

Meta-analysis of individualized treatment rules via sign-coherency

Jay Jojo Cheng

JAY.CHENG@WISC.EDU

Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison

Jared D. Huling

HULING@UMN.EDU

Division of Biostatistics, University of Minnesota

Guanhua Chen

GCHEN25@WISC.EDU

Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison

Abstract

Medical treatments tailored to a patient’s baseline characteristics hold the potential of improving patient outcomes while reducing negative side effects. Learning individualized treatment rules (ITRs) often requires aggregation of multiple datasets/sites; however, current ITR methodology does not take between-site heterogeneity into account, which can hurt model generalizability when deploying back to each site. To address this problem, we develop a method for individual-level meta-analysis of ITRs, which jointly learns site-specific ITRs while borrowing information about feature sign-coherency via a scientifically-motivated directionality principle. We also develop an adaptive procedure for model tuning, using information criteria tailored to the ITR learning problem. We study the proposed methods through numerical experiments to understand their performance under different levels of between-site heterogeneity and apply the methodology to estimate ITRs in a large multi-center database of electronic health records. This work extends several popular methodologies for estimating ITRs (A-learning, weighted learning) to the multiple-sites setting.

Keywords: Individualized treatment rule; Personalized medicine; Meta-analysis; Causal inference

1. Introduction

Two major trends in research and development across established pharmaceutical, biotech, and medical devices markets are the increasing use of real-world evidence (RWE) - observational data generated through routine clinical practice - and treatment benefit-risk analyses personalized to patients. Regulatory agencies in the U.S., Europe, and Japan have all issued guidance on the usage of RWE in supporting regulatory submissions and have moved to encourage the role of RWE throughout the entire drug development life-cycle, beyond the current use focused in the post-authorization phase (Nishioka et al., 2022; European Medicines Agency, 2020; FDA, 2019).

Concurrently, widespread recognition that the best treatment on average for a group of trial participants may not be the best treatment for a particular individual has led to a literature on estimating treatment decision rules specific to individuals, called *individualized treatment rules* (ITRs) or individualized policies. These decision functions map patient baseline covariates to treatment

options (Murphy, 2003; Zhao et al., 2012; Kosorok and Laber, 2019). Learning optimal ITRs relies on detecting subgroup heterogeneities in interactions between covariates and treatment, and thus requires large samples. However, the main limitation in greater RWE utilization is data availability (Flynn et al., 2020), which can be even more pronounced in many therapeutic areas with small patient populations (e.g. rare disease, elderly, pregnant women). The need for large samples coupled with limited data availability leads to a core tension in ITR estimation: pooling datasets between sites is common practice, but existing ITR methodology assumes an idealized sample, wherein all observations are drawn independently from a single population. Furthermore, a single globally estimated treatment rule may not be flexible enough to reflect the optimal ITRs at each clinical site.

In light of these trends, we developed an individual-level meta-analysis (ILMA) framework for adapting existing ITR methodologies to the heterogeneous data setting. Our framework uses a scientifically motivated ‘directionality principle,’ operationalized as an optimization penalty, to balance heterogeneity and information borrowing across sites, as well as a data-adaptive procedure for tuning site-specific models. Under our experiments, when sites are moderately heterogeneous, i.e. covariate shift and concept drift, our method outperforms both pooled approaches that estimate a single global model and separate learning approaches that use only site-specific information. Our analysis of transthoracic echocardiography use in a multi-center database also showcases its performance in sites with low data availability. Code for our implementation and scripts for reproducing all analyses are available at <https://github.com/jay-jojo-cheng/metacoop>.

1.1. Related Work

ITR estimation is closely related to the problem of conditional average treatment effect estimation (CATE), for which there is a large and successful literature (Wager and Athey, 2018; Dorie et al., 2019; Foster and Syrgkanis, 2019; Kennedy, 2020; Curth and van der Schaar, 2021; Fan et al., 2022). Indeed, with access to the true CATE, the Bayes optimal ITR is just the sign of the CATE. This motivates a class of ITR estimation methods, called *indirect methods*, to first estimate the CATE and then invert these regression estimates for estimating ITRs. One potential drawback of this approach is that if the outcome models are misspecified, the estimators may not be consistent for estimating the optimal ITR within a given class. But ITR estimation and CATE estimation are different learning tasks, requiring different trade-offs. For an extended perspective on the differences between CATE estimation and ITR estimation, see Fernández-Loría and Provost (2022).

Recognition of this has led to a literature on *direct methods* for estimating ITRs. These methods are also called policy learning (Beygelzimer and Langford, 2009) or value-search (Zhang et al., 2015) methods which target the clinical outcome and optimize for a decision rule with the best performance in a given class (Zhao et al., 2012; Zhu et al., 2017). Direct methods decouple the choice of models for the conditional outcome distribution from the choice of models for the ITR class so that the outcome distribution can be modeled flexibly while still controlling the complexity of treatment policy that will be implemented. Directly targeting treatment assignments over first estimating the CATE can also avoid misspecification issues while maximizing the performance of the ITR within the given class.

Chen et al. (2017) proposed a general statistical framework that unifies many direct methods in the binary treatment regime and our method can be seen as extending this framework to jointly learning different ITRs over multiple datasets. Qi et al. (2020) proposed a method for ITR estimation in the multi-treatment regime and Chen et al. (2016) studied a continuous treatment (e.g. dose-finding).

The only other work to study ITR estimation under penalization that we are aware of is Mbakop and Tabord-Meehan (2021) from the econometrics literature, which derives regret bounds of policy learning estimators with statistical penalties. However, differences with the current work include the zero-one loss function, single-dataset (IID) setting, and low-dimensional covariate in its application. The work also differs in spirit because the authors use penalization to achieve model selection, while our penalization is motivated by borrowing strength between studies.

We are not aware of any ITR methodology that addresses site heterogeneity arising from pooling multiple datasets. Within the literature on CATE estimation, Tan et al. (2022) proposed a model-averaging method for improving CATE estimation in this regime. Danieli and Moodie (2022) proposed an indirect method for learning a single ITR from several sites that focuses on preserving data privacy at each site. Although the main goal of the current is borrowing strength between estimated ITRs to adapt to heterogeneity, in the appendix, we also show how our algorithms can be adapted to the setting of distributed datasets and compare it with existing literature on privacy-preserving meta-analyses.

1.2. Notation and causal setup

We denote the number of datasets (e.g. sites, hospitals, studies) as $K \in \mathbb{N}$. Within each of the K studies, indexed by lower-case $k \in [K] = \{1, \dots, K\}$, we observe an IID sample of triplets (Y_i^k, X_i^k, T_i^k) , drawn from some joint distribution \mathbb{P}^k . Here, Y_i^k is the outcome variable (assume without loss of generality that smaller Y is better), and $X_i^k \in \mathcal{X}^k$ is a p -dimensional random vector of covariate data, and $T_i^k \in \{-1, 1\}$ is control or treatment that is observed. We use n^k to denote the total number of patients in the k th site and i indexes the observation within the site. Distributions may differ between datasets, that is, $\mathbb{P}^k \neq \mathbb{P}^{k'}$. To be clear, we allow for heterogeneity in both $\mathbb{P}^k(x) \neq \mathbb{P}^{k'}(x)$ (covariate shift) and $\mathbb{P}^k(y^{(1)}, y^{(-1)}|x) \neq \mathbb{P}^{k'}(y^{(1)}, y^{(-1)}|x)$ (concept drift). When context allows, we may drop the superscript k for notational brevity.

The p features in X_i^k are indexed by j and indexing is such that the j th feature across the K studies are similar. We will sometimes refer to the j th variables across all studies as the j th variable group. For simplicity, we only consider the case when each site has exactly one feature in the j th variable group, but extensions of our methodology to the general case are straightforward (Chiquet et al., 2012). To summarize, we use indices (and total number) i (n^k) for observations, j (p) for variables, and k (K) for studies.

A treatment rule d is a deterministic function over the space of covariates, mapping covariates to control or treatment: $d : \mathcal{X} \rightarrow \{-1, 1\}$. We wish to estimate an individualized treatment rule d^k for each site. We also adopt the potential outcome framework for causal inference (Rosenbaum and Rubin, 1983; Rubin, 2005). $Y^{k(1)}$ denotes the potential outcome related to receiving the treatment, and $Y^{k(-1)}$ denotes the potential outcome related to receiving the control. Let

$\pi^k(t, x) = P(T^k = t | X^k = x)$ be the true propensity score for the k th site. We also adopt the following standard causal assumptions:

1. Consistency/stable unit treatment value (SUTVA): $Y^k = I(T = 1)Y^{k(1)} + I(T = -1)Y^{k(-1)}$
2. Strong ignorability (no unmeasured founders): $T^k \perp (Y^{k(1)}, Y^{k(-1)}) | X^k$
3. Positivity: $0 < \pi^k(t, x) < 1$ for all x that are observed.

2. ITR problem setup

We briefly review the ITR learning problem in a single dataset setting via the [Chen et al. \(2017\)](#) framework before extending the formulation to the distributed learning over multiple datasets. We define the *conditional average treatment effect* as $\Delta(x) = \mathbb{E}[Y^{(1)} - Y^{(-1)} | X = x] / 2$. This quantity reflects the average part of the outcome that is affected by treatment among patients with covariates x . We similarly define the *conditional average main effect* as $\mu(x) = \mathbb{E}[Y^{(1)} + Y^{(-1)} | X = x] / 2$, which reflects the average part of the outcome that is not influenced by treatment. Algebraically, the conditional expected potential outcome can be expressed $\mathbb{E}[Y | T = t, X = x] = \mu(x) + t\Delta(x)$. We emphasize that this decomposition holds without any *modeling* assumption on the relationship between features and outcome.

