Two aspects of data fusion: Causal survival analysis and Surrogate-assisted conformal inference

Larry Han
Northeastern → Brown

IMSI Workshop: New Horizons on Model Transportability and Data Integration

June 24, 2026

Data fusion: my past and current research

- In multi-$\{$*source*, *institution*, *hospital*$\}$ settings, how can one make optimal use of available data to make causal inferences for a target population of interest?

Data fusion: my past and current research

- In multi-$\{$*source*, *institution*, *hospital*$\}$ settings, how can one make optimal use of available data to make causal inferences for a target population of interest?
  under multiple real-world constraints...

Data fusion: my past and current research

- In multi-{*source*, *institution*, *hospital*} settings, how can one make optimal use of available data to make causal inferences for a target population of interest?
  under multiple real-world constraints...
  - confounding
  - covariate shift
  - varying treatment guidelines
  - varying outcome models
  - privacy constraints
  - communication costs

Data fusion: my past and current research

- In multi-{*source*, *institution*, *hospital*} settings, how can one make optimal use of available data to make causal inferences for a target population of interest?
  under multiple real-world constraints...
  - confounding
  - covariate shift
  - varying treatment guidelines
  - varying outcome models
  - privacy constraints
  - communication costs

  Han, L., Hou, J., Cho, K., Duan, R., & Cai, T. (2025). Federated Adaptive Causal Estimation (FACE) of Target Treatment Effects. *JASA*.

Data fusion: my past and current research

- What if we believe the models are too restrictive?
  - **Han, L.**, Zhu, S., Zubizarreta, J. Multiply robust federated estimation of TATEs. *NeurIPS 2024.*

Data fusion: my past and current research

- What if we believe the models are too restrictive?
  - **Han, L.**, Zhu, S., Zubizarreta, J. Multiply robust federated estimation of TATEs. *NeurIPS 2024.*
- What if we care to profile institutions?
  - **Han, L.**, Li, Y., Niknam, B., Zubizarreta, J. Privacy-preserving, communication-efficient, and target-flexible hospital quality measurement. *AOAS 2024.*

Data fusion: my past and current research

- What if we believe the models are too restrictive?
  - **Han, L.**, Zhu, S., Zubizarreta, J. Multiply robust federated estimation of TATEs. *NeurIPS 2024.*
- What if we care to profile institutions?
  - **Han, L.**, Li, Y., Niknam, B., Zubizarreta, J. Privacy-preserving, communication-efficient, and target-flexible hospital quality measurement. *AOAS 2024.*
- What if we desire distribution-free, model-agnostic, finite-sample valid uncertainty quantification of predictions?
  - Liu, Y., Levis, A., Normand, S-L., **Han, L.** Multi-source conformal inference under distribution shift. *ICML 2024.*

Data fusion: my past and current research

- What if we believe the models are too restrictive?
  - **Han, L.**, Zhu, S., Zubizarreta, J. Multiply robust federated estimation of TATEs. *NeurIPS 2024.*
- What if we care to profile institutions?
  - **Han, L.**, Li, Y., Niknam, B., Zubizarreta, J. Privacy-preserving, communication-efficient, and target-flexible hospital quality measurement. *AOAS 2024.*
- What if we desire distribution-free, model-agnostic, finite-sample valid uncertainty quantification of predictions?
  - Liu, Y., Levis, A., Normand, S-L., **Han, L.** Multi-source conformal inference under distribution shift. *ICML 2024.*
- What if we do not observe outcomes in the target pop?
  - Guo, Z., Li, X., **Han, L.**, Cai, T. Robust inference for federated meta-learning. *JASA 2025.*

Data fusion: my past and current research

- What if we believe the models are too restrictive?
  - **Han, L.**, Zhu, S., Zubizarreta, J. Multiply robust federated estimation of TATEs. *NeurIPS 2024.*
- What if we care to profile institutions?
  - **Han, L.**, Li, Y., Niknam, B., Zubizarreta, J. Privacy-preserving, communication-efficient, and target-flexible hospital quality measurement. *AOAS 2024.*
- What if we desire distribution-free, model-agnostic, finite-sample valid uncertainty quantification of predictions?
  - Liu, Y., Levis, A., Normand, S-L., **Han, L.** Multi-source conformal inference under distribution shift. *ICML 2024.*
- What if we do not observe outcomes in the target pop?
  - Guo, Z., Li, X., **Han, L.**, Cai, T. Robust inference for federated meta-learning. *JASA 2025.*
- What if outcomes are time-to-event?
  - Liu, Y. Levis, A., Zhu, K. Yang, S., Gilbert, P., **Han, L.** Privacy-protected causal survival analysis under distribution shift. *ICLR 2026.*

Targeted Data Fusion for Region-Specific Survival Effects in the AMP HIV Prevention Trials

Larry Han

IMSI Workshop on New Horizons on Model Transportability and Data Integration

Brown University
Department of Biostatistics
Brown Data Science Institute

School of Public Health & BROWN Data Science Institute
Department of Biostatistics

June 24, 2026

Published as a conference paper at ICLR 2026

PRIVACY-PROTECTED CAUSAL SURVIVAL ANALYSIS UNDER DISTRIBUTION SHIFT

Yi Liu[1], **Alexander W. Levis**[2], **Ke Zhu**[1,3], **Shu Yang**[1], **Peter B. Gilbert**[4], **Larry Han**[4,5*]
[1]North Carolina State University [2]University of Pennsylvania [3]Duke University
[4]Fred Hutchinson Cancer Center [5]Northeastern University

Motivating Case Study

The NEW ENGLAND JOURNAL of MEDICINE

SPECIALTIES TOPICS MULTIMEDIA CURRENT ISSUE LEARNING/CME AUTHOR CENTER PUBLICATIONS

ORIGINAL ARTICLE

Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition

Authors: Lawrence Corey, M.D. , Peter B. Gilbert, Ph.D., Michal Juraska, Ph.D., David C. Montefiori, Ph.D., Lynn Morris, Ph.D., Shelly T. Karuna, M.D., Srilatha Edupuganti, M.D., +38 , for the HVTN 704/HPTN 085 and HVTN 703/HPTN 081 Study Teams* Author Info & Affiliations

Published March 17, 2021 | N Engl J Med 2021;384:1003-1014 | DOI: 10.1056/NEJMoa2031738 | VOL. 384 NO. 11

Motivating Case Study

Effects of bnAb in Preventing HIV-1 Acquisition

- Data: Two coordinated randomized trials, HVTN 704/HPTN 085 and HVTN 703/HPTN 081 ($n = 4,611$).
- Data from 4 regions: (i) South Africa; (ii) Other sub-Saharan African countries; (iii) Brazil or Peru; (iv) the United States or Switzerland.
- Target population: can be any one of the above regions (akin to a subgroup analysis).
- Outcome: Being HIV-positive by week-80 (overall incidence rate: 3.77%).

