_i = (\boldsymbol{X}_i, S_iA_i, S_iY_i, S_i)\}_{i=1}^n$, where $S_i = 1$ indicates that $\boldsymbol{O}_i|(S_i = 1) \sim \mathbb{P}_{\text{train}}$, and the ith data point belongs to the training data; $S_i = 0$ indicates that $\boldsymbol{O}_i|(S_i = 0) \sim \mathbb{P}_{\text{test}}$, and the ith data point belongs to the calibrating data. For the Type-1 dataset $\mathcal{O}^{(1)}$, we observe covariates, treatment assignments and outcomes in both training and calibrating data. For the Type-2 dataset $\mathcal{O}^{(2)}$, treatment assignments and outcomes $\{(A_i, Y_i) : S_i = 0\}$ in calibrating data are missing. Let $Y_i(1)$ and $Y_i(-1)$ be the potential outcomes, and denote $Y_i(d) := \sum_{a\in\{1,-1\}} Y_i(a)\mathbb{1}[d(\boldsymbol{X}_i) = a]$. For a fixed ITR $d : \mathcal{X} \to \{1, -1\}$, the goal for policy evaluation is to estimate the following values of d under the specific testing distribution:

$$\theta := \mathcal{V}_{\text{test}}(d) = \mathbb{E}_{\text{test}}[Y_i(d)];$$
$$\theta_1 := \mathcal{V}_{1,\text{test}}(d) = \mathbb{E}_{\text{test}}[Y_i(d) - Y_i(-d)]. \quad (1)$$

To identify the potential outcomes from observed outcomes, we make the following assumptions (Rubin 1974).

Table 1. Comparison of DRITR (Mo, Qi, and Liu 2020) and Retargeted policy learning (Kallus 2020).

	DRITR	Retargeted policy learning
Assumption	$\{\boldsymbol{x} \mapsto \text{sign}[C(\boldsymbol{x})]\} \nsubseteq \mathcal{D}$	$\{\boldsymbol{x} \mapsto \text{sign}[C(\boldsymbol{x})]\} \subseteq \mathcal{D}$
Source of variation	Covariate $\boldsymbol{X}$	Conditional-on-covariate $(A, Y)|\boldsymbol{X}$
Weight dependency	CTE function $C(\boldsymbol{x})$; underlying ITR $d(\boldsymbol{x})$	Propensity score function $\pi_A(a|\boldsymbol{x})$; variance function $\sigma^2(\boldsymbol{x}, a)$
Optimality	Maximizing worst-case value	Minimizing conditional-on-covariate variance
Main applications	Covariate changes	Limited overlap and heteroscedasticity

Assumption 1 (Consistency). $Y_i = Y_i(A_i) = \sum_{a\in\{1,-1\}} Y_i(a)\mathbb{1}(A_i = a)$.

Assumption 2 (Exchangeability over treatment). For $\boldsymbol{x} \in \mathcal{X}$, $s \in \{1, 0\}$ and $a \in \{1, -1\}$, we have $Y_i(a) \perp\!\!\!\perp A_i|(\boldsymbol{X}_i = \boldsymbol{x}, S_i = s)$.

Assumption 3 (Positivity of treatment probability). For $\boldsymbol{x} \in \mathcal{X}$, $s \in \{1, 0\}$ and $a \in \{1, -1\}$, we have $\pi_A(\boldsymbol{a}|\boldsymbol{x}, s) := \mathbb{P}(A_i = a|\boldsymbol{X}_i = \boldsymbol{x}, S_i = s) \geq \tau > 0$.

Notice that the policy evaluation problem can be cast as a parameter estimation problem. The key challenge here is that the estimand θ or θ_1 is evaluated under the target population $\mathbb{P}_{\text{test}}$ that can be different from the distribution of the observed data $\mathcal{O}^{(1)}$ or $\mathcal{O}^{(2)}$. To identify θ and θ_1 from the pooled data $\mathcal{O}^{(1)}$ or $\mathcal{O}^{(2)}$, we consider the following *mean exchangeability assumption* used in Pearl and Bareinboim (2014) to transport information from $\mathbb{P}_{\text{train}}$ to $\mathbb{P}_{\text{test}}$.

Assumption 4 (Mean exchangeability over selection). For $\boldsymbol{x} \in \mathcal{X}$, $a \in \{1, -1\}$ and $s \in \{1, 0\}$, we have $\mathbb{E}[Y_i(a)|\boldsymbol{X}_i = \boldsymbol{x}, S_i = s] = \mathbb{E}[Y_i(a)|\boldsymbol{X}_i = \boldsymbol{x}] := Q(\boldsymbol{x}, a)$.

Assumption 4 implies that $\mathbb{E}[Y_i(1)-Y_i(-1)|\boldsymbol{X}_i, S_i] = \mathbb{E}[Y_i(1)-Y_i(-1)|\boldsymbol{X}_i] = Q(\boldsymbol{X}_i, 1) - Q(\boldsymbol{X}_i, -1) := C(\boldsymbol{X}_i)$, which was the required condition in Mo, Qi, and Liu (2020, Remark 2). Covariate changes in Mo, Qi, and Liu (2020, Assumption 1) is sufficient for Assumption 4. To identify θ and θ_1 from the training and pooled data, we further consider the *strong ignorability assumption* on the conditional-on-covariate selection probability function.

Assumption 5 (Positivity of selection probability). For $\boldsymbol{x} \in \mathcal{X}$ and $s \in \{1, 0\}$, we have $\pi_S(s|\boldsymbol{x}) := \mathbb{P}(S_i = s|\boldsymbol{X}_i = \boldsymbol{x}) \geq \delta > 0$.

Assumption 5 also implies the positivity of the marginal selection probability: $\rho_S := \mathbb{P}(S_i = 1) \geq \delta$, and $1 - \rho_S = \mathbb{P}(S_i = 0) \geq \delta$. Under Assumption 5, we can express $\mathbb{E}_{\text{test}}(\cdot) = \mathbb{E}(\cdot|S = 0)$ under $\mathbb{E}_{\text{train}}(\cdot) = \mathbb{E}(\cdot|S = 1)$ by reweighting, which is given in the following Lemma 1.

Lemma 1. Consider the data $(\boldsymbol{X}, S) \in \mathcal{X} \times \{1, 0\}$ satisfying $\mathbb{P}(S = s|\boldsymbol{X}) \geq \delta > 0$ for $s \in \{1, 0\}$. Then for any $g : \mathcal{X} \to \mathbb{R}$ such that $\mathbb{E}[|g(\boldsymbol{X})|] < +\infty$, we have

$$\mathbb{E}[g(\boldsymbol{X})|S = 0] = \mathbb{E}[w(\boldsymbol{X})g(\boldsymbol{X})|S = 1],$$

where

$$w(\boldsymbol{X}) = \frac{\mathbb{P}(S = 1)\mathbb{P}(S = 0|\boldsymbol{X})}{\mathbb{P}(S = 0)\mathbb{P}(S = 1|\boldsymbol{X})}.$$

In particular, $w(\boldsymbol{X}) \geq 0$ and $\mathbb{E}[w(\boldsymbol{X})|S = 1] = 1$.

Lemma 1 can be regarded as a parallel version of the weighted representation in Mo, Qi, and Liu (2020, Assumption 1), and further suggests the form of the theoretical weighting function $w(\boldsymbol{x}) = \frac{\rho_S}{1-\rho_S}\frac{\pi_S(0|\boldsymbol{x})}{\pi_S(1|\boldsymbol{x})}$ for $\boldsymbol{x} \in \mathcal{X}$.

2.1. Semiparametric Inference

To consider estimates for θ and θ_1, we first study the semiparametric inference properties for θ and θ_1 based on two types of data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, respectively. In Proposition 1, we establish the identification of θ and θ_1 from the observed data. Then we derive the efficient influence functions for θ and θ_1 in Theorem 1, which is consistent with results from Rudolph and van der Laan (2017), Dahabreh et al. (2019), and Uehara, Kato, and Yasui (2020). For $\boldsymbol{x} \in \mathcal{X}$, we denote $Q(\boldsymbol{x}, d) := \sum_{a\in\{1,-1\}} Q(\boldsymbol{x}, a)\mathbb{1}[d(\boldsymbol{x}) = a]$.