Our goal is to estimate a treatment rule, d , that minimizes the average loss in potential outcome (since lower is better) if used to assign treatment. We define the *value function* $V(d)$ as $V(d) := \mathbb{E}[Y^{(d(X))}]$, which is the expected potential outcome from using decision rule d to assign treatment. We define the optimal ITR as a decision function that minimizes the value function over a suitable

class of decision rules, \mathcal{D} ,

$$d^{\text{opt}} = \arg \min_{d \in \mathcal{D}} V(d). \quad (1)$$

We note that $-\text{sign}(\Delta(x))$ is the unconstrained minimizer of the value function over all possible decision rules (not just the ones in the class \mathcal{D}), but searching for this is combinatorially hard without any assumptions on the data distribution or form of $\Delta(x)$. Instead of using equation (1) to directly minimize for d , we identify $d^{\text{opt}}(\cdot) \in \{-1, 1\}$ as the sign of a *scoring function* f^{opt} of the covariates $d^{\text{opt}}(x) = -\text{sign}(f^{\text{opt}}(x))$, with $f^{\text{opt}} : \mathcal{X} \rightarrow \mathbb{R}$. We can estimate f^{opt} by using the following lemma due to [Chen et al. \(2017\)](#):

Lemma 1 *Under the three causal assumptions above, $-\text{sign}(f^{\text{opt}})$ is an optimal individualized treatment rule if f^{opt} is a minimizer of the population loss function,*

$$\mathcal{L}(f) = \mathbb{E} \left[\frac{(Y - Tf(X))^2}{\pi(T, X)} \right]. \quad (2)$$

We provide a full proof of this lemma in [Appendix A](#) for completeness. This lemma motivates the empirical two-step M-estimator

$$\begin{aligned} & \arg \min_f \mathcal{L}_W(f) \\ &= \arg \min_f \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - T_i f(X_i))^2}{\hat{\pi}(T_i, X_i)}, \end{aligned} \quad (3)$$

where $\hat{\pi}(T_i, X_i)$ is a consistent estimator of the propensity score, corresponding to *weighted (covariate) learning* of [Tian et al. \(2014\)](#). Using similar arguments as in the proof of [Lemma 1](#), one can also conduct a similar derivation for *A-learning* ([Murphy, 2003](#)) and arrive at its corresponding loss function,

$$\begin{aligned} & \mathcal{L}_A(f) \\ &= \frac{1}{n} \sum_{i=1}^n \left(Y_i - \left\{ \frac{T_i + 1}{2} - \hat{\pi}(1, X_i) \right\} f(X_i) \right)^2. \end{aligned} \quad (4)$$

In the multiple site setting, our goal is to estimate separate ITRs d^1, \dots, d^k that minimize the *multi-sites value function*, the sum of the expected potential outcomes under site-specific treatment rules:

$$\sum_{k=1}^K V^k(d^k) = \sum_{k=1}^K \mathbb{E}[Y^{k(d^k(X^k))}]. \quad (5)$$

Recalling the identification of the optimal treatment rules as the negative sign of the conditional average treatment effect, the unconstrained minimizer of (5) can be identified as the minimizer of the *multi-site loss function* (cf. Lemma 1), since we can carry out the minimization for each summand separately:

$$\sum_{k=1}^K \mathbb{E} \left[\frac{(Y^k - T^k f^k(X^k))^2}{\pi^k(T^k, X^k)} \right]. \quad (6)$$

In the analysis of multiple sites, the conditional distribution of potential outcomes given baseline measurements (concept drift) and the distributions of X (covariate shift) can be different; thus, the value functions can be different between sites. For example, consider the joint ILMA of two medical claims datasets, one with payment information from a private insurer and the other containing payment information from Medicare. Even though both outcomes are measured in dollars, they are best considered as different random variables. One Medicare dollar is not equal to one private payer dollar, with Medicare payments generally less than private insurer payments for the same procedure. Minimizing ITR loss functions over each of these datasets are related, but different, learning tasks, and the additive structure in equation (5) gives the flexibility to take this heterogeneity into account.

2.1. Efficiency augmentation

Minimization of equations (3) or (4) yield valid estimators of site-specific ITRs, but for

a good choice of $\hat{a}(x)$, the augmented loss function

$$\frac{1}{n} \sum_{i=1}^n \frac{(Y_i - \hat{a}(X_i) - T_i f(X_i))^2}{\hat{\pi}(T_i, X_i)}, \quad (7)$$

can improve estimation efficiency in finite sample settings. The theoretically best augmentation function depends on f^{opt} , but because this optimal decision function is unknown and itself a target of estimation, in practice a working estimate (Tian et al., 2014) or a marginalization over the treatment is used (Chen et al., 2017). We pursue the latter here and improve estimation by using an ‘adequate’ augmentation function that approximates $E[Y|X]$ over the samples with the form

$$\hat{a}(X_i) = \sum_{t \in \{-1, 1\}} \hat{\pi}(t, X_i) \hat{a}(t, X_i),$$

where $\hat{a}(t, x)$ is an estimate of $E[Y|X = x, T = t]$. This choice is adequate in the sense that it is similar to the optimal augmentation function, thus it will generally lead to improvements in efficiency. In the multiple sites setting, this augmentation function is estimated locally at each site, that is

$$\hat{a}^k(X_i^k) = \sum_{t \in \{-1, 1\}} \hat{\pi}^k(t, X_i^k) \hat{a}^k(t, X_i^k).$$

We do so with random forests, but other regression approaches, e.g. splines, neural nets, can also be used for main effect augmentation in this framework.

2.2. Comparison baselines

Two natural comparison baselines to consider are locally learned models, where each site ignores data available at other sites and computes a ‘separate’ model

$$\hat{f}_{\text{separate}}^k = \arg \min L_W^k(f^k),$$

where L_W^k is the site-specific weighted learning loss function from (3). This gives each site the flexibility to learn its own ITR and should perform well when the optimal ITRs are quite different at each site. When each estimator is consistent, the joint estimator $\hat{f}_{\text{separate}} = (\hat{f}_{\text{separate}}^1, \dots, \hat{f}_{\text{separate}}^K)$ is also consistent for the multi-site loss function. The second comparison is a ‘pooled’ model, where all samples are treated as a single dataset and a single ITR is jointly learned:

$$\hat{f}_{\text{pooled}}^k = \arg \min_f \sum_{k=1}^K \sum_{i \in [n^k]} \frac{(Y_i^k - T_i^k f(X_i^k))^2}{\hat{\pi}(T_i^k, X_i^k)}.$$

The pooled model achieves optimal statistical error in the case where populations and decision functions are homogeneous across studies, but not when they can differ. Lastly, we note that local models may be computationally infeasible due to the high dimensional feature space and an insufficient number of samples in a single site.

3. Directionality principle

3.1. Sign-coherent penalization

To strike a balance between the flexibility of multiple treatment rules and the efficiency gains in pooling samples, we propose a directionality principle for borrowing information:

Locally, the same feature across studies should have the same directional contribution to its corresponding scoring function.

As a concrete example, we have intuitively expect optimal site-specific decision rules to be smooth and directionally similar. We do not expect that small increases to blood pressure leads to a greater preference for the treatment in one location while small increases to blood pressure in another clinical location leads to a preference for the control.

We first formalize the above ‘directionality principle’ in the special case of a linear scoring function, say $f^k(x) = \langle \beta^k, x \rangle$ for some parameter β^k , before we generalizing this to the case when f comes from a weakly differentiable function class. We denote the concatenated parameter vector $\beta = [\beta^{1\top} \dots \beta^{K\top}]^\top$. The multi-site loss function for linear scoring functions becomes

$$L_W(\beta) = \sum_{k=1}^K \frac{1}{n^k} \sum_{i=1}^{n^k} \frac{(Y_i^k - T_i^k \langle \beta^k, X_i^k \rangle)^2}{\pi(T_i^k, X_i^k)} \quad (8)$$

for weighted learning and

$$L_A(\beta) = \sum_{k=1}^K \frac{1}{n^k} \sum_{i=1}^{n^k} (Y_i^k - \{ \frac{T_i^k + 1}{2} - \pi(X_i^k) \} \langle \beta^k, X_i^k \rangle)^2 \quad (9)$$

for A-learning. We add to the multi-site loss functions the following directionality penalization term:

$$\lambda P(\beta) = \lambda \sum_{j=1}^J (||(\beta_j)_+|| + ||(\beta_j)_-||), \quad (10)$$

where the norm is Euclidean and $(v)_+$ and $(v)_-$ return the componentwise positive and negative parts of the vector matching the particular sign. For example, if $\beta_1 = [1, 2, 3, -1, -2, -3]$, then $(\beta_1)_+ = [1, 2, 3, 0, 0, 0]$. The objective for weighted learning becomes

$$L_W(\beta) + \lambda P(\beta) \quad (11)$$

and for A-learning becomes

$$L_A(\beta) + \lambda P(\beta). \quad (12)$$

Here, we use β_j to denote the length K vector of the j th element from each site $\beta_j = [\beta_j^1 \dots \beta_j^K]^\top$, which we call the j th coordinate block. In the specialization

of our directionality principle to linear scoring functions, the optimization objective and penalty, $L(\beta) + \lambda P(\beta)$, can be interpreted as solving a weighted least squares objective with the cooperative LASSO penalty (Chiquet et al., 2012; Huling et al., 2021). The geometry of the cooperative LASSO constraint encourages the j th coefficient of each site to match in sign such that they are zero and positive or zero and negative. This constraint encourages the same group of features to have an effect in the same direction and it penalizes large effects with opposing signs among similar features.

3.2. Directionality principle for general function classes

In the general case where f comes from arbitrary function classes, we formulate the directionality principle with function norms of gradients. Here, the penalty added to the loss function takes the form:

$$\lambda \|f\|_{\mathbb{H}} \equiv \lambda \sum_{j=1}^p \left[\left\| \begin{pmatrix} \partial_j f^{\{1\}} \\ \vdots \\ \partial_j f^{\{K\}} \end{pmatrix} \right\|_{L_2} \right]_+ + \left\| \begin{pmatrix} \partial_j f^{\{1\}} \\ \vdots \\ \partial_j f^{\{K\}} \end{pmatrix} \right\|_{L_2} \right]_-,$$

where $\|g\|_{L_2}$ is the L_2 -norm $(E[g(X)^\top g(X)])^{1/2}$, λ is a tuning parameter, and $(v)_+$ and $(v)_-$ denote positive and negative parts of the vector v . Here we integrate the positive and negative parts of the gradient. When f is a linear model, we recover the special case of the previous section. It is straightforward to check that this is indeed a norm and as suggested by the notation, it can be identified as the Hilbert norm of a reproducing kernel Hilbert space since it is bounded by two times the usual L^2 norm. Our learning problem with this generalized directionality penalty can

be interpreted as a particular instance of kernel ridge regression, leading to statistical properties that are well understood (Vovk, 2013; Steinwart et al., 2009).

3.3. Adaptive Information Criteria

The optimal choice of tuning parameter λ reflects the degree to which sign-incoherence is penalized. Typically, hyperparameter tuning requires computationally intensive model re-fitting, as in cross-validation. Here, we use an adaptive variant of the concordance information criterion (CIC) and value information criterion approaches (VIC) (Shi et al., 2021) which have the advantages of reusing computation from the augmentation functions and weights (thus skipping computationally intensive cross-validation) and directly targeting the ITR value function, as opposed to optimizing prediction error. We provide more details in Appendix F.

4. Evaluation

4.1. Simulation study

In evaluating the proposed methodology, we consider three main scenarios with low, medium, and high site heterogeneity. The basic scenario has moderate inter-site treatment effect heterogeneity in order not to favor either the pooled or separate learning approach. Scenarios two and three have high and low levels of inter-site treatment effect heterogeneity, respectively, where the separate analysis or a pooled analysis are alternatively favored. Each scenario is simulated in 18 sub-scenarios, for a total of 54 sub-scenarios, varying in the number of samples per site ($n = 50, 100, 200$), the number of total predictors ($p = 100, 500$), and the number of studies ($K = 3, 5, 10$). We pay special attention to these “low” data per site settings to mimic the situation in many aggregated databases: small amounts

of site-specific data, but large in the aggregate. Moreover, this is the “difficult” regime for ITR methods. As will be seen in the results, sensible estimators generally perform well with enough data. Each sub-scenario is replicated 500 times. For space considerations, we show one sub-scenario for each of the scenarios in the main text and show the full figures in the appendix.

Model parameters for the data generation process are randomly drawn from heavy-tailed distributions and our goal is to see how the meta-analysis approach balances the homogeneity and heterogeneity between sites to either borrow strength or increase flexibility. We compare the methodology on relative value function, or the fraction of the maximum potential value captured by the treatment rule, and the accuracy of the treatment rule compared to the true sign of the treatment effect.