Motivating Case Study

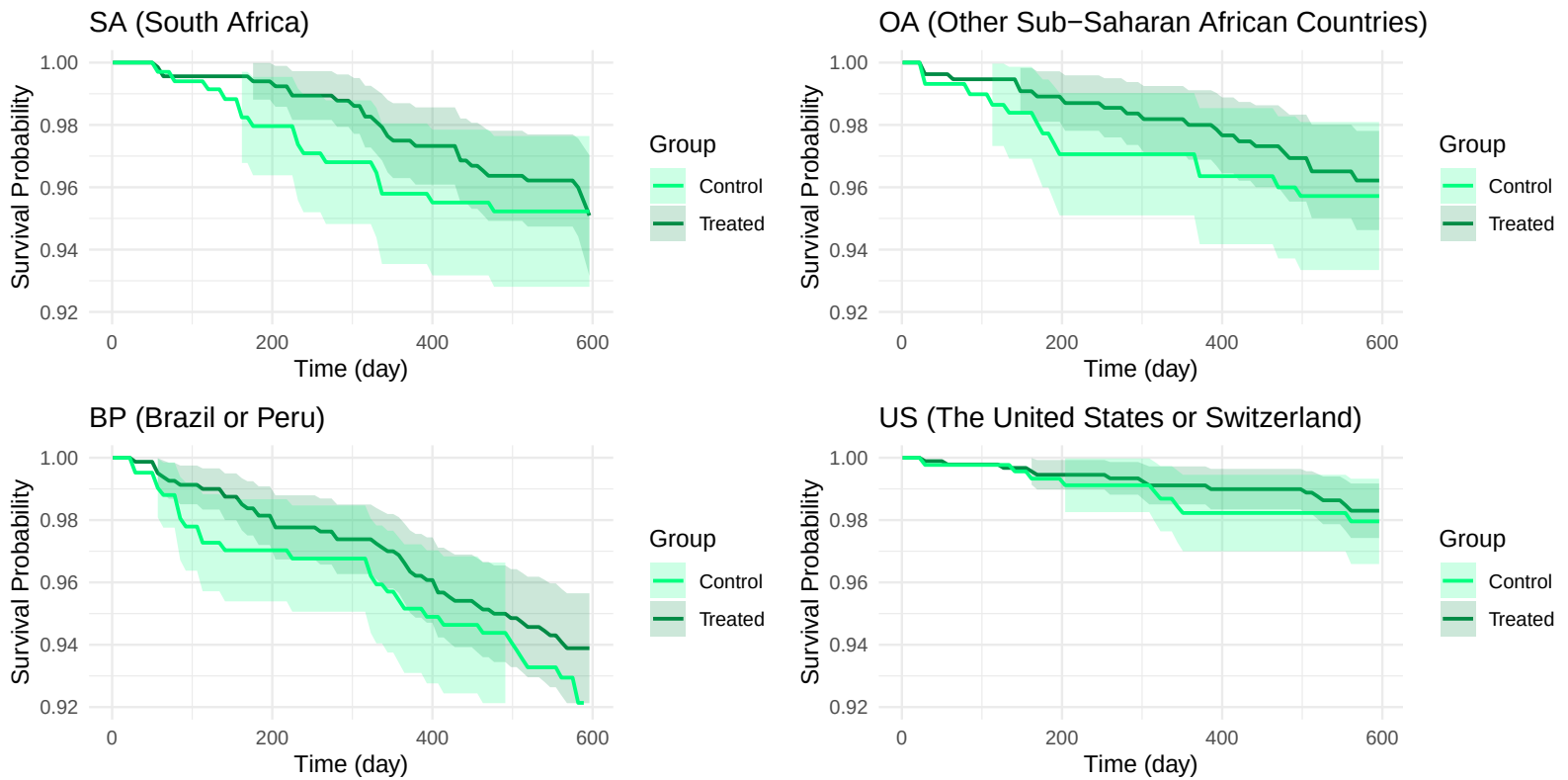


Figure: Site-specific survival curves fitted by target-only data, using their semiparametrically efficient (doubly robust) estimator with some baseline covariates. But for each of these curves, can we also leverage data from other sites (regions) for more accurate estimates $\longrightarrow$ efficiency gain?

Notation

We consider data from K study sites, each of them can be randomized or observational $(K > 0)$.

- $R \in \{0, 1, ..., K-1\}$ denotes study site. $R = 0$ indicates the target, and the rest are source sites.
- Binary treatment $A \in \{0, 1\}$. Potential event and right-censoring times: $T^{(1)}$, $T^{(0)}$, $C^{(1)}$ and $C^{(0)}$.
- Consistency: the event and censoring times of a participant are $T = T^{(A)}$, $C = C^{(A)}$.
- We observe a right-censored outcome $Y = \min(T, C)$ with an event indicator $\Delta = \mathbb{I}(T \leq C)$.
- The observed i.i.d. data: $\{\mathcal{O}_i = (\mathbf{X}_i, A_i, Y_i, \Delta_i, R_i), i = 1, \ldots, n\}$.

Estimand

- We are interested in the target-site treatment-specific survival function:

$$\theta^0(t, a) = \mathbb{P}(T^{(a)} > t \mid R = 0)$$

for $a \in \{0, 1\}$ and $t \in [0, \tau]$ for some positive $\tau < \infty$.

Nuisance Functions

For $k \in \{0, 1, \ldots, K-1\}$, $a \in \{0, 1\}$ and $t \in [0, \tau]$, we write the site- and treatment-specific nuisance functions for time t as follows, conditioning on covariate $\mathbf{X} = \mathbf{x}$:

- Conditional survival function for the event:
  $S^k(t \mid a, \mathbf{x}) = \mathbb{P}(T > t \mid A = a, \mathbf{X} = \mathbf{x}, R = k)$.
- Conditional cumulative hazard function:
  $$\Lambda^k(t \mid a, \mathbf{x}) = \int_0^t \frac{\mathbb{P}(Y \leq t, \Delta = 1 \mid A = a, \mathbf{X} = \mathbf{x}, R = k)}{\mathbb{P}(Y \geq t \mid A = a, \mathbf{X} = \mathbf{x}, R = k)}.$$
- Conditional survival function for the censoring:
  $G^k(t \mid a, \mathbf{x}) = \mathbb{P}(C > t \mid A = a, \mathbf{X} = \mathbf{x}, R = k)$.
- Propensity score for the treatment: $\pi^k(a \mid \mathbf{x}) = \mathbb{P}(A = a \mid \mathbf{X} = \mathbf{x}, R = k)$.

Riemann-Stieltjes Product-integral $\prod(1 + du)$[1]

$\sum$	$\prod$
$\int$	π

Table: A 2×2 table for understanding the nature of the product-integral $\prod$.

- Then, we write the conditional survival function of the event by

$$S^k(t \mid a, \mathbf{x}) = \prod_{(0,t]} \{1 - \Lambda^k(du \mid a, \mathbf{x})\},$$

  where Λ^k is the cumulative hazard function.
- It can be shown that $\prod$ reduces to $\prod$ in case of discrete event time, and to $\exp(-\Lambda^k(t \mid a, \mathbf{x}))$ in case of continuous event time.
- Our framework is designed for unifying both discrete and continuous time settings.

[1] Gill, R. D. & Johansen, S. A survey of product-integration with a view toward application in survival analysis. *The annals of statistics* **18,** 1501–1555 (1990).

Assumptions

For $k \in \{0, 1, \ldots, K-1\}$, $a \in \{0, 1\}$ and $t \in [0, \tau]$, we assume

- Unconfoundedness: $A \perp\!\!\!\perp T^{(a)} \mid \mathbf{X}, R$ and $A \perp\!\!\!\perp C^{(a)} \mid \mathbf{X}, R$.
- Non-informative censoring: $C^{(a)} \perp\!\!\!\perp T^{(a)} \mid A = a, \mathbf{X}, R$.
- Positivity: there exists an $\eta \in (0, \infty)$, such that $\mathbb{P}(R = k) \geq 1/\eta$, and for $\mathbb{P}$-almost all $\mathbf{x}$, $\min \cup_{k=0}^{K-1} \{\pi^k(a \mid \mathbf{x}), G^k(t \mid a, \mathbf{x})\} \geq 1/\eta$ and $\min_k S^k(t \mid a, \mathbf{x}) > 0$.

Efficient estimation by $R = 0$ data

Proposition 1 (Westling et al.[a])

[a] Westling, T. *et al.* Inference for treatment-specific survival curves using machine learning. *Journal of the American Statistical Association,* 1–13 (2023).