Proposition 1. Consider the observed data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, and the parameters θ and θ_1 in (1). Under Assumptions 1–5, we have:

(I)

$$\begin{aligned}\theta &= \mathbb{E}\left\{\frac{\mathbb{1}(S_i = 0)}{1-\rho_S}Q(\boldsymbol{X}_i, d)\right\}\\ &= \mathbb{E}\Big\{[w(\boldsymbol{X}_i)\mathbb{1}(S_i = 1) + \mathbb{1}(S_i = 0)]Q(\boldsymbol{X}_i, d)\Big\} \quad (\text{on } \mathcal{O}^{(1)})\\ &= \mathbb{E}\left\{\frac{w(\boldsymbol{X}_i)\mathbb{1}(S_i = 1)}{\rho_S}Q(\boldsymbol{X}_i, d)\right\} \quad (\text{on } \mathcal{O}^{(2)}).\end{aligned}$$

The expressions for θ_1 can be obtained by replacing $Q(\boldsymbol{X}_i, d)$ by $C(\boldsymbol{X}_i)d(\boldsymbol{X}_i)$;

(II) For $\boldsymbol{x} \in \mathcal{X}$,

$$\begin{aligned}Q(\boldsymbol{X}_i, d) &= \mathbb{E}\left\{\left(\frac{\mathbb{1}(S_i = 1)}{\pi_A(A_i|\boldsymbol{X}_i, 1)} + \frac{\mathbb{1}(S_i = 0)}{\pi_A(A_i|\boldsymbol{X}_i, 0)}\right)\mathbb{1}[d(\boldsymbol{X}_i) = A_i]Y_i\middle|\boldsymbol{X}_i\right\}\\ &\quad (\text{on } \mathcal{O}^{(1)})\\ &= \mathbb{E}\left\{\frac{1}{\pi_A(A_i|\boldsymbol{X}_i, 1)}\mathbb{1}[d(\boldsymbol{X}_i) = A_i]Y_i\middle|\boldsymbol{X}_i, S_i = 1\right\}\\ &\quad (\text{on } \mathcal{O}^{(2)}).\end{aligned}$$

The expressions for θ_1 can be obtained by replacing $\mathbb{1}[d(\boldsymbol{X}_i) = A_i]$ by $d(\boldsymbol{X}_i)A_i$.

Theorem 1 (Efficient influence functions). Consider the observed data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, and the parameters θ and θ_1 in (1). Under Assumptions 1–5 and some regularity conditions, the corresponding semiparametric *efficient influence functions (EIFs)* are:

$$\begin{aligned}\text{EIF}^{(1)}(\theta) &= \left(\tfrac{w(\boldsymbol{X})\mathbb{1}(S=1)}{\pi_A(A|\boldsymbol{X},1)} + \tfrac{\mathbb{1}(S=0)}{\pi_A(A|\boldsymbol{X},0)}\right)\\ &\quad\times \mathbb{1}[d(\boldsymbol{X}) = A][Y - Q(\boldsymbol{X}, A)]\\ &\quad+ \tfrac{\mathbb{1}(S=0)}{1-\rho_S}[Q(\boldsymbol{X}, d) - \theta];\\ \text{EIF}^{(2)}(\theta) &= \tfrac{w(\boldsymbol{X})\mathbb{1}(S=1)}{\rho_S\pi_A(A|\boldsymbol{X},1)}\\ &\quad\times \mathbb{1}[d(\boldsymbol{X}) = A][Y - Q(\boldsymbol{X}, A)]\\ &\quad+ \tfrac{\mathbb{1}(S=0)}{1-\rho_S}[Q(\boldsymbol{X}, d) - \theta]; \text{EIF}^{(1)}(\theta_1)\\ &= \left(\tfrac{w(\boldsymbol{X})\mathbb{1}(S=1)}{\pi_A(A|\boldsymbol{X},1)} + \tfrac{\mathbb{1}(S=0)}{\pi_A(A|\boldsymbol{X},0)}\right)\\ &\quad\times d(\boldsymbol{X})A[Y - Q(\boldsymbol{X}, A)]\\ &\quad+ \tfrac{\mathbb{1}(S=0)}{1-\rho_S}[C(\boldsymbol{X})d(\boldsymbol{X}) - \theta_1];\\ \text{EIF}^{(2)}(\theta_1) &= \tfrac{w(\boldsymbol{X})\mathbb{1}(S=1)}{\rho_S\pi_A(A|\boldsymbol{X},1)}\\ &\quad\times d(\boldsymbol{X})A[Y - Q(\boldsymbol{X}, A)]\\ &\quad+ \tfrac{\mathbb{1}(S=0)}{1-\rho_S}[C(\boldsymbol{X})d(\boldsymbol{X}) - \theta_1].\end{aligned}$$

Here, $\text{EIF}^{(k)}$ represents the EIF based on $\mathcal{O}^{(k)}$ for $k = 1, 2$.

Theorem 1 can be connected to existing literature. Rudolph and van der Laan (2017, sec. 4) obtained the EIF for θ_1 on the Type-2 data $\mathcal{O}^{(2)}$. Dahabreh et al. (2019, Appendix D) first obtained the EIF for θ based on the Type-2 data $\mathcal{O}^{(2)}$. Then they further argued that if Assumption 2 is relaxed to that $Y_i(a) \perp\!\!\!\perp A_i | (\boldsymbol{X}_i = \boldsymbol{x}, S_i = 1)$ holds on the training data only, then the observed treatment assignments and outcomes $\{(A_i, Y_i) : S_i = 0\}$ from the calibrating data cannot be used. In that case, the EIFs based on $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$ are the same. Uehara, Kato, and Yasui (2020, Theorem 11) obtained the EIF for θ on the Type-2 data $\mathcal{O}^{(2)}$ using a stratified sampling formulation for $\boldsymbol{O}_i | (S_i = 1)$ and $\boldsymbol{O}_i | (S_i = 0)$, and treating $w(\boldsymbol{X}_i)$ as a general density ratio function of covariates. Finally, we would like to point out that Kallus (2020, Lemma 5.1) considered the EIF for a fixed weight function, while the weight function $w(\boldsymbol{x})$ as in Lemma 1 is model-endogenous in the sense that it corresponds to the conditional-on-covariate selection odds or the covariate density ratio in the semiparametric model. The EIF in Kallus (2020, Lemma 5.1) assumes that $w(\boldsymbol{x})$ is known or the associated nuisance function $\pi_S(s|\boldsymbol{x}, a)$ is known, while the EIFs from our Theorem 1 account for additional variances contributed by estimating $w(\boldsymbol{x})$.

Suppose $w(\boldsymbol{x})$, $\pi_A(a|\boldsymbol{x}, s)$ and $Q(\boldsymbol{x}, a)$ are known as oracle. Recall that $C(\boldsymbol{x}) = Q(\boldsymbol{x}, 1) - Q(\boldsymbol{x}, -1)$. Denote $n_s := \#\{i : S_i = s\}$ for $s \in \{1, 0\}$. Then we have $n = n_1 + n_0$ and $n_1/n \to \rho_S$ as $n \to \infty$. Theorem 1 can imply the following oracle semiparametric efficient estimates of θ and θ_1 based on $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, respectively:

$$
\begin{aligned}
\widehat{\theta}^{(1)}_{\text{eff}} = {} & \frac{1}{n_1} \sum_{i:S_i=1} \frac{n_1}{n} w(\boldsymbol{X}_i) \frac{\mathbb{1}[d(\boldsymbol{X}_i) = A_i]}{\pi_A(A_i|\boldsymbol{X}_i, 1)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \\
& (:= \phi^{(1)}_{n_1}) \\
& + \frac{1}{n_0} \sum_{i:S_i=0} \left\{ \frac{n_0}{n} \frac{\mathbb{1}[d(\boldsymbol{X}_i) = A_i]}{\pi_A(A_i|\boldsymbol{X}_i, 0)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \right. \\
& \left. + Q(\boldsymbol{X}_i, d) \right\} \\
& (:= \psi^{(1)}_{n_0} + \theta); \\
\widehat{\theta}^{(2)}_{\text{eff}} = {} & \frac{1}{n_1} \sum_{i:S_i=1} w(\boldsymbol{X}_i) \frac{\mathbb{1}[d(\boldsymbol{X}_i) = A_i]}{\pi_A(A|\boldsymbol{X}, 1)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \\
& (:= \phi^{(2)}_{n_1}) \\
& + \frac{1}{n_0} \sum_{i:S_i=0} Q(\boldsymbol{X}_i, d) \\
& (:= \psi^{(2)}_{n_0} + \theta); \\
\widehat{\theta}^{(1)}_{1,\text{eff}} = {} & \frac{1}{n_1} \sum_{i:S_i=1} \frac{n_1}{n} w(\boldsymbol{X}_i) \frac{d(\boldsymbol{X}_i) A_i}{\pi_A(A_i|\boldsymbol{X}_i, 1)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \\
& (:= \phi^{(1)}_{1,n_1}) \\
& + \frac{1}{n_0} \sum_{i:S_i=0} \left\{ \frac{n_0}{n} \frac{\mathbb{1}[d(\boldsymbol{X}_i) = A_i]}{\pi_A(A_i|\boldsymbol{X}_i, 0)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \right. \\
& \left. + C(\boldsymbol{X}_i) d(\boldsymbol{X}_i) \right\} \\
& (:= \psi^{(1)}_{1,n_0} + \theta_1); \\
\widehat{\theta}^{(2)}_{1,\text{eff}} = {} & \frac{1}{n_1} \sum_{i:S_i=1} w(\boldsymbol{X}_i) \frac{d(\boldsymbol{X}_i) A_i}{\pi_A(A_i|\boldsymbol{X}_i, 1)} [Y_i - Q(\boldsymbol{X}_i, A_i)] \\
& (:= \phi^{(2)}_{1,n_1}) \\
& + \frac{1}{n_0} \sum_{i:S_i=0} C(\boldsymbol{X}_i) d(\boldsymbol{X}_i) \\
& (:= \psi^{(2)}_{1,n_0} + \theta_1). \qquad (2)
\end{aligned}
$$

Here, we use n_1/n and n_0/n to replace ρ_S and $1 - \rho_S$, respectively. The asymptotic variances of the oracle semiparametric estimates in (2) are the smallest among all *regular and asymptotic linear (RAL)* estimates (Tsiatis 2007), which we establish in the following Theorem 2. For $\boldsymbol{x} \in \mathcal{X}$, $s \in \{1, 0\}$ and $a \in \{1, -1\}$, we denote $\sigma^2(\boldsymbol{x}, s, a) := \mathsf{var}[Y_i(a) | \boldsymbol{X}_i = \boldsymbol{x}, S_i = s]$, $\sigma^2(\boldsymbol{x}, s, d) := \sum_{a \in \{1,-1\}} \sigma^2(\boldsymbol{x}, s, a) \mathbb{1}[d(\boldsymbol{x}_i) = a]$ and $\pi_A(d|\boldsymbol{x}, s) := \sum_{a \in \{1,-1\}} \pi_A(a|\boldsymbol{x}, s)$. To obtain asymptotic results for all these estimates, we make the following integrability assumption.

Assumption 6 (Squared integrability). Assume that

$$
\mathbb{E}[Q(\boldsymbol{X}_i, 1)^2], \mathbb{E}[Q(\boldsymbol{X}_i, -1)^2] < +\infty;
$$

$$
\sup_{\boldsymbol{x} \in \mathcal{X}, s \in \{1,0\}, a \in \{1,-1\}} \sigma^2(\boldsymbol{x}, s, a) < +\infty.
$$

Theorem 2 (Semiparametric efficiency). Consider the observed data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, the parameters θ and θ_1 in (1), and the corresponding oracle efficient estimates in (2). Under Assumptions 1–6, we have

$$
\begin{aligned}
& \sqrt{n_1} \phi^{(1)}_{n_1} \overset{n_1 \to \infty}{\Longrightarrow} Z^{(1)}_1; \quad \sqrt{n_0} \psi^{(1)}_{n_0} \overset{n_0 \to \infty}{\Longrightarrow} Z^{(1)}_0; \\
& Z^{(1)}_1 \sim \mathcal{N}(0, \nu^{(1)}_{\text{eff}}); \quad Z^{(1)}_0 \sim \mathcal{N}(0, \zeta^{(1)}_{\text{eff}}); \quad Z^{(1)}_1 \perp\!\!\!\perp Z^{(1)}_0; \\
& \sqrt{n_1} \phi^{(2)}_{n_1} \overset{n_1 \to \infty}{\Longrightarrow} Z^{(2)}_1; \quad \sqrt{n_0} \psi^{(2)}_{n_0} \overset{n_0 \to \infty}{\Longrightarrow} Z^{(2)}_0; \\
& Z^{(2)}_1 \sim \mathcal{N}(0, \nu^{(2)}_{\text{eff}}); \quad Z^{(2)}_0 \sim \mathcal{N}(0, \zeta^{(2)}_{\text{eff}}); \quad Z^{(2)}_1 \perp\!\!\!\perp Z^{(2)}_0; \\
& \sqrt{n_1} \phi^{(1)}_{1,n_1} \overset{n_1 \to \infty}{\Longrightarrow} Z^{(1)}_{11}; \quad \sqrt{n_0} \psi^{(1)}_{1,n_0} \overset{n_0 \to \infty}{\Longrightarrow} Z^{(1)}_{10}; \\
& Z^{(1)}_{11} \sim \mathcal{N}(0, \nu^{(1)}_{1,\text{eff}}); \quad Z^{(1)}_{10} \sim \mathcal{N}(0, \zeta^{(1)}_{1,\text{eff}}); \quad Z^{(1)}_{11} \perp\!\!\!\perp Z^{(1)}_{10}; \\
& \sqrt{n_1} \phi^{(2)}_{1,n_1} \overset{n_1 \to \infty}{\Longrightarrow} Z^{(2)}_{11}; \quad \sqrt{n_0} \psi^{(2)}_{1,n_0} \overset{n_0 \to \infty}{\Longrightarrow} Z^{(2)}_{10}; \\
& Z^{(2)}_{11} \sim \mathcal{N}(0, \nu^{(2)}_{1,\text{eff}}); \quad Z^{(2)}_{10} \sim \mathcal{N}(0, \zeta^{(2)}_{1,\text{eff}}); \quad Z^{(2)}_{11} \perp\!\!\!\perp Z^{(2)}_{10},
\end{aligned}
$$

where

$$\begin{aligned}
\nu_{\text{eff}}^{(1)} &= \rho_S^2 \mathbb{E}\left\{ w(\boldsymbol{X}_i)^2 \frac{\sigma^2(\boldsymbol{X}_i, 1, d)}{\pi_A(d|\boldsymbol{X}_i, 1)} \middle| S_i = 1 \right\}; \\
\zeta_{\text{eff}}^{(1)} &= (1-\rho_S)^2 \mathbb{E}\left\{ \frac{\sigma^2(\boldsymbol{X}_i, 0, d)}{\pi_A(d|\boldsymbol{X}_i, 0)} \middle| S_i = 0 \right\} \\
&\quad + \mathsf{var}[Q(\boldsymbol{X}_i, d)|S_i = 0]; \\
\nu_{\text{eff}}^{(2)} &= \mathbb{E}\left\{ w(\boldsymbol{X}_i)^2 \frac{\sigma^2(\boldsymbol{X}_i, 1, d)}{\pi_A(d|\boldsymbol{X}_i, 1)} \middle| S_i = 1 \right\}; \\
\zeta_{\text{eff}}^{(2)} &= \mathsf{var}[Q(\boldsymbol{X}_i, d)|S_i = 0]; \\
\nu_{1,\text{eff}}^{(1)} &= \rho_S^2 \mathbb{E}\left\{ w(\boldsymbol{X}_i)^2 \sum_{a\in\{1,-1\}} \frac{\sigma^2(\boldsymbol{X}_i, 1, a)}{\pi_A(a|\boldsymbol{X}_i, 1)} \middle| S_i = 1 \right\}; \\
\zeta_{1,\text{eff}}^{(1)} &= (1-\rho_S)^2 \mathbb{E}\left\{ \sum_{a\in\{1,-1\}} \frac{\sigma^2(\boldsymbol{X}_i, 0, a)}{\pi_A(a|\boldsymbol{X}_i, 0)} \middle| S_i = 0 \right\} \\
&\quad + \mathsf{var}[C(\boldsymbol{X}_i) d(\boldsymbol{X}_i)|S_i = 0]; \\
\nu_{1,\text{eff}}^{(2)} &= \mathbb{E}\left\{ w(\boldsymbol{X}_i)^2 \sum_{a\in\{1,-1\}} \frac{\sigma^2(\boldsymbol{X}_i, 1, a)}{\pi_A(a|\boldsymbol{X}_i, 1)} \middle| S_i = 1 \right\}; \\
\zeta_{1,\text{eff}}^{(2)} &= \mathsf{var}[C(\boldsymbol{X}_i) d(\boldsymbol{X}_i)|S_i = 0].
\end{aligned}$$