The estimation methods compared are two meta-analysis methods, with weighted and A-learning as the base learners; three centrally pooled methods, pooled weighted learning for a single model, pooled A-learning for a single model, A-learning with group lasso penalty to fit site-specific models; and two local methods, separate weighted learning, separate A-learning.

The first two meta-analysis methods use our overall method and a CIC step. The pooled weighted and A-learning analyses represent an idealized analysis with individual-level data sharing but a single global model is learned. We combine the K datasets into a single dataset and estimate a single ITR using weighted or A-learning. A-learning with a group lasso penalty fits site-specific models constrained with a group lasso penalty, representing a pooled approach that attempts to incorporate site-specific tailoring under an alternative penalization. The separate analyses estimate K ITRs on the datasets separately and represent local anal-

yses done without any information sharing between models. The meta-analyses refer to the method described in this article.

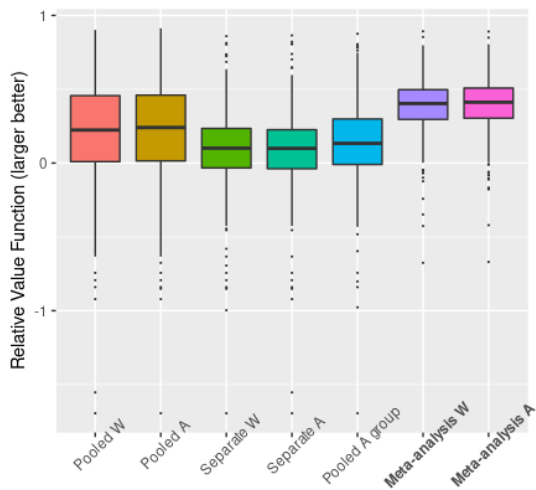
Propensity scores and main effect augmentations are both fit with random forests using 5-fold cross-fitting (Appendix D). The same weights and augmentations are used across all methods in order to compare the treatment rule estimation scheme rather than nuisance parameter estimation. A detailed description of the data generation process is given in Appendix G.

4.1.1. EXPERIMENTAL FINDINGS

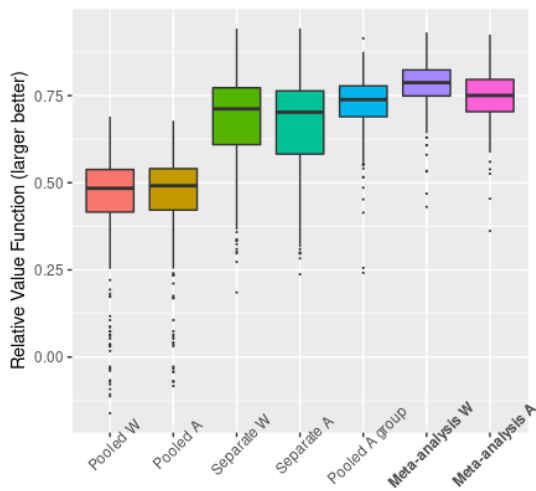
Under low site heterogeneity, pooled methods are expected to outperform separate methods, because they are able to leverage a larger sample size. The sub-scenario shown in the main text typifies the general trends in the full set of simulations. Here we see that the meta-analysis methods perform comparably or even better. In particular, the variance seems to be smaller. In the full results of Figure 3, we can see that sometimes the performance of meta-analysis is able to quickly catch up as the amount of data increases. Both vastly outperform local separate learning analyses.

Figure 4 shows a setting where the true treatment rules are moderately heterogeneous between sites. The blue dotted line indicates the average performance of the pooled oracle predictor within the rule class - fitting a site-specific treatment rule with access to infinite samples at each site. In a moderately heterogeneous setting, the centralized estimator that fits a single model still outperforms the site-specific local estimators. As the number of sites or dimensionality increases, the advantage of the meta-analysis over the local models improves.

Under high site heterogeneity, the pooled estimator underperforms against the local models. Separate learning methods are expected to outperform the pooled methods,



(a) Low site heterogeneity



(b) High site heterogeneity

Figure 1: Meta-analysis adapts to the level of site heterogeneity compared to base learners.

because the optimal ITR at each site is quite different. Here the increase in sample size does not provide an advantage for estimation compared to the heterogeneity of the datasets. Here, we also see that meta-analysis is flexible enough to fit site-specific ITRs, while still borrowing strength to improve estimation, even though the sites are quite different. The full results in Figure 5 show that although the true treatment rules are very different between studies, the meta-analysis method is still able to outperform the local models by taking advantage of sign-coherency.

The accuracy results (reported in the appendix) are similar to the relative value function but more pronounced. In terms of treatment accuracy, local approaches lag behind until the sample size increases. The accuracy of pooled methods also is extremely variable in low-data settings. This suggests that the pooled methods are more accurate on patients with larger conditional average treatment effects, but accuracy may be unstable in others.

4.2. Individual-level meta-analysis of transthoracic echocardiography

We apply the proposed methods above to reanalyze the ECHO study (Feng et al., 2018) on the use of transthoracic echocardiography (TTE) in ICU patients with sepsis. Echocardiography can be used in critically ill patients to detect unsuspected cardiac abnormalities, but it is unclear whether or not care management arising from the echocardiography result alters patient outcomes. Feng et al. (Feng et al., 2018) studied whether or not TTE use improves outcomes for patients on average, using the MIMIC III database of ICU patients from the Beth Israel Deaconess Medical Center in Boston, Massachusetts (Johnson et al., 2016; Goldberger et al., 2000). Their analytic sam-

ple consists of 6361 patients and 40 baseline variables, including demographics, admissions data, vitals, labs, and other medical interventions, and their goal was to estimate the average causal treatment effect of using TTE.

In our analysis, we replicate the ECHO study cohort within the publicly available eICU database (Goldberger et al., 2000). Replicating the ECHO study with the eICU database typifies standard meta-analyses because eICU consists of 208 hospitals categorized into Midwest, Northeast, South, and West geographical regions. We thus split the data into four ‘sites’ based on these regions. Noting that the MIMIC database does not contribute any patients toward eICU, our sample consists of 21,317 new patients and 38 baseline variables chosen based on the original study. We use our methodology to estimate a decision rule that determines whether or not TTE should be used on a given patient to minimize hospital length of stay. To account for outcome censoring in 11.4% of patients, we use random survival forests (Ishwaran et al., 2008) to impute discharge times with a procedure described in detail in Appendix J. Less than 5% of data cells in the baseline features were missing, and these were imputed using random forests (Tang and Ishwaran, 2017). We reemphasize that our estimation target is no longer an average causal treatment effect as in the original study (causal effect estimation), but an *individualized* treatment rule (causal decision-making).

We vary the amount of training data used from 2% to 25% and let the test set be the remainder of the data. This was done to assess performance in small data settings that commonly appear in biomedical applications. At each level of training data used, for 500 replications, the data is randomly split into training and test data. On the training sets, we run the proposed methodologi-

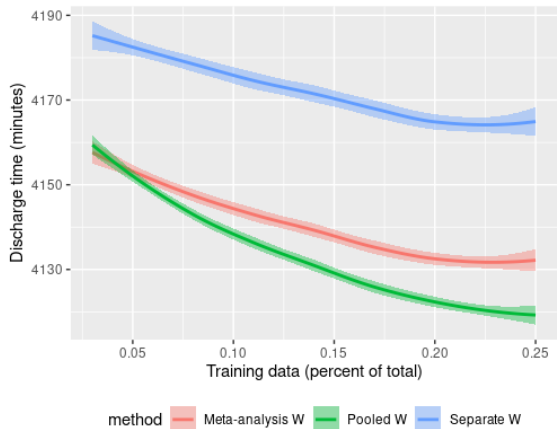


Figure 2: Estimated ECHO study discharge times from using estimated treatment rules (lower is better).

cal pipeline, using energy balancing weights, weighted learning, and adaptive VIC to learn a meta-analysis model. We compare with local analysis, separate weighted learning, and a pooled weighted learning method with access to training samples from all sites. We compare the methodologies on the AIPW estimate of their value functions over the test data. The results are presented in Figure 2. The graph also suggests that our methodology considerably outperforms a locally learned model and results in improved discharge times of about 20-25 minutes compared to local methods and performs comparably to a method that pools all the data.

5. Discussion

In this article, we extended the ITR learning problem to the meta-analysis setting. Our proposed method leverages a directionality principle to borrow strength across multiple sites for learning ITRs by encouraging sign-coherency among site-specific ITRs and

maintains flexibility by allowing them to be different.

In simulation studies, we compared how our method performs against local site-specific models that do not leverage global information and global models with have access to all data but do not tailor decision rules to sites. Even in scenarios where settings favor these idealized models, our method performs comparably or even better. We further demonstrated the usefulness of our method by creating individualized transthoracic echocardiography rules on ICU discharge data from a network of hospitals.

The methodology has several natural extensions worth further investigation. The generalized penalization strategy given in the appendix can be applied to any learning algorithm that relies on optimizing a loss function (e.g. neural nets, support vector machines, etc.). One can also “mix-and-match” estimators within this more general formulation, say if separate sites model their data differently. The general formulation also has an equivalent expression as kernel ridge regression and this can be a convenient way to learn a more flexible treatment rule.

One weakness of the method is that sign-coherency is a difficult property to verify in practice. However, if sign-coherency does not hold, it may be an indication that these separate datasets should not be analyzed jointly to begin with, and one can fall back to the local estimator. Regardless, our approach seems to strike a balance when facing this uncertainty. In the future, we would also be interested in extending the work to the on-line learning regime.

Acknowledgments

JJC was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359). Research reported

in this work was partially funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-2018C2-13180). The views in this work are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

References

- Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.
- Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov. How to backdoor federated learning. In *Artificial Intelligence and Statistics*, pages 2938–2948. PMLR, 2020.
- Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.
- Eli Ben-Michael, Avi Feller, David A Hirshberg, and José R Zubizarreta. The balancing act in causal inference. *arXiv preprint arXiv:2110.14831*, 2021.
- Alina Beygelzimer and John Langford. The offset tree for learning with partial labels. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 129–138, 2009.
- Guanhua Chen, Donglin Zeng, and Michael R Kosorok. Personalized dose finding using outcome weighted