If we only use the target site ($R = 0$) data, the semiparametric efficient influence function (EIF) for $\theta^0(t, a)$ is $\varphi^{*0}_{t,a}(\mathcal{O}; S^0, G^0, \pi^0) = \varphi^{*0}_{t,a}(\mathcal{O}; \mathbb{P}) = \varphi^0_{t,a}(\mathcal{O}; \mathbb{P}) - \theta^0(t, a)$, *where* $\varphi^0_{t,a}(\mathcal{O}; \mathbb{P}) =$

$$\frac{\mathbb{I}(R=0)}{\mathbb{P}(R=0)}\left[1 - \frac{\mathbb{I}(A=a)}{\pi^0(a \mid \mathbf{X})}\left\{\frac{\mathbb{I}(Y \le t, \delta = 1)}{S^0(Y \mid a, \mathbf{X})G^0(Y \mid a, \mathbf{X})} - \int_0^{t \wedge Y} \frac{\Lambda^0(du \mid a, \mathbf{X})}{S^0(u \mid a, \mathbf{X})G^0(u \mid a, \mathbf{X})}\right\}\right] S^0(t \mid a, \mathbf{X}).$$

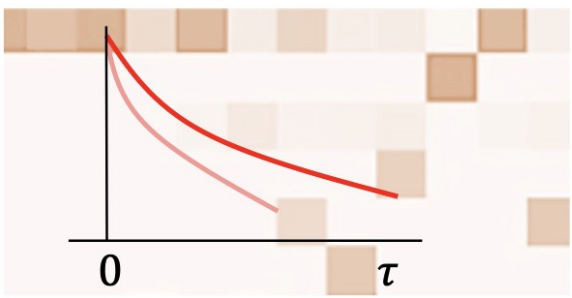
- A doubly robust estimator is given by $\widehat{\theta}^0_n(t, a) = \mathbb{P}_n[\widehat{\varphi}^0_{t,a}(\mathcal{O}; \widehat{\mathbb{P}})]$.

Leveraging Other Data Sources

- Recall the estimand is target-site treatment-specific survival function:

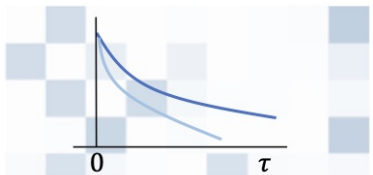
$$\theta^0(t, a) = \mathbb{P}(T^{(a)} > t \mid R = 0)$$

for $a \in \{0, 1\}$ and $t \in [0, \tau]$ for some positive $\tau < \infty$.



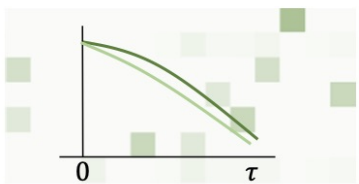
Target Site

$\hat{\theta}^0(t, a)$



Source Site 1

.
.
.



Source Site K

Similar as the target

What are the criteria for integrating these sites?

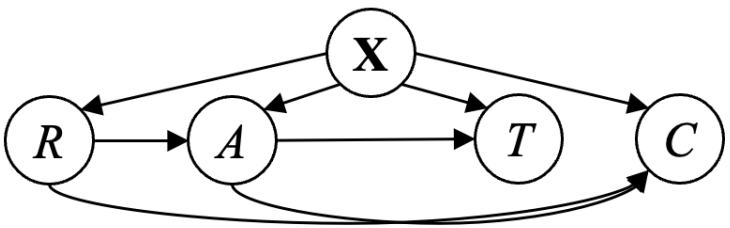
Different to the target

A Key Consideration

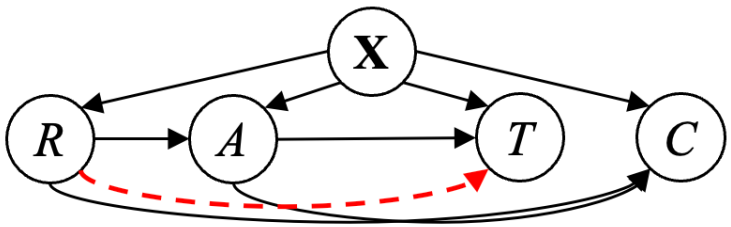
CCOD Assumption

Common conditional outcome distribution (CCOD) across study sites

$$R \perp\!\!\!\perp T^{(a)} \mid \mathbf{X} \text{ for } a \in \{0, 1\}.$$



(a) CCOD holds



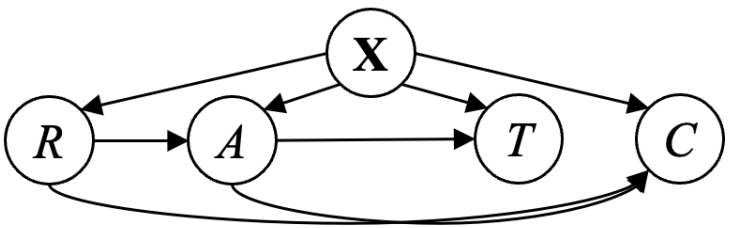
(b) CCOD possibly violated

A Key Consideration

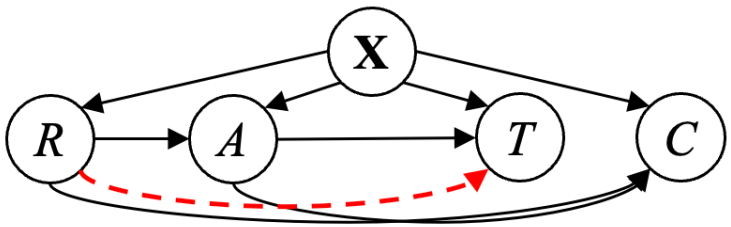
CCOD Assumption

Common conditional outcome distribution (CCOD) across study sites

$$R \perp\!\!\!\perp T^{(a)} \mid \mathbf{X} \text{ for } a \in \{0, 1\}.$$



(a) CCOD holds



(b) CCOD possibly violated

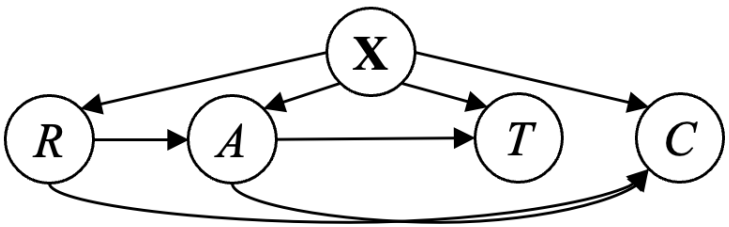
- CCOD implies $S^k(t \mid a, \mathbf{X}) = \mathbb{P}(T^{(a)} > t \mid \mathbf{X}, \cancel{R = k}) := \bar{S}(t \mid a, \mathbf{X})$ has the same form for all k.
- CCOD permits $\mathbf{X}$ shift across sites at any levels.

A Key Consideration

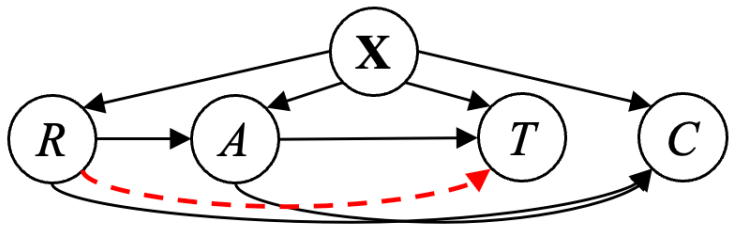
CCOD Assumption

Common conditional outcome distribution (CCOD) across study sites

$$R \perp\!\!\!\perp T^{(a)} \mid \mathbf{X} \text{ for } a \in \{0, 1\}.$$



(a) CCOD holds



(b) CCOD possibly violated

- CCOD implies $S^k(t \mid a, \mathbf{X}) = \mathbb{P}(T^{(a)} > t \mid \mathbf{X}, \cancel{R = k}) := \bar{S}(t \mid a, \mathbf{X})$ has the same form for all k.
- CCOD permits $\mathbf{X}$ shift across sites at any levels.
- In some context, a weaker assumption might be useful:
  - Partial CCOD: $T^{(a)} \mid \mathbf{X}, R = 0 \sim_d T^{(a)} \mid \mathbf{X}, R = k$ for $a \in \{0, 1\}$ and some k.

A Key Consideration

- ✓ CCOD holds: we propose a semiparametrically efficient estimator under target + source data.
- × CCOD does not hold: we propose a data-adaptive strategy to upweight sources providing consistent estimates (efficiency gain) while downweight biased sources (bias reduction).