Moreover, for $\{\alpha_n\}$ such that $\alpha_n/\sqrt{n_1} \to \gamma_1$ and $\alpha_n/\sqrt{n_0} \to \gamma_0$ as $n \to \infty$, we have

$$\begin{aligned}
\alpha_n(\widehat{\theta}_{\text{eff}}^{(1)} - \theta) &= (\gamma_1 + \mathcal{O}(1)) \times \sqrt{n_1}\phi_{n_1}^{(1)} + (\gamma_0 + \mathcal{O}(1)) \\
&\quad \times \sqrt{n_0}\psi_{n_0}^{(1)} \overset{n_1, n_0 \to \infty}{\Longrightarrow} \gamma_1 Z_1^{(1)} + \gamma_0 Z_0^{(1)}; \\
\alpha_n(\widehat{\theta}_{\text{eff}}^{(2)} - \theta) &= (\gamma_1 + \mathcal{O}(1)) \times \sqrt{n_1}\phi_{n_1}^{(2)} + (\gamma_0 + \mathcal{O}(1)) \\
&\quad \times \sqrt{n_0}\psi_{n_0}^{(2)} \overset{n_1, n_0 \to \infty}{\Longrightarrow} \gamma_1 Z_1^{(2)} + \gamma_0 Z_0^{(2)}; \\
\alpha_n(\widehat{\theta}_{1,\text{eff}}^{(1)} - \theta_1) &= (\gamma_1 + \mathcal{O}(1)) \times \sqrt{n_1}\phi_{1,n_1}^{(1)} + (\gamma_0 + \mathcal{O}(1)) \\
&\quad \times \sqrt{n_0}\psi_{1,n_0}^{(1)} \overset{n_1, n_0 \to \infty}{\Longrightarrow} \gamma_1 Z_{11}^{(1)} + \gamma_0 Z_{10}^{(1)}; \\
\alpha_n(\widehat{\theta}_{1,\text{eff}}^{(1)} - \theta_1) &= (\gamma_1 + \mathcal{O}(1)) \times \sqrt{n_1}\phi_{1,n_1}^{(2)} + (\gamma_0 + \mathcal{O}(1)) \\
&\quad \times \sqrt{n_0}\psi_{1,n_0}^{(2)} \overset{n_1, n_0 \to \infty}{\Longrightarrow} \gamma_1 Z_{11}^{(2)} + \gamma_0 Z_{10}^{(2)}.
\end{aligned}$$

In particular, for $\alpha_n = \sqrt{n}$ (resp. $\sqrt{n_1}$ or $\sqrt{n_0}$), the corresponding $\sqrt{n}$ (resp. $\sqrt{n_1}$ or $\sqrt{n_0}$) asymptotic variances achieve the semiparametric $\sqrt{n}$-(resp. $\sqrt{n_1}$- or $\sqrt{n_0}$-)variance lower bounds for θ and θ_1 on $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, respectively.

We notice that the efficient estimates (2) can depend on the unknown nuisance functions $w(\boldsymbol{x})$, $\pi_A(a|\boldsymbol{x}, s)$, and $Q(\boldsymbol{x}, a)$. Implementable efficient estimates of θ and θ_1 are the plug-in versions with the corresponding sample-dependent nuisance function estimates. Uehara, Kato, and Yasui (2020) took the following cross-fitting strategy (Chernozhukov et al. 2018) when plugging in the nuisance function estimates: the pooled data are stratified into $\{i : S_i = 1\}$ and $\{i : S_i = 0\}$, and the sample points within each stratum are randomly divided into K bags. For $k \in \{1, 2, \ldots, K\}$, we obtain out-of-bag estimates of the nuisance functions from the pooled dataset that rules out the kth bag of data points. When constructing the cross-fitting estimates (2) of θ and θ_1, if the ith data point belongs to the kth bag, then it utilizes the out-of-kth-bag nuisance function estimates. Using the cross-fitting strategy, we can follow Uehara, Kato, and Yasui (2020, Theorem 2) and establish $\sqrt{n}$-equivalences for plug-in efficient estimates. In Theorem 3, we only consider the cross-fitting estimates for θ, and the same argument can be applied to θ_1.

Theorem 3 ($\sqrt{n}$-Equivalence). Consider the observed data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$ and the parameter θ in (1). Denote $\Phi^{(1)}(\eta) := \widehat{\theta}_{\text{eff}}^{(1)} - \theta$ and $\Phi^{(2)}(\eta) = \widehat{\theta}_{\text{eff}}^{(2)} - \theta$ from (2) with $\eta = \big(w(\boldsymbol{x}), \pi_A(a|\boldsymbol{x}, s), Q(\boldsymbol{x}, a)\big)$ as the nuisance functions. For $k \in \{1, 2, \ldots, K\}$, let $\widehat{\eta}_{\text{cross}} := \{(\widehat{w}^{(k)}, \widehat{\pi}_A^{(k)}, \widehat{Q}^{(k)})\}_{k=1}^K$ be the out-of-bag nuisance function estimates, and $\Phi^{(1)}(\widehat{\eta}_{\text{cross}})$ and $\Phi^{(2)}(\widehat{\eta}_{\text{cross}})$ be the corresponding cross-fitting versions. Define

$$\begin{aligned}
\alpha_n^{(2)} &:= \max_{a\in\{1,-1\}} \max_{1\le k\le K} \left\| \frac{\widehat{w}^{(k)}(\cdot)}{\widehat{\pi}_A^{(k)}(a|\cdot, 1)} - \frac{w(\cdot)}{\pi_A(a|\cdot, 1)} \right\|_{L^2(\mathbb{P})}; \\
\alpha_n^{(1)} &:= \alpha_n^{(2)} + \max_{1\le k\le K} \left\| \widehat{\pi}_A^{(k)}(1|\cdot, 0) - \pi_A(1|\cdot, 0) \right\|_{L^2(\mathbb{P})}; \\
\beta_n &:= \max_{a\in\{1,-1\}} \max_{1\le k\le K} \left\| \widehat{Q}^{(k)}(\cdot, a) - Q(\cdot, a) \right\|_{L^2(\mathbb{P})},
\end{aligned}$$

where $\|\cdot\|_{L^2(\mathbb{P})}$ is the $L^2(\mathbb{P})$-norm with respect to the covariate vector $\boldsymbol{X}$. Then under Assumptions 1–6 and that $\alpha_n^{(1)}, \alpha_n^{(2)}, \beta_n = \mathcal{O}_{\mathbb{P}}(1)$, we have

$$\begin{aligned}
\sqrt{n}[\Phi^{(1)}(\widehat{\eta}_{\text{cross}}) - \Phi^{(1)}(\eta)] &= \mathcal{O}_{\mathbb{P}}(\sqrt{n}\alpha_n^{(1)}\beta_n); \\
\sqrt{n}[\Phi^{(2)}(\widehat{\eta}_{\text{cross}}) - \Phi^{(2)}(\eta)] &= \mathcal{O}_{\mathbb{P}}(\sqrt{n}\alpha_n^{(2)}\beta_n).
\end{aligned}$$

Remark 1. Notice that in Theorem 3, by the fact that $\pi_A(a|\boldsymbol{x}, s) \ge \tau$, we further have

$$\begin{aligned}
\alpha_n^{(2)} &\le \text{constant} \times \max_{1\le k\le K} \left\| \widehat{w}^{(k)} - w \right\|_{L^2(\mathbb{P})} \\
&\quad + \text{constant} \times \max_{1\le k\le K} \left\| \widehat{\pi}_A^{(k)}(1|\cdot, 1) - \pi_A(1|\cdot, 1) \right\|_{L^2(\mathbb{P})}.
\end{aligned}$$

Then a dominating rate of $\alpha_n^{(2)}$ can be chosen as the slower one of $\widehat{w}(\cdot)$ and $\widehat{\pi}_A(1|\cdot, 1)$.