- learning. *J. Amer. Statist. Assoc.*, 111(516):1509–1521, 2016.
- Shuai Chen, Lu Tian, Tianxi Cai, and Menggang Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4): 1199–1209, 2017.
- Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *Am. Econ. Rev.*, 107(5): 261–65, 2017.
- Julien Chiquet, Yves Grandvalet, Camille Charbonnier, et al. Sparsity with sign-coherent groups of variables via the cooperative-lasso. *Ann. Appl. Stat.*, 6(2): 795–830, 2012.
- Yifan Cui and Eric Tchetgen Tchetgen. A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173, 2021.
- Alicia Curth and Mihaela van der Schaar. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 1810–1818. PMLR, 2021.
- Coraline Danieli and Erica EM Moodie. Preserving data privacy when using multi-site data to estimate individualized treatment rules. *Statistics in Medicine*, 41(9):1627–1643, 2022.
- Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott, and Dan Cervone. Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68, 2019.
- European Medicines Agency. Ema regulatory science to 2025–strategic reflection, 2020.
- Qingliang Fan, Yu-Chin Hsu, Robert P Lieli, and Yichong Zhang. Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics*, 40(1):313–327, 2022.
- FDA. Submitting documents using real-world data and real-world evidence to fda for drugs and biologics guidance for industry-may 9, 2019, 2019.
- Mengling Feng, Jakob I McSparron, Dang Trung Kien, David J Stone, David H Roberts, Richard M Schwartzstein, Antoine Vieillard-Baron, and Leo Anthony Celi. Transthoracic echocardiography and mortality in sepsis: analysis of the mimic-iii database. *Intensive Care Med.*, 44(6):884–892, 2018.
- Carlos Fernández-Loría and Foster Provost. Causal decision making and causal effect estimation are not the same... and why it matters. *INFORMS Journal on Data Science*, 2022.
- Robert Flynn, Karin Hedenmalm, Tarita Murray-Thomas, Alexandra Pacurariu, Peter Arlett, Hilary Shepherd, Puja Myles, and Xavier Kurz. Ability of primary care health databases to assess medicinal products discussed by the european union pharmacovigilance risk assessment committee. *Clinical Pharmacology & Therapeutics*, 107(4):957–965, 2020.
- Dylan J Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *arXiv preprint arXiv:1901.09036*, 2019.
- Ankit Garg, Tengyu Ma, and Huy Nguyen. On communication cost of distributed statistical estimation and dimensionality. *Ad-*

- vances in Neural Information Processing Systems*, 27, 2014.
- Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in Neural Information Processing Systems*, 33: 16937–16947, 2020.
- Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiokit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- Jared D Huling and Simon Mak. Energy balancing of covariate distributions. *arXiv preprint arXiv:2004.13962*, 2020.
- Jared D Huling, Maureen A Smith, and Guanhua Chen. A two-part framework for estimating individualized treatment rules from semicontinuous outcomes. *Journal of the American Statistical Association*, 116(533):210–223, 2021.
- Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.
- Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Sci. data*, 3(1):1–9, 2016.
- Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. *Advances in neural information processing systems*, 31, 2018.
- Edward H Kennedy. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020.
- Michael R Kosorok and Eric B Laber. Precision medicine. *Annual review of statistics and its application*, 6:263–286, 2019.
- Owen E Leete, Nathan Kallus, Michael G Hudgens, Sonia Napravnik, and Michael R Kosorok. Balanced policy evaluation and learning for right censored data. *arXiv preprint arXiv:1911.05728*, 2019.
- Eric Mbakop and Max Tabord-Meehan. Model selection for treatment choice: Penalized welfare maximization. *Econometrica*, 89(2):825–848, 2021.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
- Susan A Murphy. Optimal dynamic treatment regimes. *J. R. Stat. Soc. Series B Stat. Methodol.*, 65(2):331–355, 2003.
- Kinuo Nishioka, Tomomi Makimura, Akihiro Ishiguro, Takahiro Nonaka, Mitsune Yamaguchi, and Yoshiaki Uyama. Evolving acceptance and use of rwe for regulatory decision making on the benefit/risk assessment of a drug in Japan. *Clinical Pharmacology & Therapeutics*, 111(1):35–43, 2022.
- Hongming Pu and Bo Zhang. Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *arXiv preprint arXiv:2002.02579*, 2020.
- Zhengling Qi, Dacheng Liu, Haoda Fu, and Yufeng Liu. Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various

- outcomes. *Journal of the American Statistical Association*, 115(530):678–691, 2020.
- Wei Qian, Yi Yang, and Hui Zou. Tweedie’s compound poisson model with grouped elastic net. *J. Comput. Graph. Stat.*, 25(2):606–625, 2016.
- Hongxiang Qiu, Marco Carone, Ekaterina Sadikova, Maria Petukhova, Ronald C Kessler, and Alex Luedtke. Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, 116(533):174–191, 2021.
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.*, 100(469):322–331, 2005.
- Chengchun Shi, Rui Song, and Wenbin Lu. Concordance and value information criteria for optimal treatment decision. *Ann. Stat.*, 49(1):49–75, 2021.
- Ingo Steinwart, Don R Hush, Clint Scovel, et al. Optimal rates for regularized least squares regression. In *Conference on Learning Theory*, pages 79–93, 2009.
- Xiaoqing Tan, Chung-Chou H Chang, Ling Zhou, and Lu Tang. A tree-based model averaging approach for personalized treatment effect estimation from heterogeneous data sources. In *International Conference on Machine Learning*, pages 21013–21036. PMLR, 2022.
- Fei Tang and Hemant Ishwaran. Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(6):363–377, 2017.
- Lu Tian, Ash A Alizadeh, Andrew J Gentles, and Robert Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *J. Amer. Statist. Assoc.*, 109(508):1517–1532, 2014.
- Robert Tibshirani, Jacob Bien, Jerome Friedman, Trevor Hastie, Noah Simon, Jonathan Taylor, and Ryan J Tibshirani. Strong rules for discarding predictors in lasso-type problems. *J. R. Stat. Soc. Series B Stat. Methodol.*, 74(2):245–266, 2012.
- Vladimir Vovk. Kernel ridge regression. In *Empirical inference*, pages 105–116. Springer, 2013.
- Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.
- Hongyi Wang, Kartik Sreenivasan, Shashank Rajput, Harit Vishwakarma, Saurabh Agarwal, Jy-yong Sohn, Kangwook Lee, and Dimitris Papailiopoulos. Attack of the tails: Yes, you really can backdoor federated learning. *Advances in Neural Information Processing Systems*, 33:16070–16084, 2020.
- Bo Zhang, Jordan Weiss, Dylan S Small, and Qingyuan Zhao. Selecting and ranking individualized treatment rules with unmeasured confounding. *Journal of the American Statistical Association*, 116(533):295–308, 2021.
- Yichi Zhang, Eric B Laber, Anastasios Tsiatis, and Marie Davidian. Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904, 2015.

Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.*, 107(499):1106–1118, 2012.

Wenjing Zheng and Mark J van der Laan. Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pages 459–474. Springer, 2011.

Ruoqing Zhu, Ying-Qi Zhao, Guanhua Chen, Shuangge Ma, and Hongyu Zhao. Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, 73(2):391–400, 2017.

Appendix A. Proof of Lemma 1

A similar argument is sketched out in [Chen et al. \(2017\)](#). We give an explicit proof here for completeness. In addition to the three causal assumptions, we will need to assume squared integrability of Y , that is $E[Y^2] < \infty$.

Proof Let $L(\phi, \chi) = \mathbb{E} \left[\frac{(Y-T\phi)^2}{\pi(T, X)} \middle| X = \chi \right]$ be a function of two arguments from $\mathbb{R}^2 \rightarrow \mathbb{R}$. This is well defined by positivity and squared integrability of Y . By the variational Euler-Lagrange equation, $f^{\text{opt}}(x)$ is a stationary trajectory of $\mathbb{E}_X[L(\phi, X)]$ if and only if $\frac{\partial L(\phi, \chi)}{\partial \phi}$ evaluated at $\phi = f^{\text{opt}}(x)$ and $\chi = x$ vanishes for all x . Calculating this derivative, we have that $\frac{\partial L(\phi, \chi)}{\partial \phi}$ is

$$\begin{aligned} &\equiv \lim_{h \rightarrow 0} \frac{L(\phi, \chi) - L(\phi + h, \chi)}{h} \\ &= \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{(Y-T(\phi+h))^2 - (Y-T\phi)^2}{h\pi(T, X)} \middle| X = \chi \right] \\ &= \mathbb{E} \left[\frac{-2YT + 2\phi}{\pi(T, X)} \middle| X = \chi \right]. \end{aligned}$$

Then by Euler-Lagrange, we see that f^{opt} satisfies

$$\begin{aligned} 0 &= \mathbb{E} \left[\frac{-YT + f^{\text{opt}}(x)}{\pi(T, X)} \middle| X = x \right] \\ &= \mathbb{E}_T \left[\mathbb{E} \left[\frac{-YT + f^{\text{opt}}(x)}{\pi(T, X)} \middle| T, X = x \right] \right] \\ &= \mathbb{E}_T \left[\frac{1}{\pi(T, x)} (f^{\text{opt}}(x) - \mathbb{E}[YT|T, X = x]) \right] \\ &= \sum_{t=-1,1} (f^{\text{opt}}(x) - \mathbb{E}[YT|T = t, X = x]) \\ &\quad \times \frac{P(T = t|X = x)}{\pi(t, x)} \\ &= 2f^{\text{opt}}(x) - \mathbb{E}[Y^{(1)}|T = 1, X = x] \\ &\quad + \mathbb{E}[Y^{(-1)}|T = -1, X = x] \quad (\text{SUTVA}) \\ &= 2f^{\text{opt}}(x) - \mathbb{E}[Y^{(1)}|X = x] \\ &\quad + \mathbb{E}[Y^{(-1)}|X = x]. \quad (\text{str. ignor.}) \\ &\implies f^{\text{opt}}(x) = \Delta(x). \end{aligned}$$

Thus, $f^{\text{opt}}(x)$ is an optimal scoring function and $-\text{sign}(f^{\text{opt}}(x))$ is an optimal ITR. \blacksquare

In general, there can be many optimal scoring functions for any joint distribution $X, Y, T \sim \mathbb{P}$. For example, for any transformation h that is positive on $\mathbb{R}_{>0}$ and negative on $\mathbb{R}_{<0}$, $\text{sign}(h(\tau(x)))$ is optimal for the value function when there is no assumption on the class of rules. This lemma gives a means for identifying one such scoring function. We note that restricting the class of scoring functions to be linear does not mean that we assume the underlying CATE is linear.

Choosing the class of scoring functions/decision rules is usually more a question about policy that clinical practitioners are willing to implement, rather than modeling choice. A linear CATE implies that there is a linear scoring function that matches the unconstrained best value function. But there are also other (nonlinear) scoring functions that can also achieve the best value function. In general, a nonlinear CATE means that a

linear scoring function cannot achieve the optimal value function, but the set of best linear decision rules is well defined.