Method I: CCOD Estimator

- Define some "global" nuisance functions (free of conditioning on R):
  - $\bar{\pi}(a \mid \mathbf{x}) = \mathbb{P}(A = a \mid \mathbf{X} = \mathbf{x})$;
  - $\bar{S}(t \mid a, \mathbf{x}) = \mathbb{P}(T > t \mid A = a, \mathbf{X} = \mathbf{x})$; and
  - $\bar{G}(t \mid a, \mathbf{x}) = \mathbb{P}(C > t \mid A = a, \mathbf{X} = \mathbf{x})$.
- Positivity: There exists an $\eta \in (0, \infty)$, such that for $\mathbb{P}$-almost all $\mathbf{x}$, $\min\{\bar{\pi}(a \mid \mathbf{x}), \bar{G}(t \mid a, \mathbf{x})\} \geq 1/\eta$ and $\bar{S}(t \mid a, \mathbf{x}) > 0$, for all $a \in \{0, 1\}$ and $t \in [0, \tau]$.

Method I: CCOD Estimator

Theorem 1

With source data from other $K-1$ sites, the semiparametric EIF of $\theta^0(t,a)$ given $t \in [0,\tau]$ and $a \in \{0,1\}$ is given by $\varphi^{*CCOD}_{t,a}(\mathcal{O};\bar{S},\bar{G},\bar{\pi}) = \varphi^{*CCOD}_{t,a}(\mathcal{O};\mathbb{P}) = \varphi^{CCOD}_{t,a}(\mathcal{O};\mathbb{P}) - \theta^0(t,a)$, *where* $\varphi^{CCOD}_{t,a}(\mathcal{O};\mathbb{P}) =$

$$\frac{\mathbb{I}(R=0)}{\mathbb{P}(R=0)}\bar{S}(t \mid a,\mathbf{X}) - \frac{q^0(\mathbf{X})}{\mathbb{P}(R=0)}\frac{\mathbb{I}(A=a)}{\bar{\pi}(a \mid \mathbf{X})}\left\{\frac{\mathbb{I}(Y \le t, \Delta=1)}{\bar{S}(Y \mid a,\mathbf{X})\bar{G}(Y \mid a,\mathbf{X})} - \int_0^{t\wedge Y}\frac{\bar{\Lambda}(du \mid a,\mathbf{X})}{\bar{S}(u \mid a,\mathbf{X})\bar{G}(u \mid a,\mathbf{X})}\right\}\bar{S}(t \mid a,\mathbf{X}),$$

where $q^0(\mathbf{X}) = \mathbb{P}(R=0 \mid \mathbf{X})$ *is the target site propensity.*

- Based on the derivation, the terms $q^0(\mathbf{X})$ and $\mathbb{I}(R=0)$ are NOT exchangeable.
- Models for $\bar{S}, \bar{G}, \bar{\pi}$ and q^0 are trained using all data from all sites.

Method I: CCOD Estimator

Based on the EIF, we propose the following CCOD estimator:

$$\widehat{\theta}_n^{\text{CCOD}}(t, a) = \mathbb{P}_n[\varphi_{t,a}^{\text{CCOD}}(\mathcal{O}; \widehat{\mathbb{P}})].$$

Method I: CCOD Estimator

Based on the EIF, we propose the following CCOD estimator:

$$\widehat{\theta}_n^{\text{CCOD}}(t, a) = \mathbb{P}_n[\varphi_{t,a}^{\text{CCOD}}(\mathcal{O}; \widehat{\mathbb{P}})].$$

- Double robustness: when either (i) $\bar{S}$ is correctly specified; or (ii) $\bar{\pi}$, $\bar{G}$ and q^0 are correctly specified, $\widehat{\theta}_n^{\text{CCOD}}(t, a)$ is consistent to $\theta^0(t, a)$.

Method I: CCOD Estimator

Based on the EIF, we propose the following CCOD estimator:

$$\widehat{\theta}_n^{\text{CCOD}}(t,a) = \mathbb{P}_n[\varphi_{t,a}^{\text{CCOD}}(\mathcal{O};\widehat{\mathbb{P}})].$$

- Double robustness: when either (i) $\bar{S}$ is correctly specified; or (ii) $\bar{\pi}$, $\bar{G}$ and q^0 are correctly specified, $\widehat{\theta}_n^{\text{CCOD}}(t,a)$ is consistent to $\theta^0(t,a)$.
- Under some regularity conditions, the CCOD estimator is **uniformly** RAL and achieves the semiparametric efficiency bound.

Method I: CCOD Estimator

Based on the EIF, we propose the following CCOD estimator:

$$\widehat{\theta}_n^{\text{CCOD}}(t,a) = \mathbb{P}_n[\varphi_{t,a}^{\text{CCOD}}(\mathcal{O};\widehat{\mathbb{P}})].$$

- Double robustness: when either (i) $\bar{S}$ is correctly specified; or (ii) $\bar{\pi}$, $\bar{G}$ and q^0 are correctly specified, $\widehat{\theta}_n^{\text{CCOD}}(t,a)$ is consistent to $\theta^0(t,a)$.
- Under some regularity conditions, the CCOD estimator is **uniformly** RAL and achieves the semiparametric efficiency bound.
- The process $\{n^{1/2}(\widehat{\theta}_n^{\text{CCOD}}(u,a) - \theta^0(u,a))\}_{u\in[0,t]}$ converges weakly in $\ell^\infty([0,t])$ to a tight, mean-zero Gaussian process with covariance function $(u,v) \mapsto \mathbb{P}\big(\varphi_{u,a}^{*\text{CCOD}}\varphi_{v,a}^{*\text{CCOD}}\big)$.

Method II: Federated Learning

- It can be too strong to assume CCOD in some scenarios, when the knowledge about the site heterogeneity is limited.
- The CCOD estimator is expected to be biased under CCOD violation.

Method II: Federated Learning

- It can be too strong to assume CCOD in some scenarios, when the knowledge about the site heterogeneity is limited.
- The CCOD estimator is expected to be biased under CCOD violation.
- However, some source sites may still provide relevant information for constructing the target-site survival estimate, while some may not, e.g., under partial CCOD.

Method II – Step 1: Local Estimation

Theorem 2

A semiparametric EIF under partial CCOD between sites $R = 0$ and $R = k$ is given by $\varphi^{*k,0}_{t,a}(\mathcal{O}; S^k, S^0, G^k, \pi^k, \omega^{k,0}) = \varphi^{*k,0}_{t,a}(\mathcal{O}; \mathbb{P}) = \varphi^{k,0}_{t,a}(\mathcal{O}; \mathbb{P}) - \theta^0(t, a)$, *where* $\varphi^{k,0}_{t,a}(\mathcal{O}; \mathbb{P}) =$

$$\frac{\mathbb{I}(R=0)}{\mathbb{P}(R=0)} S^0(t \mid a, \mathbf{X}) - \frac{\mathbb{I}(R=k)}{\mathbb{P}(R=k)} S^k(t \mid a, \mathbf{X}) \omega^{k,0}(\mathbf{X})$$
$$\times \frac{\mathbb{I}(A=a)}{\pi^k(a \mid \mathbf{X})} \left\{ \frac{\mathbb{I}(Y \le t, \Delta = 1)}{S^k(Y \mid a, \mathbf{X}) G^k(Y \mid a, \mathbf{X})} - \int_0^{t \wedge Y} \frac{\Lambda^k(du \mid a, \mathbf{X})}{S^k(u \mid a, \mathbf{X}) G^k(u \mid a, \mathbf{X})} \right\},$$

where $\omega^{k,0}(\mathbf{X}) = \dfrac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}$ *is a density ratio function of covariates* $\mathbf{X}$ *under target site to source site* k.

- Source site estimator for $\theta^0(t, a)$: $\widehat{\theta}^{k,0}_n(t, a) = \mathbb{P}_n[\widehat{\varphi}^{k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}})]$.

Method II – Step 1: Local Estimation

- We impose the partial CCOD between sites 0 and k as a "working assumption" for deriving the form of the EIF and the downstream analysis, even it may be violated.

Method II – Step 1: Local Estimation

- We impose the partial CCOD between sites 0 and k as a "working assumption" for deriving the form of the EIF and the downstream analysis, even it may be violated.
- Looking back to the form of the EIF, the first term uses data from $R = 0$ (target site), while the second term (the augmented term) uses data from $R = k$ (source site).