One important implication of the remainder terms in Theorem 3 is that $\alpha_n^{(1)}$, $\alpha_n^{(2)}$ and β_n can be in slower orders than the usual requirement $\mathcal{O}_{\mathbb{P}}(n^{-1/2})$ for negligibility. For example, if $\alpha_n^{(1)} = \mathcal{O}_{\mathbb{P}}(n^{-1/4})$ and $\beta_n = \mathcal{O}_{\mathbb{P}}(n^{-1/4})$, then the plug-in effect for $\Phi^{(1)}(\widehat{\eta}_{\text{cross}})$ is negligible.

Finally, we notice that the $\sqrt{n}$-asymptotic results in Theorem 2 rely on the implication from Assumption 5 that $n_1/n \to \rho_S \ge \delta$ and $n_0/n \to 1 - \rho_S \ge \delta$ as $n \to \infty$, so that both n_1 and n_0 grow linearly in n. When the calibrating sample size n_0 is of a smaller order in the training sample size n_1, the leading order terms for the estimates in (2) are of order $\sqrt{n_0}$. We establish the corresponding asymptotic results in the following Corollary 1 that are different from those in Theorem 2.

Corollary 1 (Semiparametric efficiency with small calibrating data). Consider the observed data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$, the parameters θ and θ_1 in (1), and the corresponding oracle efficient

estimates in (2). Suppose $n_0/n_1 \to 0$ as $n_1, n_0 \to \infty$. Then under Assumptions 1–4, 6 and that $\sup_{\boldsymbol{x}\in\mathcal{X}} w(\boldsymbol{x}) < +\infty$, we have

$$\sqrt{n_0}(\widehat{\theta}^{(2)}_{\text{eff}} - \theta) \xrightarrow[n_0\to\infty]{\mathcal{D}} \mathcal{N}(0, \text{var}[Q(\boldsymbol{X}_i, d)|S_i = 0]);$$
$$\sqrt{n_0}(\widehat{\theta}^{(2)}_{\text{eff}} - \theta) \xrightarrow[n_0\to\infty]{\mathcal{D}} \mathcal{N}(0, \text{var}[Q(\boldsymbol{X}_i, d)|S_i = 0]);$$
$$\sqrt{n_0}(\widehat{\theta}^{(1)}_{1,\text{eff}} - \theta_1) \xrightarrow[n_0\to\infty]{\mathcal{D}} \mathcal{N}(0, \text{var}[C(\boldsymbol{X}_i)d(\boldsymbol{X}_i)|S_i = 0]);$$
$$\sqrt{n_0}(\widehat{\theta}^{(2)}_{1,\text{eff}} - \theta_1) \xrightarrow[n_0\to\infty]{\mathcal{D}} \mathcal{N}(0, \text{var}[C(\boldsymbol{X}_i)d(\boldsymbol{X}_i)|S_i = 0]),$$

which attain the corresponding semiparametric $\sqrt{n_0}$-asymptotic variance lower bounds for θ and θ_1, respectively.

Notice that the $\sqrt{n_0}$-asymptotic variances from Corollary 1 are the same as $\frac{1}{n_0}\sum_{i:S_i=0} Q(\boldsymbol{X}_i, d)$ and $\frac{1}{n_0}\sum_{i:S_i=0} C(\boldsymbol{X}_i)d(\boldsymbol{X}_i)$, respectively, so that all efficiency augmentation terms in (2) are negligible with respect to the $\sqrt{n_0}$-order. Moreover, any estimate $\widehat{Q}(\boldsymbol{x}, a)$ of $Q(\boldsymbol{x}, a)$ that satisfies $\|\widehat{Q}(\cdot, a) - Q(\cdot, a)\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}(n_0^{-1/2})$ for $a \in \{1, -1\}$ will not affect these $\sqrt{n_0}$-asymptotic results. Since we have assumed that $n_0/n_1 \to 0$, it can be relatively easier than Theorem 3 to obtain an estimate $\widehat{Q}$ of order $\mathcal{O}_{\mathbb{P}}(n_0^{-1/2})$ based on the training data.

2.2. Estimating Weights

In Section 2.1, we obtain the efficient estimates for θ and θ_1 on two types of pooled data $\mathcal{O}^{(1)}$ and $\mathcal{O}^{(2)}$ as in (2). We also establish the asymptotic properties in Theorems 2 and 3 for the oracle and cross-fitting efficient estimates, respectively. In this section, we discuss different methods from literature for estimating training sample weights $\{w(\boldsymbol{X}_i) : S_i = 1\}$ or the nuisance function $\boldsymbol{x} \mapsto w(\boldsymbol{x})$. In particular, *augmented inverse probability of sampling weight (AIPSW)* and density ratio estimation are related to the discussions in Li, Li, and Luedtke (2020) and Liang and Zhao (2020), respectively.

2.2.1. Augmented Inverse Probability of Sampling Weight (AIPSW)

Let $\widehat{\pi}_S(1|\boldsymbol{x})$ be an estimate of $\pi_S(1|\boldsymbol{x})$. Define the estimated weights at training data points $\{i : S_i = 1\}$ as

$$\widehat{w}(\boldsymbol{X}_i) := \frac{n_1\widehat{\pi}_S(0|\boldsymbol{X}_i)}{n_0\widehat{\pi}_S(1|\boldsymbol{X}_i)}; \quad \forall i : S_i = 1.$$

Then the *AIPSW estimates* (Stuart et al. 2011; Buchanan et al. 2018) are defined as in (2) with the AIPSWs $\{\widehat{w}(\boldsymbol{X}_i) : S_i = 1\}$ that are obtained from the estimated conditional-on-covariate sampling probability function $\widehat{\pi}_S(s|\boldsymbol{x})$. We further denote the AIPSW estimates as $\widehat{\theta}^{(1)}_{\text{AIPSW}}$, $\widehat{\theta}^{(2)}_{\text{AIPSW}}$, $\widehat{\theta}^{(1)}_{1,\text{AIPSW}}$, and $\widehat{\theta}^{(2)}_{1,\text{AIPSW}}$, respectively. Notice that the AIPSW estimates have one-to-one correspondence to the estimators proposed in Li, Li, and Luedtke (2020, sec. 3.2): $\widehat{\theta}^{(1)}_{\text{AIPSW}} = \widehat{V}^*_{\text{eff}}(d)$, $\widehat{\theta}^{(2)}_{\text{AIPSW}} = \widehat{V}^*_{\text{onlyX}}(d)$, $\widehat{\theta}^{(1)}_{1,\text{AIPSW}} = \widehat{R}^*_{\text{eff}}(d)$, and $\widehat{\theta}^{(2)}_{1,\text{AIPSW}} = \widehat{R}^*_{\text{onlyX}}(d)$.