Appendix B. Optimization

In this section, we describe how to solve optimization problems (11) and (12) when using weighted learning or A-learning as the base learners. We will also discuss how to solve them in a distributed setting but to first fix ideas, we first consider solving these without any data-sharing constraints. The main idea of a block coordinate minimization scheme is to take a Newton step in each coordinate block by minimizing the following objective for each coordinate block β_j :

$$\begin{aligned} \min_{\beta_j} L(\beta^{cur}) + U^{cur\top}(\beta - \beta^{cur}) \\ + (\beta - \beta^{cur})^\top \frac{H}{2}(\beta - \beta^{cur}) + \lambda P(\beta), \end{aligned}$$

Here β^{cur} are the current estimates of β , U^{cur} is the gradient vector at the current estimates, $\frac{\partial L(\beta^{cur})}{\partial \beta}$, and H is the Hessian matrix, $\frac{\partial^2 L}{\partial \beta \partial \beta^\top}$, which happens to be fixed when using weighted learning or A-learning as the base learners. Here $\beta = \beta^{cur}$ outside of the j th block. This can be thought of as the second-order Taylor series expansion of the loss around β^{cur} . Since the blocks outside of the j th block are fixed, we can minimize it via the following approximation of the Hessian:

$$\begin{aligned} \min_{\beta_j} L(\beta^{cur}) + U_j^{cur\top}(\beta_j - \beta_j^{cur}) \\ + \frac{\gamma_j}{2} \|\beta_j - \beta_j^{cur}\|^2 + \lambda P_j(\beta_j), \quad (13) \end{aligned}$$

where U_j^{cur} is the gradient with respect to the j th coordinate block evaluated at β^{cur} and γ_j is the largest eigenvalue of the j th block of H . We can compute the gradients $U_{W_j}^{cur} =$

$\frac{\partial L_W(\beta^{cur})}{\partial \beta_j}$ explicitly for weighted learning as

$$\begin{bmatrix} -\frac{2}{n^1} \sum_{i=1}^{n^1} w_i^1 X_{ij}^1 (T_i^1 Y_i^1 - \langle \beta^{1,cur}, X_i^1 \rangle) \\ \vdots \\ -\frac{2}{n^K} \sum_{i=1}^{n^K} w_i^K X_{ij}^K (T_i^K Y_i^K - \langle \beta^{K,cur}, X_i^K \rangle) \end{bmatrix} \quad (14)$$

where the weights correspond to either the inverse propensity or energy balancing weights discussed in Appendix E. The corresponding gradient, $U_{A_j}^{cur} = \frac{\partial L_A(\beta^{cur})}{\partial \beta_j}$, for the A-learning method is

$$\begin{bmatrix} -\frac{2}{n^1} \sum_{i=1}^{n^1} v_i^1 X_{ij}^1 (Y_i^1 - v_i^1 \langle \beta^{1,cur}, X_i^1 \rangle) \\ \vdots \\ -\frac{2}{n^K} \sum_{i=1}^{n^K} v_i^K X_{ij}^K (Y_i^K - v_i^K \langle \beta^{K,cur}, X_i^K \rangle) \end{bmatrix} \quad (15)$$

where $v_i^k = (T_i^k + 1)/2 - \pi(1, X_i^k)$. The Hessian approximations that we use are $\gamma_{W_j} = \max_{k \in [K]} \{ \frac{1}{n^k} \sum_{i=1}^{n^k} w_i^k (X_{ij}^k)^2 \}$ and $\gamma_{A_j} = \max_{k \in [K]} \{ \frac{1}{n^k} \sum_{i=1}^{n^k} (v_i^k X_{ij}^k)^2 \}$. Based on a similar argument in Qian et al. (2016), minimizing problem (13) yields a decrease in the loss function at each step. By checking the KKT conditions, the solution of problem (13) is

$$\begin{aligned} \beta_j^{k,new} &= \left(\beta_j^{k,cur} - \frac{U_j^{k,cur}}{\gamma_j} \right) \\ &\times \left(1 - \frac{\lambda}{\|\varphi_k(\gamma_j \beta_j^{cur} - U_j^{cur})\|} \right)_+ \end{aligned} \quad (16)$$

$$(17)$$

where $\varphi_m(v)$ returns the component-wise positive or negative part of v , according to the sign of the m th element (if it is positive or negative).

Now we consider the perspective of computing this step in a distributed setting. In our setup, we assume that communication occurs in rounds: within each round, sites can exchange messages with the centralized

coordinating center (spoke and wheel topology) and between each round, sites compute based on the data at the site and information received from previous rounds. For data privacy reasons, we cannot share data or models between sites.

Clearly, the usual strategy (e.g. federated learning) of transferring gradients U_j from sites to servers allows us to compute this step at a central server. However, in the case where sharing of parameters and gradients is sensitive (Wang et al., 2020; Bagdasaryan et al., 2020; Geiping et al., 2020), we note that to compute (17) locally, only the shrinkage term depends on information from other sites. Thus, it is sufficient to compute this step by only transferring $db_j^k \equiv \gamma_j \beta_j^k - u_j^k$ to the central server, which in turn broadcasts the step sizes

$$\left(1 - \frac{\lambda}{\|(\gamma_j \beta_j^{cur} - U_j^{cur})_+\|}\right)_+, \quad (18)$$

$$\left(1 - \frac{\lambda}{\|(\gamma_j \beta_j^{cur} - U_j^{cur})_-\|}\right)_+ \quad (19)$$

back to each site, depending on the sign of the k th bit received. This motivates the algorithm presented in Algorithm 1. This algorithm achieves the optimal communication lower bound of $O(pK)$ bits per iteration as in the homogeneous populations, homogeneous concept case (Garg et al., 2014).

In our implementation, we also use a strong rules approach to discarding predictors (Tibshirani et al., 2012) for efficient optimization, so for each iteration, only a small ‘active’ fraction of the p predictors need to be cycled through. The worst case complexity of using strong rules does not yield any improvements, but in practice, we find that the optimization with strong rules is quite fast. We give the details of the strong rules filtering in the next section.

Compared to a federated learning approach (McMahan et al., 2017), each site

Algorithm 1: Optimization for individual-level meta-analysis

Input: Data from each site
 $\{(Y_i, X_i, T_i)\}_{i \in [n_k]}$ for $k \in [r]$

Output: ITR coefficients from each site
 β^1, \dots, β^K

for Site k in $[K]$ **do**

- Compute site-specific weights $\{w_i\}^k$ and augmentation \hat{a}^k ;
- Initialize $\beta^{k,0}$ to solution to local subproblem L_W^k ;
- Compute Hessian eigenvalues γ_j^k for all $j \in [p]$;
- Broadcast p eigenvalues to central server ;

end

Approximate Hessian for all j by setting $\gamma_j = \max_k \gamma_j^k$;

Broadcast γ_j for $j \in [p]$ to all sites $k \in [K]$;

for $t = 0, 1, \dots, T - 1$ **do**

- for** Feature j in $[p]$ **do**

 - for** Site k in $[K]$ **do**

 - Compute the k th element of $U_{Wj}, u_{Wj}^k = -\frac{2}{n^k} \sum_{i=1}^{n^k} w_i^k X_{ij}^k (T_i^k Y_i^k - \langle \beta^{k,t}, X_i^k \rangle)$;
 - Transmit scalar $db_j^k = \gamma_j \beta_j^k - u_{Wj}^k$ to central server ;

end

At central server, stack $dB_j = [db_j^1 \dots db_j^K]^\top$;

Broadcast step size (19) back to each site k , depending on if db_j^k was positive or negative. ;

- for** Site k in $[K]$ **do**

 - Compute $\beta_j^{k,t+1}$ using equation (17) ;

end

end

/ finished one cycle of coordinates */*

only needs to transfer a scalar during each communication round, instead of transferring gradients. For added security, a drop-in differential privacy method could be used when transmitting that scalar (Abadi et al., 2016).

Appendix C. Strong rules for cooperative LASSO

When computing model coefficients along a solution path of a hyperparameter λ_ℓ for $\ell \in L$, the strong rule strategy (Tibshirani et al., 2012) is to use a heuristic to discard a large amount of predictors at each step. This is not exact, so sometimes a mistake is made and a nonzero predictor is discarded. However, this is so rare that discarding according to the strong rule first and then checking the KKT conditions for mistakes later results in a beneficial computational trade-off. In this section, we use the notation $X = \begin{bmatrix} X^1 & & 0 \\ & \ddots & \\ 0 & & X^k \end{bmatrix}$, $Y = [Y^{1\top} \ \dots \ Y^{K\top}]^\top$, $\beta = [\beta^{1\top} \ \dots \ \beta^{K\top}]^\top$ to match the strong rules literature. We let \mathcal{G}_j be the group of j th coefficient indices from each site. Following a similar derivation as Tibshirani et al. (2012), we start from the KKT conditions of the cooperative LASSO,

$$\langle X, Y - X\beta \rangle = \lambda\theta$$

where θ is a subgradient of the coop norm. We can write out the j th index explicitly as $\langle X_j, Y - X\beta \rangle = \lambda\theta_j$. By Theorem 1 of Chiquet et al. (2012), we have

$$\begin{aligned} \langle x_k, y - X\beta \rangle &= \frac{\lambda\beta_k}{\|\varphi_k(\beta_{\mathcal{G}_j})\|} \\ &\text{for } k \in \mathcal{G}_j \text{ and } \beta_j^k \text{ is nonzero} \\ \|\varphi_k(\langle X_{\mathcal{G}_j}, Y - X\beta \rangle)\| &\leq \lambda \\ &\text{for } k \in \mathcal{G}_j \text{ and } \beta_j^k \text{ is zero} \end{aligned}$$

We let $c_{kj}(\lambda) = \varphi_k(\langle X_{\mathcal{G}_j}, Y - X\hat{\beta}(\lambda) \rangle)$ be a function of λ . The key approximation for strong rules is that we assume that c_{kj} is nonexpansive with respect to the 2-norm:

$$\|c_{kj}(\lambda) - c_{kj}(\tilde{\lambda})\| \leq |\lambda - \tilde{\lambda}|.$$

If we choose our strong rule to be

$$\|\varphi_k(\langle X_{\mathcal{G}_j}, Y - X\hat{\beta}(\lambda_{\ell-1}) \rangle)\| < 2\lambda_\ell - \lambda_{\ell-1}. \quad (20)$$

Then we get

$$\begin{aligned} \|c_{kj}(\lambda_\ell)\| &\leq \|c_{kj}(\lambda_\ell) - c_{kj}(\lambda_{\ell-1})\| \\ &\quad + \|c_{kj}(\lambda_{\ell-1})\| \\ &< (\lambda_{\ell-1} - \lambda_\ell) + (2\lambda_\ell - \lambda_{\ell-1}) \\ &= \lambda_\ell, \end{aligned}$$

implying that $\hat{\beta}_j^k(\lambda_\ell)$ vanishes. In the strict inequality above, we used the nonexpansive assumption for the first term, strong rule for the second term.

Appendix D. Cross-fitting

As we have seen, estimation of the ITR depends on both an estimated propensity score and augmentation function. These need to be estimated from the data, but reusing data for estimating these and for fitting the scoring function can lead to overfitting (Chernozhukov et al., 2017). In order to overcome this limitation, we use a cross-fitting strategy (Zheng and van der Laan, 2011) that decouples the estimation of the ITR from the estimation of the propensity score and augmentation term. Within a given site, samples are split into folds so that the augmentation and propensity functions evaluated on a given sample point were trained on out-of-fold data.

Appendix E. Energy balancing weights

The core reason that the learning problem needs to be cast as a weighted regres-

sion problem is to achieve a valid causal estimate by achieving covariate balance of confounding variables. There are two major approaches to achieving covariate balance. The first is to fit a propensity score and then invert it, as discussed previously. An alternative is to use a weighting method that achieves covariate balance by directly optimizing for balance (Ben-Michael et al., 2021). We can instead minimize $\frac{1}{n} \sum_{i=1}^n w(T_i, X_i)(Y_i - \hat{a}(X_i) - T_i f(X_i))^2$ as the weighted learning loss function, where $w(t, x)$ is a weighting function that achieves approximate covariate balance:

$$\mathbb{E}[w(T, X)g(T, X)] = \mathbb{E}\left[\frac{g(T, X)}{\pi(T, X)}\right] \quad (21)$$

for any integrable function g . In the multi-site setting, achieving balance *within* sites is sufficient for a valid approximation of the multi-site loss function (6). Many approaches attempt to balance covariate distributions by balancing a finite number of moments of the covariates, but this trades careful modeling with careful hyperparameter tuning. The energy balancing weight approach (Huling and Mak, 2020) overcomes this by using a weighted energy distance metric on the distributions to avoid both model specification and hyperparameter tuning, achieving a causal estimate by direct balancing of the distributions of confounders across treatment groups. Fundamentally, the choice of balancing method is contextual, and our implementation provides options for parametric and nonparametric propensity-score fitting and energy balancing weighting.