Method II – Step 1: Local Estimation

- We impose the partial CCOD between sites 0 and k as a "working assumption" for deriving the form of the EIF and the downstream analysis, even it may be violated.
- Looking back to the form of the EIF, the first term uses data from $R = 0$ (target site), while the second term (the augmented term) uses data from $R = k$ (source site).
- The augmented part has a density ratio term $\omega^{k,0}(\mathbf{X}) = \dfrac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}$.

Method II – Step 1: Local Estimation

- We impose the partial CCOD between sites 0 and k as a "working assumption" for deriving the form of the EIF and the downstream analysis, even it may be violated.
- Looking back to the form of the EIF, the first term uses data from $R = 0$ (target site), while the second term (the augmented term) uses data from $R = k$ (source site).
- The augmented part has a density ratio term $\omega^{k,0}(\mathbf{X}) = \dfrac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}$.
  - It serves for projecting site 0 to site k to target $\theta^0(t, a)$ using site k.
  - A quick fact by Bayes's rule:

$$\frac{\mathbb{P}(R = 0 \mid \mathbf{X})}{\mathbb{P}(R = 0)} = \underbrace{\frac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}}_{\omega^{k,0}(\mathbf{X})} \cdot \frac{\mathbb{P}(R = k \mid \mathbf{X})}{\mathbb{P}(R = k)}.$$

Method II – Step 1: Local Estimation

- We impose the partial CCOD between sites 0 and k as a "working assumption" for deriving the form of the EIF and the downstream analysis, even it may be violated.
- Looking back to the form of the EIF, the first term uses data from $R = 0$ (target site), while the second term (the augmented term) uses data from $R = k$ (source site).
- The augmented part has a density ratio term $\omega^{k,0}(\mathbf{X}) = \dfrac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}$.
  - It serves for projecting site 0 to site k to target $\theta^0(t, a)$ using site k.
  - A quick fact by Bayes's rule:

$$\frac{\mathbb{P}(R = 0 \mid \mathbf{X})}{\mathbb{P}(R = 0)} = \underbrace{\frac{\mathbb{P}(\mathbf{X} \mid R = 0)}{\mathbb{P}(\mathbf{X} \mid R = k)}}_{\omega^{k,0}(\mathbf{X})} \cdot \frac{\mathbb{P}(R = k \mid \mathbf{X})}{\mathbb{P}(R = k)}.$$

- But the first term still uses data from the target site, which helps maintaining robustness when using source site data k.

Method II – Step 2: Adaptive Weighting

For a specific $a \in \{0, 1\}$ and $t \in [0, \tau]$, and for $k = 1, \ldots, K-1$:

- Obtain the site-specific discrepancy measure $\widehat{\chi}^{k,0}_{n,t,a} = \mathbb{P}_n \left[\widehat{\varphi}^{k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \widehat{\varphi}^{0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right]$.

Method II – Step 2: Adaptive Weighting

For a specific $a \in \{0, 1\}$ and $t \in [0, \tau]$, and for $k = 1, \ldots, K-1$:

- Obtain the site-specific discrepancy measure $\widehat{\chi}^{k,0}_{n,t,a} = \mathbb{P}_n \left[\widehat{\varphi}^{k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \widehat{\varphi}^{0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right]$.
  - The difference of the two augmented terms of $\varphi^0_{t,a}$ and $\varphi^{k,0}_{t,a}$.
  - Plays a role of testing the partial CCOD.

Method II – Step 2: Adaptive Weighting

For a specific $a \in \{0, 1\}$ and $t \in [0, \tau]$, and for $k = 1, \ldots, K-1$:

- Obtain the site-specific discrepancy measure $\widehat{\chi}^{k,0}_{n,t,a} = \mathbb{P}_n \left[\widehat{\varphi}^{k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \widehat{\varphi}^{0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right]$.
  - The difference of the two augmented terms of $\varphi^0_{t,a}$ and $\varphi^{k,0}_{t,a}$.
  - Plays a role of testing the partial CCOD.
- Solve for $\widehat{\boldsymbol{\eta}}_{t,a} = (\widehat{\eta}^0_{t,a}, \widehat{\eta}^1_{t,a}, \ldots, \widehat{\eta}^{K-1}_{t,a})$ that minimize

$$Q(\boldsymbol{\eta}_{t,a}) = \mathbb{P}_n \left[\left\{ \widehat{\varphi}^{*0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \sum_{k=1}^{K-1} \eta^k_{t,a} \widehat{\varphi}^{*k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right\}^2 \right] + \frac{1}{n} \lambda \sum_{k=1}^{K-1} |\eta^k_{t,a}| (\widehat{\chi}^{k,0}_{n,t,a})^2,$$

subject to $0 \leq \eta^k_{t,a} \leq 1$ and $\sum_{k=0}^{K-1} \eta^k_{t,a} = 1$, and λ is a tuning parameter chosen by cross-validation (akin to an adaptive Lasso regression).

Method II – Step 2: Adaptive Weighting

For a specific $a \in \{0, 1\}$ and $t \in [0, \tau]$, and for $k = 1, \ldots, K-1$:

- Obtain the site-specific discrepancy measure $\widehat{\chi}^{k,0}_{n,t,a} = \mathbb{P}_n \left[\widehat{\varphi}^{k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \widehat{\varphi}^{0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right]$.
  - The difference of the two augmented terms of $\varphi^0_{t,a}$ and $\varphi^{k,0}_{t,a}$.
  - Plays a role of testing the partial CCOD.
- Solve for $\widehat{\boldsymbol{\eta}}_{t,a} = (\widehat{\eta}^0_{t,a}, \widehat{\eta}^1_{t,a}, \ldots, \widehat{\eta}^{K-1}_{t,a})$ that minimize

$$Q(\boldsymbol{\eta}_{t,a}) = \mathbb{P}_n \left[\left\{ \widehat{\varphi}^{*0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) - \sum_{k=1}^{K-1} \eta^k_{t,a} \widehat{\varphi}^{*k,0}_{t,a}(\mathcal{O}; \widehat{\mathbb{P}}) \right\}^2 \right] + \frac{1}{n} \lambda \sum_{k=1}^{K-1} |\eta^k_{t,a}| (\widehat{\chi}^{k,0}_{n,t,a})^2,$$

  subject to $0 \leq \eta^k_{t,a} \leq 1$ and $\sum_{k=0}^{K-1} \eta^k_{t,a} = 1$, and λ is a tuning parameter chosen by cross-validation (akin to an adaptive Lasso regression).
- Obtain

$$\widehat{\theta}^{\text{fed}}_n(t, a) = \sum_{k=0}^{K-1} \widehat{\eta}^k_{t,a} \widehat{\theta}^{k,0}_n(t, a).$$

Method II: Theoretical Results

Theorem 3

Under regularity conditions for local estimates and some oracle selection conditions for source sites, the federated estimator $\widehat{\theta}_n^{fed}(t,a)$ *has the following asymptotic distribution at every* $(t,a) \in [0,\tau] \times \{0,1\}$*,*

$$\sqrt{n/\widehat{\mathcal{V}}_{t,a}}\left\{\widehat{\theta}_n^{fed}(t,a) - \theta^0(t,a)\right\} \to_d \mathcal{N}(0,1),$$

with $\widehat{\mathcal{V}}_{t,a}$ *consistently estimates the asymptotic variance.*

Method II: Theoretical Results

- Let $\mathcal{S}^*_{t,a} = \{k \in \{1, \ldots, K-1\} : \theta^k(t,a) = \theta^0(t,a)\}$ (oracle selection space) and $\mathbb{R}^{\mathcal{S}^*_{t,a}} = \{\boldsymbol{\eta}_{t,a} \in \mathbb{R}^{K-1} : \eta^j_{t,a} = 0, \forall j \notin \mathcal{S}^*_{t,a}\}$ (space of weights).

Method II: Theoretical Results

- Let $\mathcal{S}^*_{t,a} = \{k \in \{1, \ldots, K-1\} : \theta^k(t,a) = \theta^0(t,a)\}$ (oracle selection space) and $\mathbb{R}^{\mathcal{S}^*_{t,a}} = \{\boldsymbol{\eta}_{t,a} \in \mathbb{R}^{K-1} : \eta^j_{t,a} = 0, \forall j \notin \mathcal{S}^*_{t,a}\}$ (space of weights).
- Our federated estimator is equivalent to solving:

$$\bar{\boldsymbol{\eta}}_{t,a} = \underset{\eta^k_{t,a}=0, \forall k \notin \mathcal{S}^*_{t,a}}{\arg\min} \; \mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a}),$$

where $\mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a})$ is the asymptotic variance of the federated estimator under a fixed weight vector $\boldsymbol{\eta}_{t,a}$.