In practice, $\widehat{\pi}_S(s|\boldsymbol{x})$ can be estimated by logistic regression of S_i on $\boldsymbol{X}_i$. Consider the simulation setup in Li, Li, and Luedtke (2020): $\boldsymbol{X}_i|(S_i = 1) \sim \mathcal{N}_p(\boldsymbol{0}, \mathbf{I}_p)$ and $\boldsymbol{X}_i|(S_i = 0) \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ for some $\boldsymbol{\mu} \in \mathbb{R}^p$. For $s \in \{1, 0\}$, denote $f(\boldsymbol{x}|s)$ as the density of $\boldsymbol{X}_i|(S_i = s)$. The theoretical weight function in this case is $w(\boldsymbol{x}) = \frac{f_X(\boldsymbol{x}|0)}{f_X(\boldsymbol{x}|1)} = \exp(\|\boldsymbol{\mu}\|_2^2/2 - \boldsymbol{\mu}^\intercal\boldsymbol{x})$. The corresponding log-odds of $S_i|(\boldsymbol{X}_i = \boldsymbol{x})$ is: $\log\left(\frac{\pi_S(1|\boldsymbol{x})}{\pi_S(0|\boldsymbol{x})}\right) = \log\left(\frac{\rho_S}{1-\rho_S}\right) - \log[w(\boldsymbol{x})] = \log\left(\frac{\rho_S}{1-\rho_S}\right) - \|\boldsymbol{\mu}\|_2^2/2 + \boldsymbol{\mu}^\intercal\boldsymbol{x}$. In this case, logistic regression can correctly specify $S_i|\boldsymbol{X}_i$. Therefore, the corresponding AIPSW estimates can enjoy semiparametric efficiency, which was illustrated in the numerical studies of Li, Li, and Luedtke (2020, sec. 3.3).

2.2.2. Density Ratio Estimation

Instead of estimating $\pi_S(s|\boldsymbol{x})$ first to obtain $w(\boldsymbol{x}) = \frac{\rho_S}{1-\rho_S}\frac{\pi_S(0|\boldsymbol{x})}{\pi_S(1|\boldsymbol{x})}$, we can directly consider $w(\boldsymbol{x})$ as the covariate density ratio function $\frac{\mathbb{P}(\boldsymbol{X}_i=\boldsymbol{x}|S_i=0)}{\mathbb{P}(\boldsymbol{X}_i=\boldsymbol{x}|S_i=1)}$. Uehara, Kato, and Yasui (2020) proposed to estimate the covariate density ratio for $\{\boldsymbol{X}_i : S_i = 1\}$ using the *kernel-based unconstrained least-squares importance fitting (KuLSIF)* (Kanamori, Suzuki, and Sugiyama 2012), which was also discussed in Liang and Zhao (2020). Specifically, let $\mathcal{G}$ be a generic function class, and consider the following least-squared problem:

$$\begin{aligned}
&\min_{g\in\mathcal{G}} \frac{1}{2}\mathbb{E}\left\{[g(\boldsymbol{X}_i) - w(\boldsymbol{X}_i)]^2\middle|S_i = 1\right\} \\
&= \min_{g\in\mathcal{G}} \left\{\frac{1}{2}\mathbb{E}[g(\boldsymbol{X}_i)^2|S_i = 1] - \mathbb{E}[g(\boldsymbol{X}_i)|S_i = 0]\right\} \\
&\quad + \frac{1}{2}\mathbb{E}[w(\boldsymbol{X}_i)^2|S_i = 1].
\end{aligned}$$

The empirical version of the above least-squared problem is as follows:

$$\min_{g\in\mathcal{G}} \left\{\frac{1}{2n_1}\sum_{i:S_i=1} g(\boldsymbol{X}_i)^2 - \frac{1}{n_0}\sum_{i:S_i=0} g(\boldsymbol{X}_i) + \frac{\lambda}{2}\|g\|^2_{\mathcal{G}}\right\}, \quad (3)$$

where $(\lambda/2)\|g\|^2_{\mathcal{G}}$ is a functional penalty on $\mathcal{G}$. Let K be a positive semidefinite kernel function on $\mathcal{X}$ and $\mathcal{G} = \mathcal{H}_K$ be the corresponding reproducing kernel Hilbert space (RKHS). Define $\mathbf{K}_{11} := [K(\boldsymbol{X}_i, \boldsymbol{X}_j) : S_i = S_j = 1]$ and $\mathbf{K}_{01} := [K(\boldsymbol{X}_i, \boldsymbol{X}_j) : S_i = 0, S_j = 1]$. The solution to (3) can be represented as $g_{\boldsymbol{\alpha}}(\cdot) = \sum_{i:S_i=1}\alpha_i K(\boldsymbol{X}_i, \cdot) + [1/(\lambda n_0)]\sum_{i:S_i=0} K(\boldsymbol{X}_i, \cdot)$ (Kanamori, Suzuki, and Sugiyama 2012, Theorem 1), and the dual optimization problem for $\boldsymbol{\alpha} = (\alpha_i : S_i = 1)^\intercal \in \mathbb{R}^{n_1}$ is

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n_1}} \left\{\frac{1}{2}\boldsymbol{\alpha}^\intercal\left(\frac{1}{n_1}\mathbf{K}_{11} + \lambda\mathbf{I}_{n_1}\right)\boldsymbol{\alpha} - \frac{1}{\lambda n_0 n_1}\mathbf{1}^\intercal_{n_0}\mathbf{K}_{01}\boldsymbol{\alpha}\right\}.$$

Let $\widehat{w}(\boldsymbol{x})$ be the KuLSIF estimate of the covariate weight function. Under certain conditions, Kanamori, Suzuki, and Sugiyama (2012, Theorem 2) showed that $\|\widehat{w} - w\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}\left((n_1 \wedge n_0)^{-1/(2+\gamma)}\right)$ for some $\gamma \in (0, 2)$.

Based on the KuLSIF weights, Liang and Zhao (2020) considered $\widehat{\theta}_1 := (1/n_1)\sum_{i:S_i=1}\widehat{w}(\boldsymbol{X}_i)\widehat{C}(\boldsymbol{X}_i)d(\boldsymbol{X}_i)$. Under Assumptions 1–6, it requires that $\|\widehat{w} - w\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ and $\|\widehat{C} - C\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ to establish the $\sqrt{n}$-consistency of $\widehat{\theta}_1$. However, the KuLSIF estimate of $w(\boldsymbol{x})$ cannot satisfy this rate condition. Even when $\|\widehat{w} - w\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ and $\|\widehat{C} - C\|_{L^2(\mathbb{P})} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ so that $\widehat{\theta}_1$ is $\sqrt{n}$-consistent, the asymptotic variance of $\widehat{\theta}_1$ is generally greater than the semiparametric efficiency bound in Theorem 2. To achieve the efficiency

bound, the KuLSIF density ratio estimate should be combined with a semiparametric efficient estimate and the cross-fitting strategy.

2.2.3. Augmented Calibration Weight (ACW)

Let $\mathcal{G}$ be a function class on $\mathcal{X}$. Lemma 1 can also motivate the following covariate balancing conditions among covariate functions in $\mathcal{G}$ for the training sample weights $\{W_i : S_i = 1\}$ (Imai and Ratkovic 2014):

$$\underbrace{\sum_{i:S_i=1} W_i g(\boldsymbol{X}_i)}_{\text{empirical version of } \mathbb{E}[w(\boldsymbol{X}_i)g(\boldsymbol{X}_i)|S_i=1]} = \underbrace{\frac{1}{n_0}\sum_{i:S_i=0} g(\boldsymbol{X}_i)}_{\text{empirical version of } \mathbb{E}[g(\boldsymbol{X}_i)|S_i=0]} \quad ; \quad \forall g \in \mathcal{G}. \tag{4}$$

Based on the balancing conditions (4), Hainmueller (2012) proposed to solve the entropy balancing problem for calibration weights:

$$\begin{aligned} \min_{W_i:S_i=1} \quad & \sum_{i:S_i=1} W_i \log W_i, \\ \text{subject to} \quad & W_i \geq 0; \qquad \text{for all } i \text{ with } S_i = 1; \\ & \sum_{i:S_i=1} W_i = 1; \\ & \sum_{i:S_i=1} W_i g(\boldsymbol{X}_i) = \frac{1}{n_0}\sum_{i:S_i=0} g(\boldsymbol{X}_i); \quad \forall g \in \mathcal{G}. \end{aligned} \tag{5}$$