Appendix F. Adaptive information criteria

The concordance function is a population-level quantity defined for single sites as

$$C(f) = \mathbb{E}[(\Delta(X) - \Delta(X')) \times I(f(X) < f(X'))],$$

where X_i and $X_{i'}$ are two independent copies of the features. The interpretation of concordance is that it measures the compatibility between a scoring function and the actual treatment contrast.

In turn, the (single site) concordance information criterion is defined as $\text{CIC}_{\kappa_n}(\beta) = n\hat{C}(\beta) - \log(n)\|\beta\|_0$, where $\hat{C}(\beta)$ is the augmented inverse probability weighting (AIPW) estimator (Bang and Robins, 2005) of the concordance

$$\sum_{i \neq j} \left\{ [w_i(Y_i - \hat{a}(X_i)) - w_j(Y_j - \hat{a}(X_j))] \times \mathbb{I}(\langle \beta, X_i \rangle > \langle \beta, X_j \rangle) \right\} \frac{1}{n(n-1)}, \quad (22)$$

and the 0-norm $\|\cdot\|_0$ is the cardinality of nonzero elements (Shi et al., 2021). The weights w_i can be any covariate balancing weights as discussed in Appendix E. The form of this information criteria resembles the Bayesian Information Criterion (BIC), with the log-likelihood replaced by estimates of the concordance or value function respectively.

One obstacle in directly using CIC (VIC) is dependence on the choice of outcome standardization. In ordinary practice, researchers often make a choice to rescale or transform variables, for reasons of interpretability or computational stability. From the forms above, we note that multiplying Y by a constant leads to a corresponding multiplication in the fitted augmentation term by the constant in the concordance term in CIC (VIC) while leaving the complexity penalization unchanged. While this is not an issue for the asymptotic theory, this ambiguity in how to scale relative terms in CIC (VIC) leads to ambiguity in model selection in practice.

In order to resolve this ambiguity and make our method robust to the choice of outcome scaling, we develop a data-adaptive

form of CIC (VIC). The intuition behind the approach is that CIC (VIC) defines a single curve in λ . With an auxiliary scaling parameter, we can then define a family of curves, indexed by γ :

$$CIC(\lambda, \gamma) = n\hat{C}(\hat{\beta}_\lambda) - \gamma \log(n) \|\hat{\beta}_\lambda\|_0. \quad (23)$$

A good choice of γ for scaling results in a curve with high discrimination between the best model and other models. In the case where curves are weakly differentiable, the total variation of a curve indicates how ‘sharp’ its derivatives are. Motivated by this, we propose Procedure 2.

Algorithm 2: Adaptive CIC (VIC)

Input: Sequence of models $\hat{\beta}(\lambda)$ for $\lambda \in \Lambda$
Output: Value of λ^{opt} and model $\hat{\beta}(\lambda^{\text{opt}})$
for Sequence $\gamma_{\min} \leq \gamma \leq \gamma_{\max}$ **do**
 Compute $CIC(\lambda, \gamma)$ (or $VIC(\lambda, \gamma)$) according to eq. (23);
 Set $\tilde{CIC}(\lambda, \gamma)$ be the normalized curve;
 Compute $TV(\tilde{CIC}(\lambda, \gamma))$;
end
Pick the curve with the greatest total variation norm;
Pick the value of λ that maximizes this curve;

The interpretation is that the CIC curve with the largest total variation norm, after renormalizing, corresponds to the curve with the best model discrimination. We implement this numerically via first-order finite differences.

Appendix G. Simulation data generation

For the simulations only, we assume larger Y is better. Since the data is generated, we

can directly compare the accuracy of treatment rules estimated to the true sign of the treatment effect, as well as compute a relative value function, which is a fraction of the maximum potential value captured by the treatment rule. Within a single replication, the data within a site is generated by a model and noise according to

$$\begin{aligned} Y_i^k &= \mu^k(X_i^k) + T_i^k \Delta^k(X_i^k) + \varepsilon_i^k \\ &= \underbrace{\langle \alpha^k, X_i^k \rangle}_{\text{linear terms}} + \underbrace{\langle \gamma^k, Z_i^k \rangle}_{\text{quadratic terms}} \\ &\quad + T_i^k \underbrace{\left(\langle \beta^k, X_i^k \rangle + \langle \zeta^k, Z_i^k \rangle \right)}_{\text{treatment effect}} + \varepsilon_i^k. \end{aligned}$$

where X_i^k is standard multivariate normal, $Z_i^k = \text{vec}(X_i^k \otimes X_i^k)$ is the vector of second order terms of covariates, $[\alpha^k]$ and $[\beta^k] \in \mathbb{R}^{(p+1) \times 1}$ are vectors of parameters (including intercepts) for the linear main and interaction effects, respectively, and $\gamma^k, \zeta^k \in \mathbb{R}^{(p+1)^2 \times 1}$ are vectors of parameters for the nonlinear main and interaction effects, respectively. The sparsity of coefficients is set so 15% of the p elements of α and β are nonzero in each replication. The sparsity of coefficients for the nonlinear terms is set so 8% of the $(p+1)^2$ elements are nonzero. The support of these coefficients is drawn uniformly at random for each replication. Each nonzero linear parameter and intercept is the sum of a shared ‘base’ component and a site-specific ‘perturbation’ component, multiplied by a sign:

$$\begin{aligned} \alpha_j^{\{k\}} &= s_{\alpha,j}(\alpha_{j,\text{base}} + \alpha_{j,\text{perturb}}^{\{k\}}), j \in [p] \cup \{0\} \\ \beta_j^{\{k\}} &= s_{\beta,j}(\beta_{j,\text{base}} + \beta_{j,\text{perturb}}^{\{k\}}), j \in [p] \cup \{0\}. \end{aligned}$$

Each sign is a Rademacher random variable and ensures sign-coherency of the linear coefficients between studies. The nonzero base and perturbation coefficients are all drawn

from a log-normal($\mu = 0, \sigma^2 = 1$) distribution in the first scenario, hence a balance between homogeneity and heterogeneity. For the nonlinear coefficients, there are no site-specific perturbation terms, and the nonzero components are drawn from lognormal(0, 1). The error term is generated from a standard normal independently of all other parameters. Each of the 500 replications for each sub-scenario is performed under a fresh set of parameters.

In the second scenario, the site-specific perturbation coefficients for the linear interaction effects are drawn from lognormal(5, 1) while the base coefficients are still drawn from lognormal(0, 1) so the site-specific component is larger than the shared component. In the third scenario, this is flipped with the base coefficients for the linear interaction effects drawn from lognormal(5, 1) while the site-specific perturbation is drawn from lognormal(0, 1) so the between-site heterogeneity is larger. In all sub-scenarios, the test set is 10,000 points generated from the above model, including both potential outcomes $T = 1, T = -1$.

Appendix H. Sensitivity to unmeasured confounders

While we take ignorability as a starting point, sensitivity of the model to unmeasured confounders is an important consideration before real clinical decisions are made. Sensitivity analysis is an important research topic in causal inference. However, most sensitivity analysis focus on the average treatment effect, the ones focus on the ITR are very sparse except Kallus and Zhou (2018) and Zhang et al. (2021). Alternatively, there are some recent attempts to relax this assumption based on instrumental variable framework including Cui and Tchetgen Tchetgen (2021); Qiu et al. (2021); Pu and

Zhang (2020). To understand the sensitivity of meta-analysis to unmeasured confounders, we conducted additional simulations, repeated for 500 replications, where 1/3 of the variables with true effects are missing.

Table 1: First quartile, mean, and third quartile of relative value function compared to the maximum potential outcome over 500 replications for many missing confounders. MA=metaanalysis, BLDR=best linear decision rule.

	1Q	mean	3Q
pooled W	0.3814	0.4554	0.6290
pooled A	0.3682	0.4479	0.6263
separate W	0.4204	0.5735	0.7662
separate A	0.4194	0.5754	0.7673
group A	0.4364	0.5811	0.7813
MA W	0.5586	0.6688	0.8176
MA A	0.5802	0.6824	0.8219
BLDR	0.6878	0.7614	0.8622

All methods are sensitive to unmeasured confounders, but it seems like metaanalysis is the most robust to unmeasured confounding.

Appendix I. Performance under Covariate Shift

Following the same data generation mechanism for $n = 50, p = 100$, and $K = 3$ we performed additional simulations where every covariate in each site are randomly shifted up to one standard deviation compared to the original data generation process. For 500 replications, we present the 1st quartile, mean, and 3rd quartile of the relative value captured.

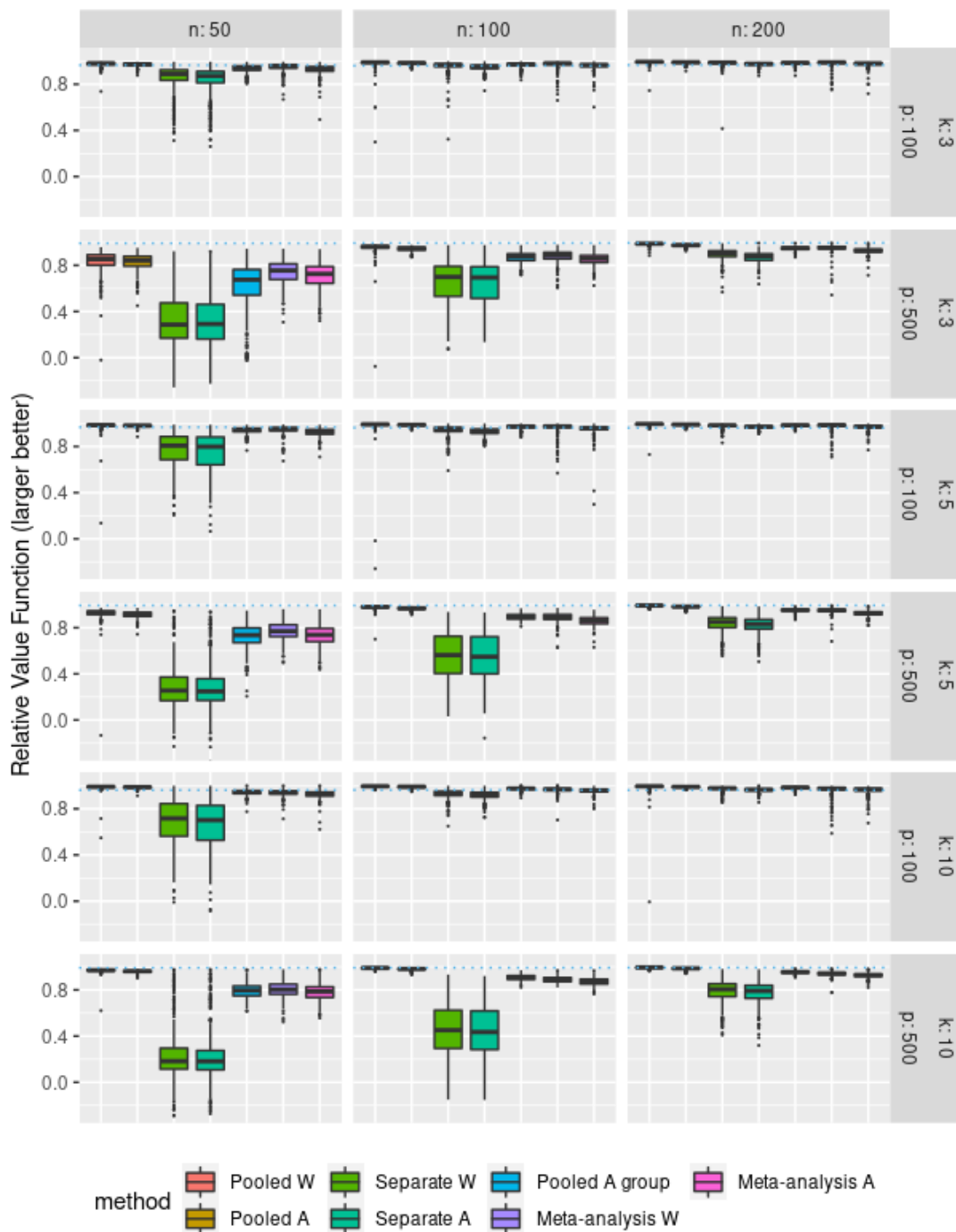


Figure 3: Relative value function under low site heterogeneity.