Method II: Theoretical Results

- Let $\mathcal{S}^*_{t,a} = \{k \in \{1, \ldots, K-1\} : \theta^k(t,a) = \theta^0(t,a)\}$ (oracle selection space) and $\mathbb{R}^{\mathcal{S}^*_{t,a}} = \{\boldsymbol{\eta}_{t,a} \in \mathbb{R}^{K-1} : \eta^j_{t,a} = 0, \forall j \notin \mathcal{S}^*_{t,a}\}$ (space of weights).
- Our federated estimator is equivalent to solving:

$$\bar{\boldsymbol{\eta}}_{t,a} = \underset{\eta^k_{t,a}=0, \forall k \notin \mathcal{S}^*_{t,a}}{\arg\min} \; \mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a}),$$

  where $\mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a})$ is the asymptotic variance of the federated estimator under a fixed weight vector $\boldsymbol{\eta}_{t,a}$.

- The federated estimator is a variance minimizer among some estimators, including the target-only estimator as a special case under $\boldsymbol{\eta}_{t,a} = (1, 0, \ldots, 0) \Longrightarrow$ the variance of the target-only estimator can't be smaller than that of the federated estimator.

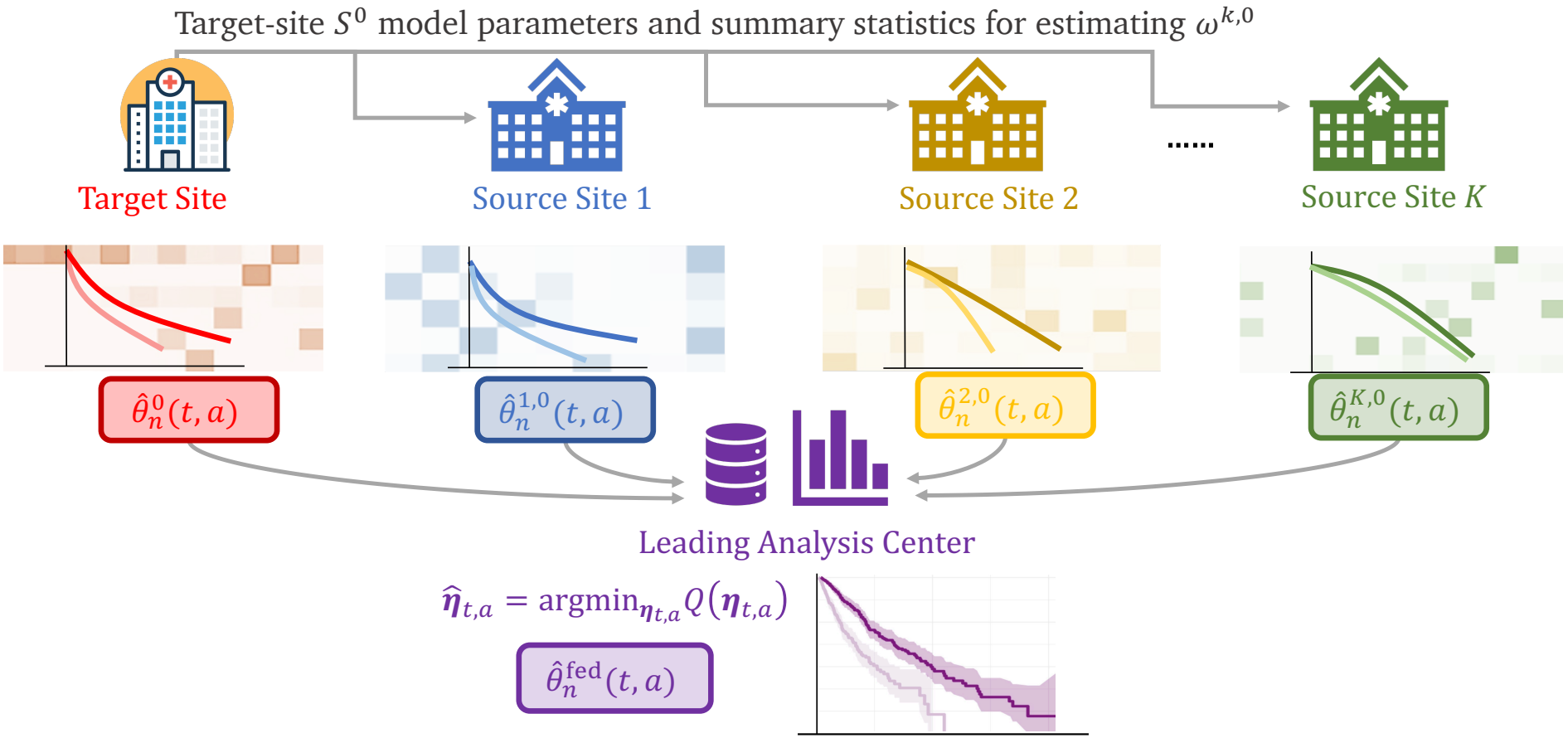
Method II: Theoretical Results

- Let $\mathcal{S}^*_{t,a} = \{k \in \{1, \ldots, K-1\} : \theta^k(t,a) = \theta^0(t,a)\}$ (oracle selection space) and $\mathbb{R}^{\mathcal{S}^*_{t,a}} = \{\boldsymbol{\eta}_{t,a} \in \mathbb{R}^{K-1} : \eta^j_{t,a} = 0, \forall j \notin \mathcal{S}^*_{t,a}\}$ (space of weights).
- Our federated estimator is equivalent to solving:

$$\bar{\boldsymbol{\eta}}_{t,a} = \underset{\eta^k_{t,a}=0, \forall k \notin \mathcal{S}^*_{t,a}}{\arg\min} \; \mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a}),$$

  where $\mathcal{V}_{t,a}(\boldsymbol{\eta}_{t,a})$ is the asymptotic variance of the federated estimator under a fixed weight vector $\boldsymbol{\eta}_{t,a}$.
- The federated estimator is a variance minimizer among some estimators, including the target-only estimator as a special case under $\boldsymbol{\eta}_{t,a} = (1, 0, \ldots, 0) \Longrightarrow$ the variance of the target-only estimator can't be smaller than that of the federated estimator.
- If $\mathcal{S}^*_{t,a} \neq \emptyset$ and regularity conditions for local estimates hold, the efficiency gain is strict: the asymptotic variance of the federated estimator $<$ the target-only estimator.