Here, the objective function is the empirical Kullback–Leibler divergence of $\mathbb{P}(\cdot|S_i = 0)$ with respect to $\mathbb{P}(\cdot|S_i = 1)$, and W_i is the covariate density ratio at $\boldsymbol{X}_i$. Therefore, the optimization problem (5) seeks for the balancing weights $\{\widehat{W}_i : S_i = 1\}$ that satisfies (4) and minimizes the discrepancy of the testing covariate distribution from training. If $\mathcal{G} = \{g_j\}_{j=1}^m$ and denote $\boldsymbol{g} = (g_1, g_2, \ldots, g_m)^\intercal$ as an m-dimensional vector of instrumental functions, then the calibration weights can be solved by the following dual problem:

$$\widehat{W}_i = \frac{\exp[\widehat{\boldsymbol{\lambda}}^\intercal \boldsymbol{g}(\boldsymbol{X}_i)]}{\sum_{j:S_j=1}\exp[\widehat{\boldsymbol{\lambda}}^\intercal \boldsymbol{g}(\boldsymbol{X}_j)]}; \quad \text{for all } i \text{ with } S_i = 1,$$

where $\widehat{\boldsymbol{\lambda}} \in \mathbb{R}^m$ is solved by the balancing equations (4). The dual solution can be further extended to the case when $\mathcal{G}$ is an RKHS (Zhao 2019). The calibration weights from (5) can also correspond to a parametric model for $w(\boldsymbol{X}_i)$. Specifically, if $w(\boldsymbol{x}; \boldsymbol{\eta}) = \exp[\boldsymbol{\eta}^\intercal \boldsymbol{g}(\boldsymbol{x})]$, then $\widehat{\boldsymbol{\lambda}} \xrightarrow{\mathbb{P}} \boldsymbol{\eta}$, and $\widehat{W}_i = \frac{w(\boldsymbol{X}_i)}{n_1} + \mathcal{O}_{\mathbb{P}}(n^{-1})$ (Dong et al. 2020, Theorem 1). The following two cases can imply such a parametric model:

(I) Logistic regression of S_i on $\boldsymbol{g}(\boldsymbol{X}_i)$ can imply such a parametric model for $w(\boldsymbol{X}_i)$;

(II) If $\boldsymbol{g}(\boldsymbol{X}_i)$ is a sufficient statistic for $\boldsymbol{X}_i$ and $\boldsymbol{g}(\boldsymbol{X}_i)|(S_i = s) \sim \mathcal{N}_m(\boldsymbol{\mu}_s, \Sigma)$ for $s \in \{1, 0\}$, then $w(\boldsymbol{x}) = \exp\left\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \left(\boldsymbol{g}(\boldsymbol{x}) - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0}{2}\right)\right\}$. Because of the additional term $\frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0}{2}$ in $w(\boldsymbol{x})$, a constant function is required in the function class $\mathcal{G}$ for balancing conditions.

Finally, the *ACW estimates* are defined as the efficient estimates (2) with $\{w(\boldsymbol{X}_i) : S_i = 1\}$ as $\{n_1 \widehat{W}_i : S_i = 1\}$ from (5) (Dong et al. 2020). One advantage of such estimates is the implicit nonparametric function class specification for $w(\boldsymbol{x})$.

2.3. Challenges for Efficient Policy Evaluation

To summarize, we have studied the policy evaluation problem when a set of calibrating data from the testing distribution can be used for training. We establish the efficient policy evaluation results based on the EIFs and discussed the properties of efficient estimates that can be related to existing literature and the discussions in Li, Li, and Luedtke (2020) and Liang and Zhao (2020). However, when the calibrating sample size n_0 is small, our discussion suggests the following two main challenges:

(I) Efficient policy evaluation requires a sufficiently large calibrating sample size for useful efficiency gain. Assumption 5 can imply that the asymptotic sampling rates $n_1/n \to \mathbb{P}(S_i = 1)$ and $n_0/n \to \mathbb{P}(S_i = 0)$ are both at least δ, which requires that both the training and calibrating sample sizes n_1 and n_0 grow linearly in n. When $n_0 = \mathcal{O}(n_1)$, Corollary 1 suggests that the $\sqrt{n_0}$-asymptotic efficient estimates in (2) are equivalent to averaging the nonparametric estimates of $Q(\boldsymbol{X}_i, d)$ or $C(\boldsymbol{X}_i)d(\boldsymbol{X}_i)$ over the calibrating data. The complicated forms in (2) may not be helpful for efficiency improvement in this case. More importantly, the resulting estimates can be unstable due to the limited calibrating sample size n_0;

(II) Nuisance function estimates can be hard to obtain in efficient policy evaluation. Specifically, the optimality of efficient estimates requires that the plug-in nuisance function estimates $\widehat{w}(\boldsymbol{x})$, $\widehat{\pi}_A(a|\boldsymbol{x}, s)$, and $\widehat{Q}(\boldsymbol{x}, a)$ are $\sqrt{n}$-negligible as in Theorem 3. However, the challenge for nuisance function estimation mainly appears in the covariate weight function $\widehat{w}(\boldsymbol{x})$, since its rate of convergence is determined by $\min\{n_1, n_0\}$. The same difficulty can appear when estimating the conditional-on-covariate selection probability function $\pi_S(s|\boldsymbol{x})$ in the AIPSW estimates, where a correctly specified parametric estimate $\widehat{\pi}_S(s|\boldsymbol{x})$ is in the $(n_1 \wedge n_0)^{1/2}$-order. Thus, it can be difficult to estimate the nuisance function when n_0 is small.

Given the challenges of efficient policy evaluation with a limited calibrating sample size n_0, DRITR can be less dependent on n_0. First of all, DRITR utilizes the training data only to obtain the set of candidates $\{\widehat{d}_c\}_{c\in\mathcal{C}}$, where $\mathcal{C}$ is a set of candidate DR-constants, and each $\widehat{d}_c$ enjoys certain robust performance guarantee. Then calibrating sample is used to choose the final DRITR. In this way, DRITR is less affected by the size of n_0 compared to the combined strategy in Section 2. Second, the value function estimates used in evaluating candidate DRITRs on the calibrating dataset can still enjoy certain properties with a small n_0. In Mo, Qi, and Liu (2020, sec. 2.4), two calibration procedures were proposed, one using the calibrating covariates only, and another one using the calibrating covariates, treatments and outcomes. In the calibration procedure based on calibrating covariates only, the testing value function estimate for a given ITR d is $\frac{1}{n_0}\sum_{i:S_i=0}\widehat{C}(\boldsymbol{X}_i)d(\boldsymbol{X}_i)$,

which is semiparametric $\sqrt{n_0}$-asymptotic efficient due to Corollary 1. For another calibration procedure based on calibrating covariates, treatments and outcomes, the testing value function estimate is $\frac{1}{n_0}\sum_{i:S_i=0}\frac{\mathbb{1}[d(\boldsymbol{X}_i)=A_i]}{\pi_A(A_i|\boldsymbol{X}_i,1)}Y_i$. Although such an estimate may not achieve the $\sqrt{n_0}$-variance lower bound, it can be robust if Assumption 4 is violated. Specifically, if Assumption 4 is violated, the estimates (2) may not be consistent, while $\frac{1}{n_0}\sum_{i:S_i=0}\frac{\mathbb{1}[d(\boldsymbol{X}_i)=A_i]}{\pi_A(A_i|\boldsymbol{X}_i,1)}Y_i$ remains $\sqrt{n_0}$-consistent. Thirdly, we would like to point out that DRITR avoids estimating the covariate weight function $w(\boldsymbol{x})$. Therefore, it can bypass the challenge of nuisance function estimation given the limited calibrating data.

3. Applicability of DRITR

In Section 1, we distinguish DRITR from retargeted policy learning as it focuses on covariate changes. In Section 2, we consider the problem of covariate changes with calibrating data from a specific testing distribution being available at the training stage. In particular, we discuss the general challenges for efficient policy evaluation when available information from testing is limited, and how DRITR can avoid such challenges. As is mentioned at the beginning of Section 2, DRITR focused on general covariate changes instead of a specific testing distribution. In this section, we discuss the general applicability of DRITR.