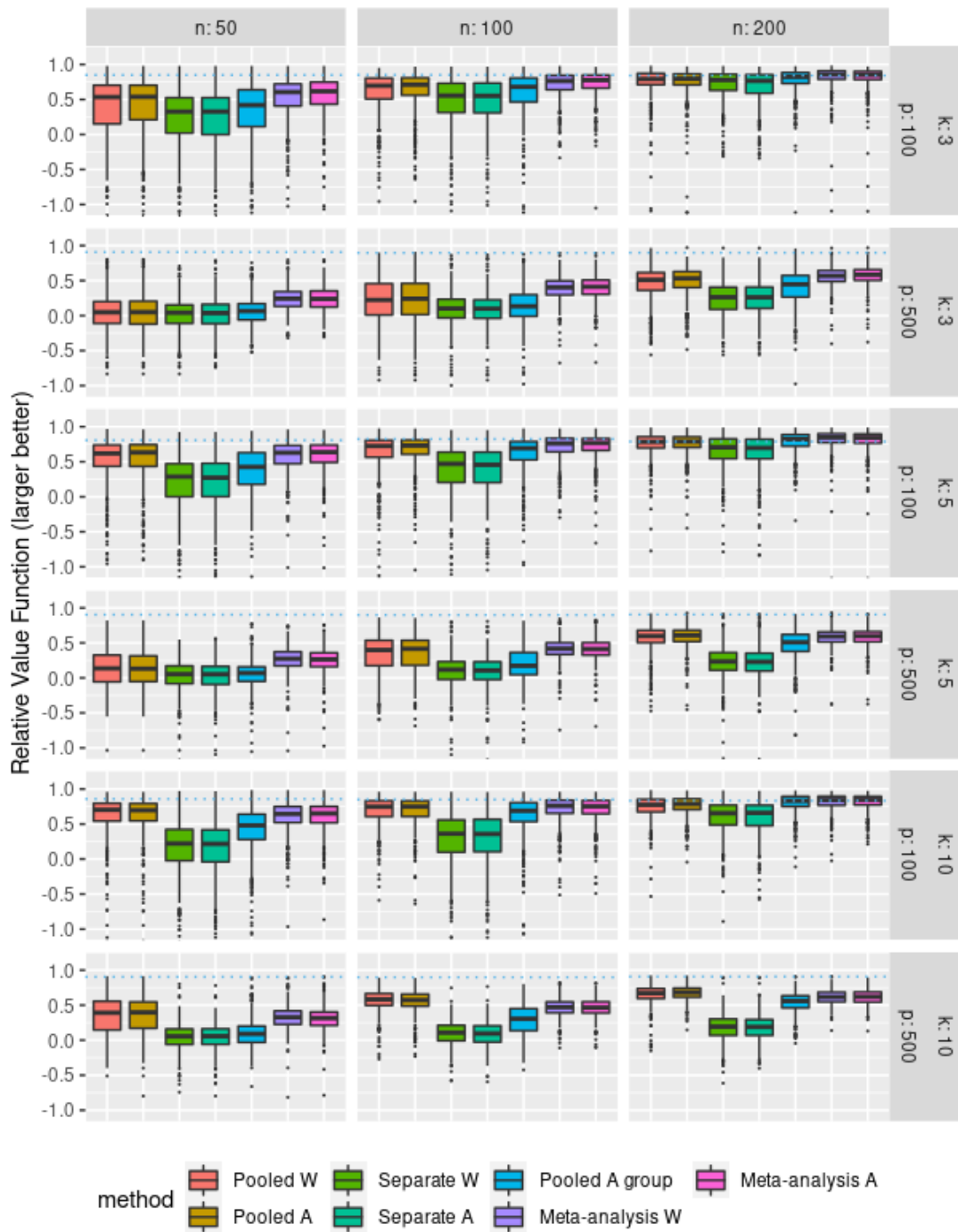


Figure 4: Relative value function under moderate site heterogeneity.

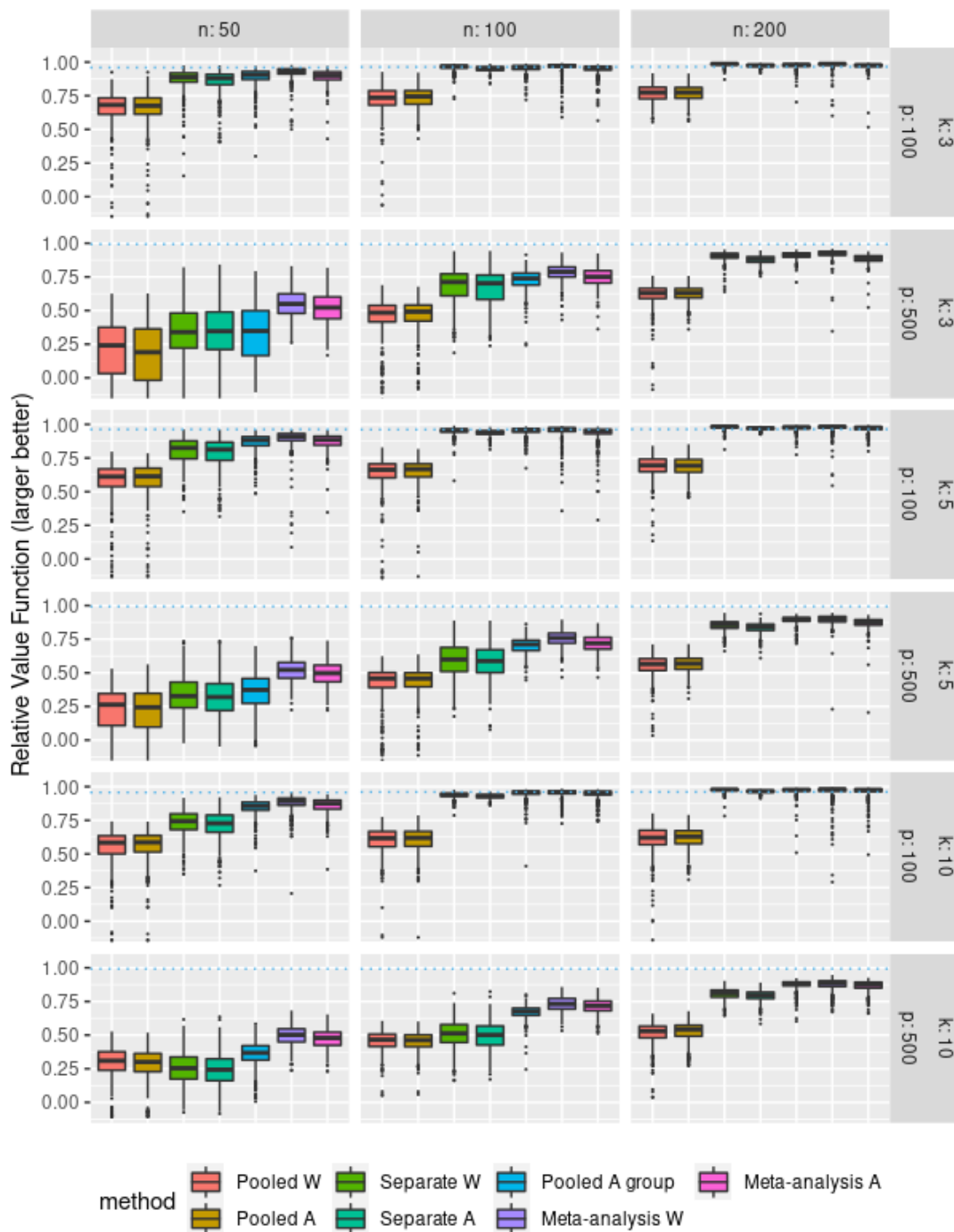


Figure 5: Relative value function under high site heterogeneity.

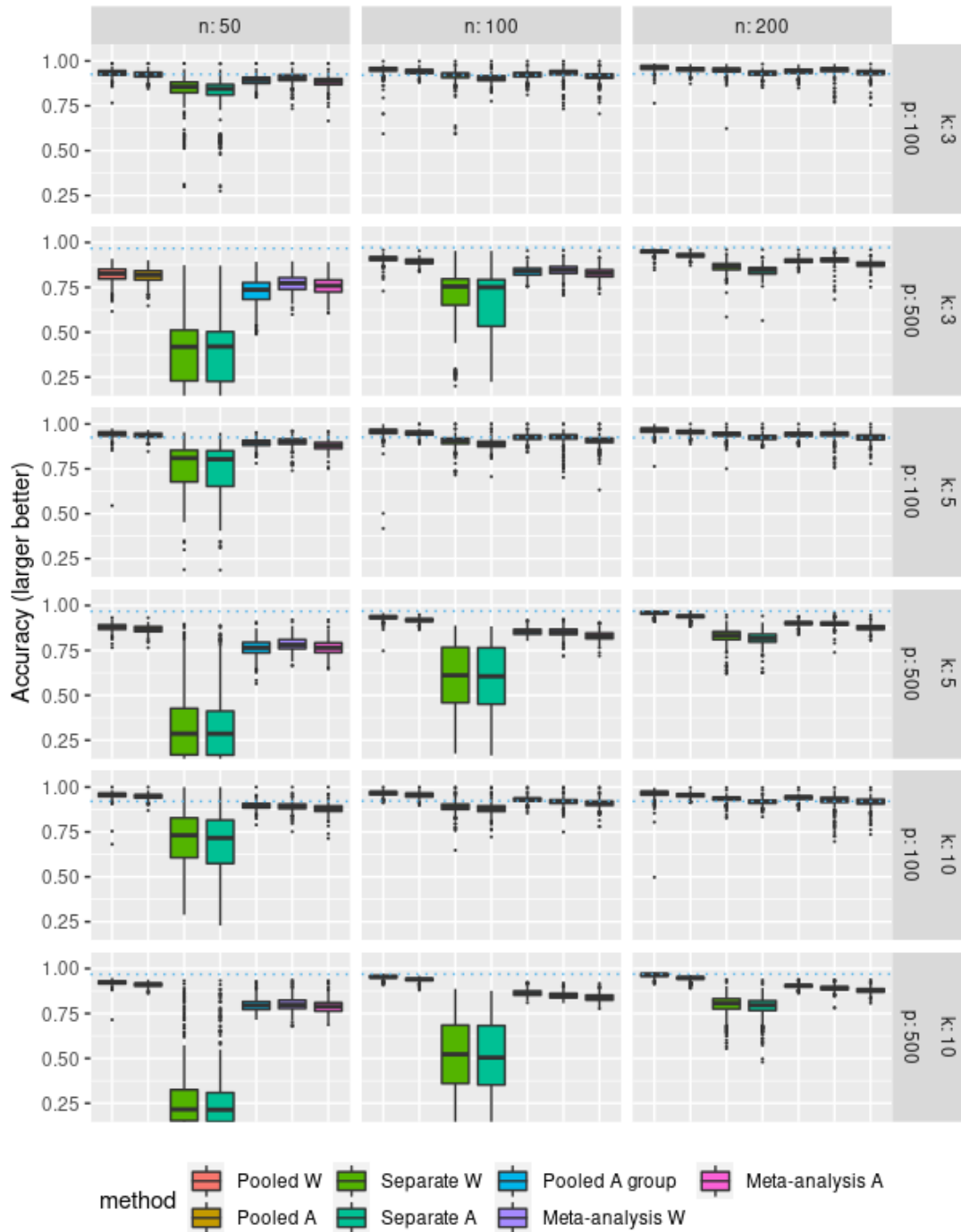


Figure 6: Treatment rule accuracy compared to true potential outcomes under low site heterogeneity.