Method II: Algorithm Summary



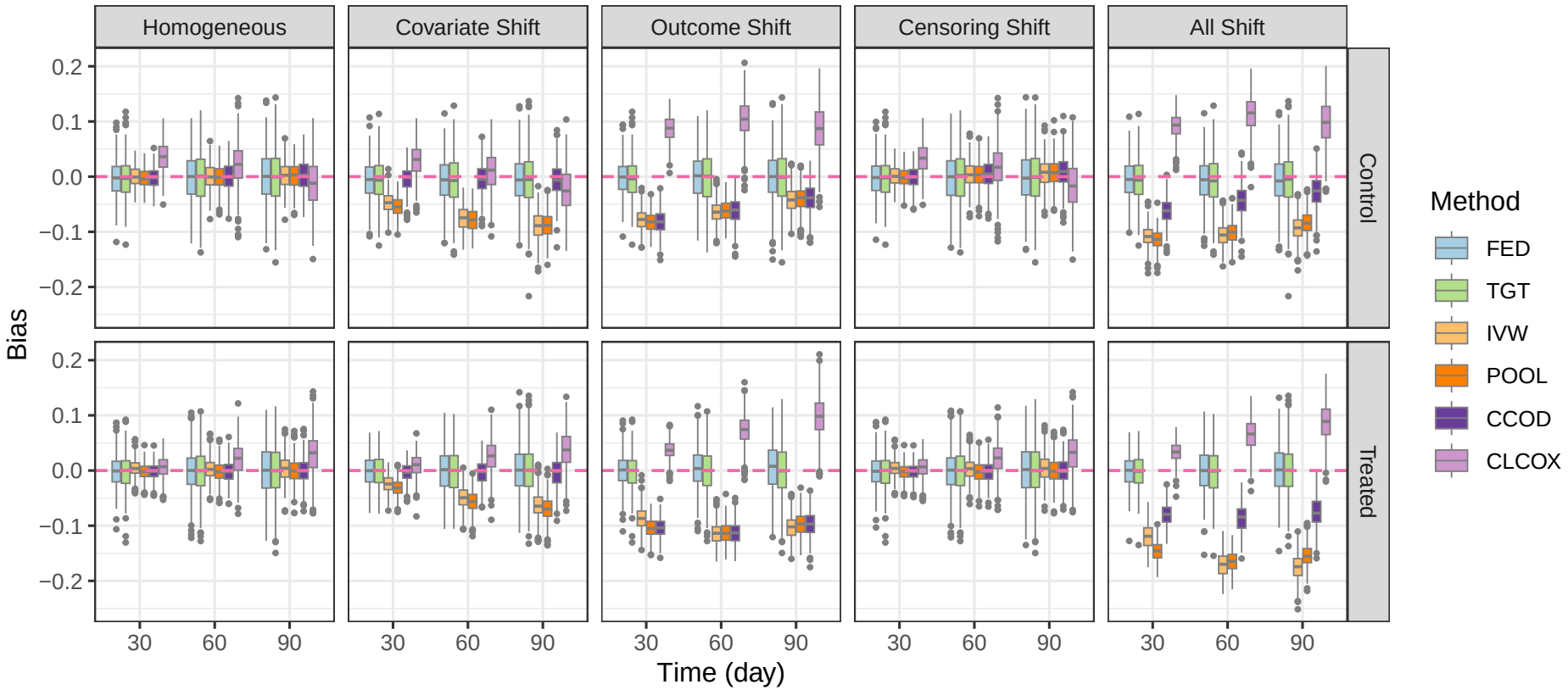
Simulation Study

- We generate data of $K = 5$ sites. We fix the target sample size $n_0 = 300$ and vary $n_k \in \{300, 600, 1000\}$ for $k = 1, \ldots, 4$.
- We model time-to-event outcomes over one year (365 days), truncating censoring at day 200. Results are evaluated at days 30, 60, and 90.

Simulation Study

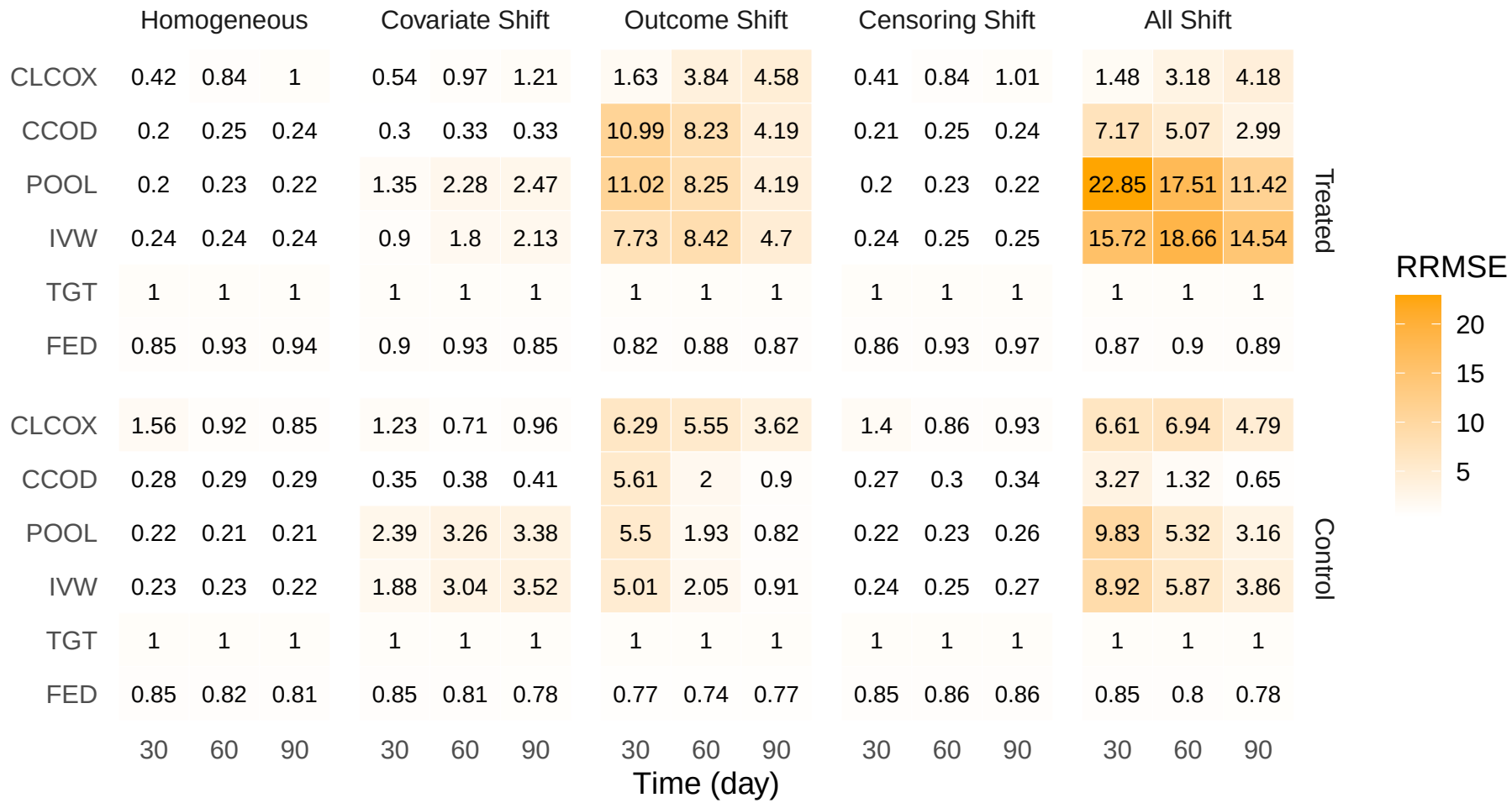
- We generate data of $K = 5$ sites. We fix the target sample size $n_0 = 300$ and vary $n_k \in \{300, 600, 1000\}$ for $k = 1, \ldots, 4$.
- We model time-to-event outcomes over one year (365 days), truncating censoring at day 200. Results are evaluated at days 30, 60, and 90.
- We designed 5 "Distribution Shift" cases across study sites:
  - Homogeneous (CCOD ✓)
  - Covariate shift (CCOD ✓)
  - Outcome shift (CCOD $\times$)
  - Censoring shift (CCOD ✓)
  - All shift (CCOD $\times$)