We first emphasize that DRITR aims for protecting scientific discoveries from the general agnostic covariate changes. This explains why, in response to Li, Li, and Luedtke (2020), we proposed to work with the least favorable case among some possible covariate changes. In fact, the concern on potential training-testing distributional changes can be important in modern prediction methodology. Efron (2020, sec. 6) discussed their analysis on the prostate cancer microarray study. If they randomly split data into training and testing, then the testing error of a random forest classifier can be as low as 2%. However, if they selected patients with the lowest ID numbers into the training dataset, with the remaining for testing purpose, then the testing error would be as high as 24%. We also performed similar analysis on the ACTG175 study. In particular, we found that when testing on the female population, several other existing methods can have poorer performance than DRITR. Such a violation of the identical training and testing distributions can undermine an existing scientific finding, and researchers may question the faithfulness of such a finding when generalizing it to a much broader scope. On the contrary, a scientific finding robust to all such violations can typically be closer to universal, eternal truths and become long-lasting (Efron 2020). The same scientific principle has also been advocated in Yu and Kumbier (2020) and Bühlmann (2020), both of which established nice connections of such a principle with adversarial perturbation and distributional robustness.

DRITR can correspond to a more "forgiving" but useful approach than precise estimation. On one hand, we agree with Dukes and Vansteelandt (2020) that making correct "causal predictions," that is, estimating the CTE function correctly, can be the most robust way of protecting from covariate changes. In fact, we highlight in our Section 1 that general applicability of DRITR relies on the assumption $\{\boldsymbol{x}\mapsto \mathsf{sign}[C(\boldsymbol{x})]\}\nsubseteq\mathcal{D}$. One example is when the true CTE function $C(\boldsymbol{x})$ is too complicated to be estimable, the ITR class $\mathcal{D}$ can be misspecified. Another example is that the CTE function $C(\boldsymbol{x})$ takes a complicated functional form, and $\mathcal{D}$ is intended for a more parsimonious class of decision rules in practice. In either of these two cases, DRITR can be a useful methodology with tolerance on incorrect "causal predictions." On the other hand, given our combined data analysis in Section 2, efficient inference of parameters of interests can have more restrictive requirements on data availability and involve more assumptions. In contrast, the requirements for accurate predictions, for example, predicting which treatment to assign in our context, can typically be less stringent than drawing efficient inference of parameter estimates, as was discussed in Efron (2020, Criteria 6). These can distinguish the prediction-driven focus and usefulness of DRITR. While an inference-based criterion can only be applicable if all required assumptions hold, a prediction-based criterion particularly focuses on some measurements of testing performance and can be less restrictive. Therefore, even though DRITR can be conservative by performing worst-case policy optimization, it can enjoy less restrictions and more general applicability.

The last point we would like to point out is that the training of candidate DRITRs can be performed before using calibrating data. This can provide more privacy protection. Specifically, DRITR can utilize the training data to obtain a class of candidate ITR estimates $\{\widehat{d}_c\}_{c\in\mathcal{C}}$, where $\mathcal{C}$ is the set of candidate DR-constants. When estimating the optimal DRITR on a specific testing distribution, we only use the testing information to choose the best ITR from $\{\widehat{d}_c\}_{c\in\mathcal{C}}$ without requesting for the complete training data. In contrast, the combined analysis in Section 2 requires at least either $\{\boldsymbol{X}_i,\widehat{Q}(\boldsymbol{X}_i,\pm1):S_i=1\}$ or $\{\boldsymbol{X}_i,\widehat{C}(\boldsymbol{X}_i):S_i=1\}$ from the training data. In this case, treatment effect information at the individual level would be exposed to the testing agents. Therefore, the individualized treatment effect information obtained from training can be kept privately when applying DRITR, but cannot when using methods based on combined data in Section 2.

Acknowledgments

The authors would like to thank the editor and the associate editor for organizing this insightful discussion.

Funding

This research was partially supported by NSF grant DMS1821231 and NIH grants R01GM126550, P01CA142538.

References

Athey, S., and Wager, S. (2020), "Policy Learning With Observational Data," *Econometrica* (to appear). [700]

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018), "Generalizing Evidence From Randomized Trials Using Inverse Probability of Sampling Weights," *Journal of the Royal Statistical Society*, Series A, 181, 1193–1209. [704]

Bühlmann, P. (2020), "Invariance, Causality and Robustness," *Statistical Science*, 35, 404–426. [706]

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21, C1–C68. [703]

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2006), "Moving the Goalposts: Addressing Limited Overlap in the Estimation of Average Treatment Effects by Changing the Estimand," Technical Report, National Bureau of Economic Research. [700]

——— (2009), "Dealing With Limited Overlap in Estimation of Average Treatment Effects," *Biometrika*, 96, 187–199. [700]

Dahabreh, I. J., Robertson, S. E., Petito, L. C., Hernán, M. A., and Steingrimsson, J. A. (2019), "Efficient and Robust Methods for Causally Interpretable Meta-Analysis: Transporting Inferences From Multiple Randomized Trials to a Target Population," arXiv no. 1908.09230. [701,702]

Dong, L., Yang, S., Wang, X., Zeng, D., and Cai, J. (2020), "Integrative Analysis of Randomized Clinical Trials With Real World Evidence Studies," arXiv no. 2003.01242. [705]

Dukes, O., and Vansteelandt, S. (2020), "Discussion of Kallus and Mo, Qi, and Liu: New Objectives for Policy Learning." [699,706]

Efron, B. (2020), "Prediction, Estimation, and Attribution," *Journal of the American Statistical Association*, 115, 636–655. [706]

Hainmueller, J. (2012), "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 20, 25–46. [705]

Imai, K., and Ratkovic, M. (2014), "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society*, Series B, 76, 243–263. [705]

Kallus, N. (2020), "More Efficient Policy Learning via Optimal Retargeting," *Journal of the American Statistical Association* (to appear). [699,700,702]

Kanamori, T., Suzuki, T., and Sugiyama, M. (2012), "Statistical Analysis of Kernel-Based Least-Squares Density-Ratio Estimation," *Machine Learning*, 86, 335–367. [704]

Li, S., Li, X., and Luedtke, A. (2020), "Discussion of Kallus (2020) and Mo, Qi, and Liu (2020): New Objectives for Policy Learning," arXiv no.2010.04805. [699,700,704,705,706]

Liang, M., and Zhao, Y. (2020), "Discussion of Kallus (2020) and Mo et al. (2020)." [699,700,704,705]

Mo, W., Qi, Z., and Liu, Y. (2020), "Learning Optimal Distributionally Robust Individualized Treatment Rules," *Journal of the American Statistical Association* (to appear). [699,700,701,705]

Molina, J., Rotnitzky, A., Sued, M., and Robins, J. (2017), "Multiple Robustness in Factorized Likelihood Models," *Biometrika*, 104, 561–581. [700]

Pearl, J., and Bareinboim, E. (2014), "External Validity: From Do-Calculus to Transportability Across Populations," *Statistical Science*, 29, 579–595. [701]

Qi, Z., Cui, Y., Liu, Y., and Pang, J.-S. (2019), "Estimation of Individualized Decision Rules Based on an Optimized Covariate-Dependent Equivalent of Random Outcomes," *SIAM Journal on Optimization*, 29, 2337–2362. [700]

Robins, J. M., Rotnitzky, A., and van der Laan, M. (2000), "On Profile Likelihood: Comment," *Journal of the American Statistical Association*, 95, 477–482. [699]

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701. [700]

Rudolph, K. E., and 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. [701,702]

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and 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. [700,704]

Tsiatis, A. (2007), *Semiparametric Theory and Missing Data*, New York: Springer-Verlag. [702]

Uehara, M., Kato, M., and Yasui, S. (2020), "Off-Policy Evaluation and Learning for External Validity Under a Covariate Shift," in *Advances in Neural Information Processing Systems* (to appear). [701,702,703,704]

Yu, B., and Kumbier, K. (2020), "Veridical Data Science," *Proceedings of the National Academy of Sciences of the United States of America*, 117, 3920–3929. [706]

Zhao, Q. (2019), "Covariate Balancing Propensity Score by Tailored Loss Functions," *The Annals of Statistics*, 47, 965–993. [705]

Zhao, Y.-Q., Zeng, D., Tangen, C. M., and Leblanc, M. L. (2019), "Robustifying Trial-Derived Optimal Treatment Rules for a Target Population," *Electronic Journal of Statistics*, 13, 1717–1743. [700]