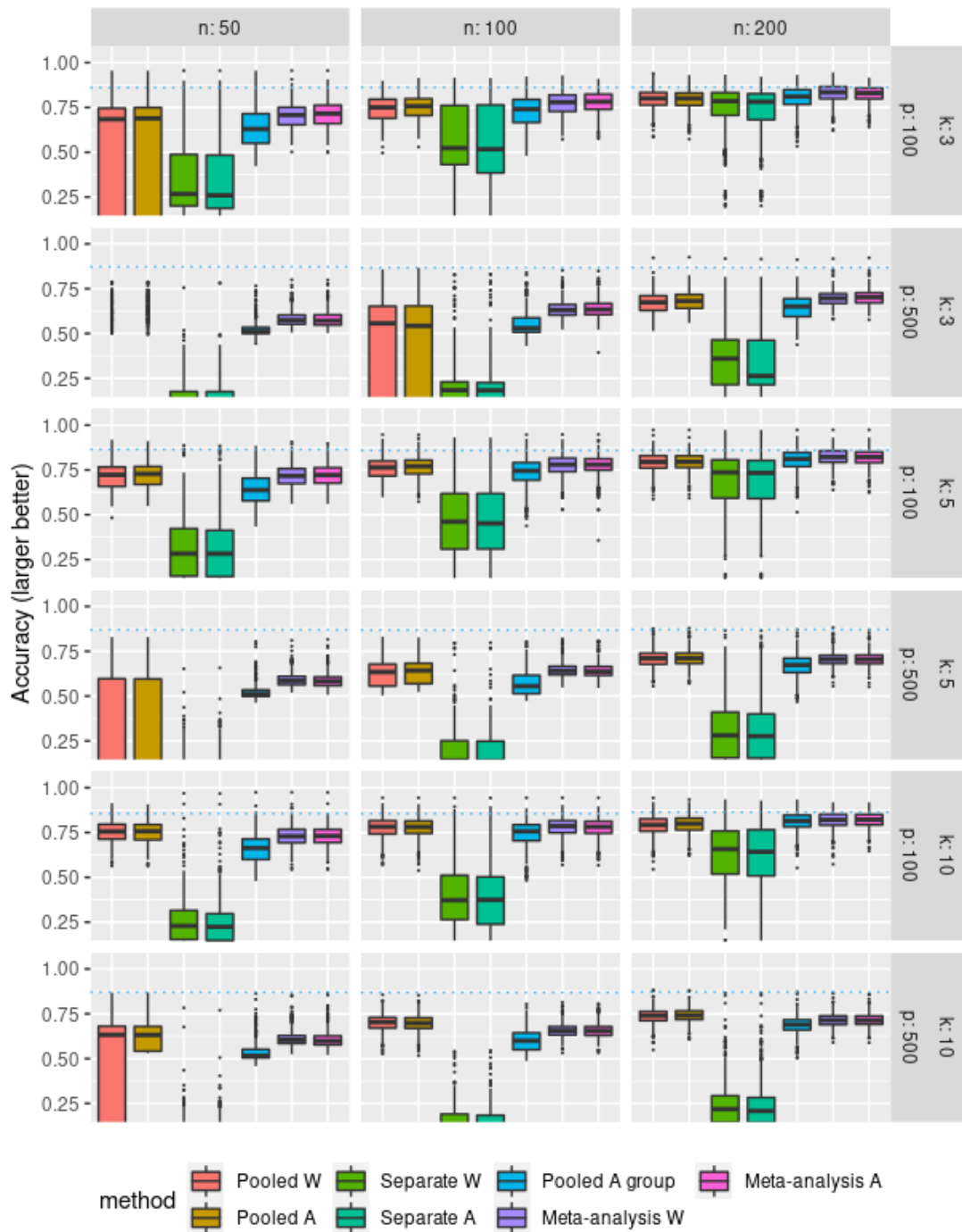


Figure 7: Treatment rule accuracy compared to true potential outcomes under moderate site heterogeneity.

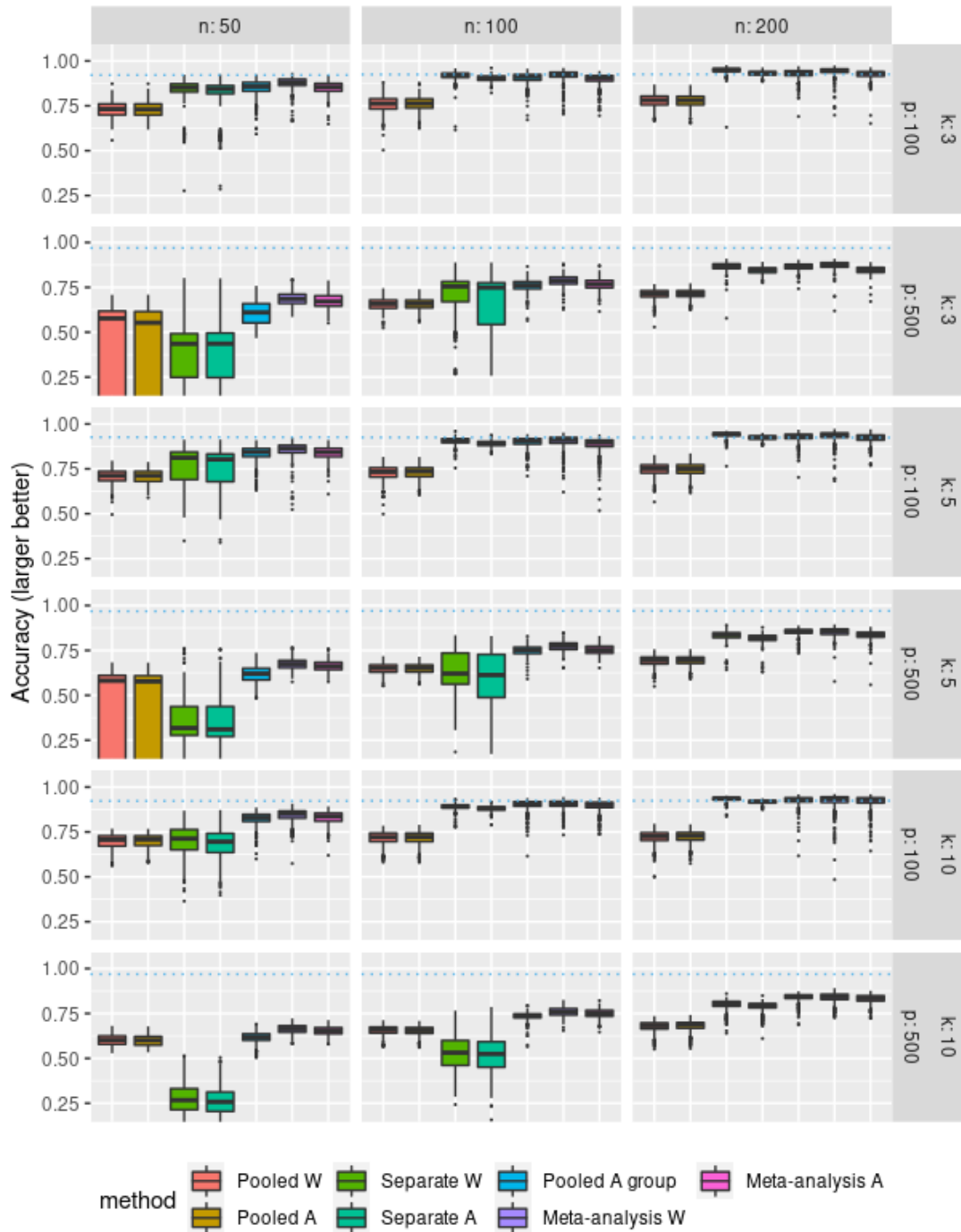


Figure 8: Treatment rule accuracy compared to true potential outcomes under high site heterogeneity.

Table 2: First quartile, mean, and third quartile of relative value function compared to the maximum potential outcome over 500 replications with large covariate drift between sites. MA=metaanalysis, BLDR=best linear decision rule.

	1Q	mean	3Q
pooled W	0.6352	0.6786	0.7629
pooled A	0.6352	0.6845	0.7616
separate W	0.8830	0.9028	0.9355
separate A	0.8631	0.8866	0.9264
group A	0.8720	0.8921	0.9327
MA W	0.9223	0.9360	0.9550
MA A	0.8982	0.9155	0.9396
BLDR	0.9620	0.9607	0.9940

where \hat{S} is an estimate of the survival function and $d\hat{F}$ is an estimate of the density function of \tilde{Y} . We estimate \hat{S} with random survival forests and use the relation $d\hat{F}(t|X, T) = \mathbb{P}(t \leq T < t + dt|X, T)$ to integrate.

Appendix J. Imputation for Censored Discharge Times

Let the true discharge time (in absence of censoring) be denoted \tilde{Y} and the censoring time be denoted C . In this case, the censoring time is the time of death. We use τ to denote the maximum observable time. Let $\delta = \mathbb{I}(\tilde{Y} < C)$ be the indicator that the time we observe is either a censoring time or survival time. The observed time is $Y = \min(\tilde{Y}, C)$, either censored time or survival time. Our setup is that we observe (Y, δ, X, T) where X are the pre-treatment covariates and T is the treatment.

To address the censoring bias, we impute a discharge time to the patients who died. The imputation approach follows the approach taken in [Leete et al. \(2019\)](#). For missing discharge times, we impute it with an estimate of the conditional average discharge time, $\hat{\mathbb{E}}[\tilde{Y}|X, T, Y, \delta = 0]$,

$$\frac{\tau \hat{S}(\tau|X, T) + \int_Y^\tau t d\hat{F}(t|X, T)}{\hat{S}(Y|X, T)},$$