Simulation Study



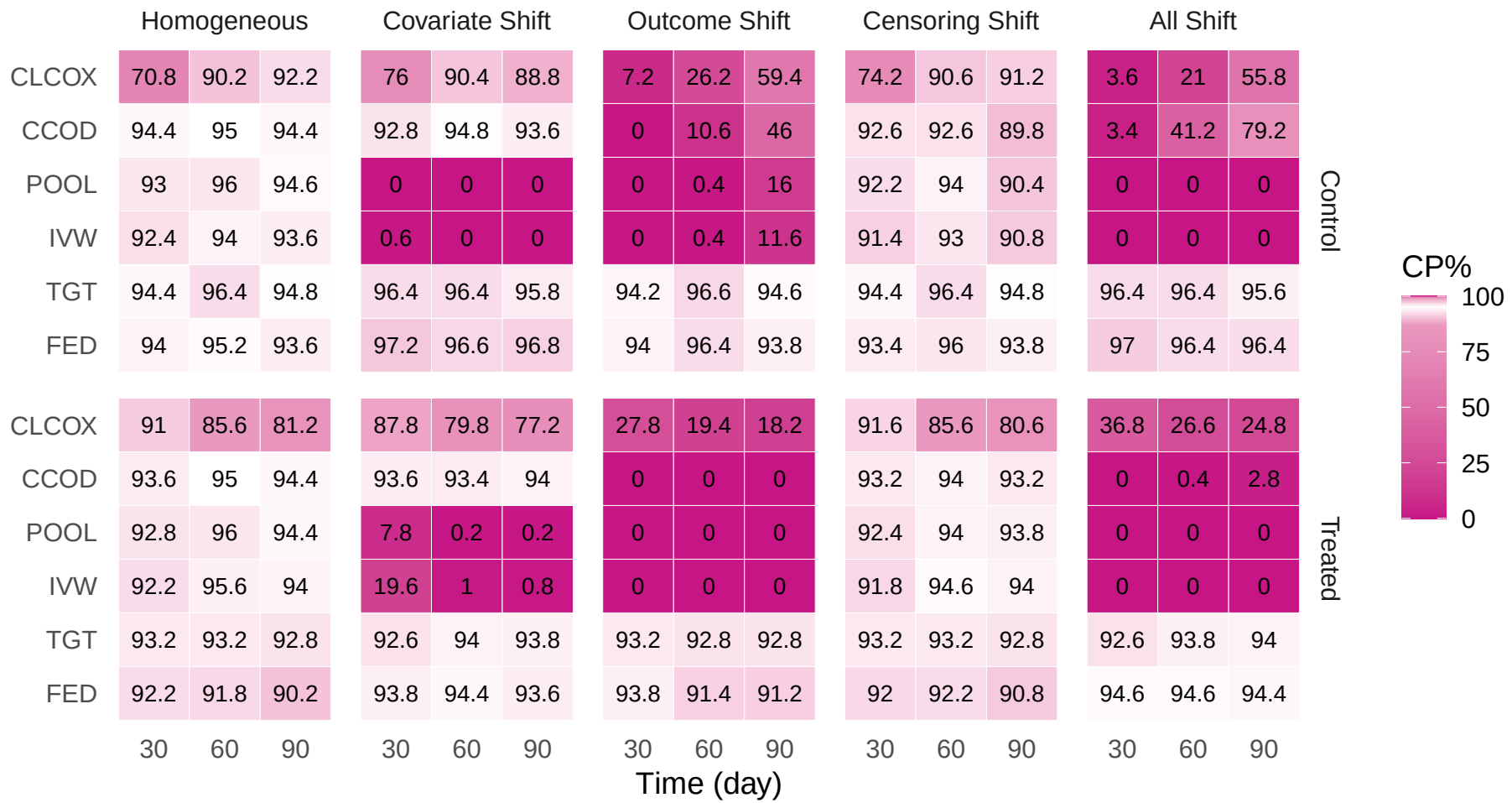
Except for cases where CCOD assumption is violated that the CCOD estimator is biased, the proposed methods (FED and CCOD) have small biases.

Simulation Study



Relative RMSE compared to target-only estimator: FED shows consistently smaller RMSE, with efficiency gains between 6–26%, and CCOD under CCOD assumption has substantially more efficiency gains. Other methods have more efficiency loss.

Simulation Study



FED has overall valid coverage probabilities for 95% CI, confirming the valid inference results.

Revisiting the Case Study

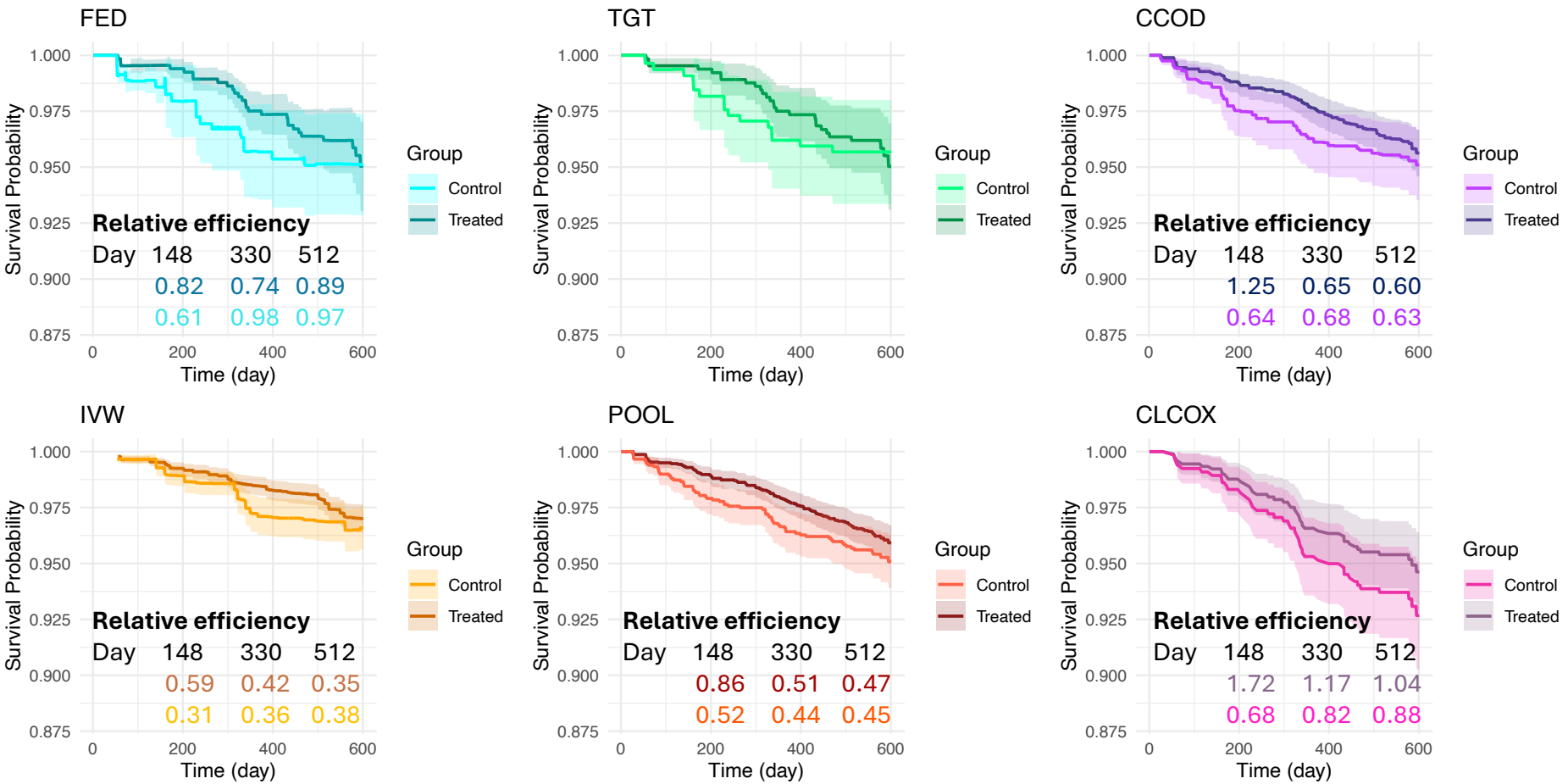


Figure: Results of data fusion when treating **South Africa** as the target. FED: Federated; TGT: target-site-only; IVW: inverse variance weighting; POOL: simple pooling. CLCOX: clustered Cox.

Limitation and Future Work

- The optimality of our federated method is not established
  $\longrightarrow$ covariate-adaptive or other efficient weighting...

Limitation and Future Work

- The optimality of our federated method is not established
  $\longrightarrow$ covariate-adaptive or other efficient weighting...
- We also aim to more efficiently capture temporal trends and appropriately smooth the time-specific federated weights.

Limitation and Future Work

- The optimality of our federated method is not established
  $\longrightarrow$ covariate-adaptive or other efficient weighting...
- We also aim to more efficiently capture temporal trends and appropriately smooth the time-specific federated weights.
- Time-varying covariates (challenging in continuous time setting).

Thank you!

Full manuscript on arXiv

`https://arxiv.org/abs/2501.18798`

"Fundamental problem of causal inference"

- Only one potential outcome is observed for each individual, while the others remain unobserved (Holland, 1986).

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1

Figure: Science table $\neq$ statistician's table (Hernán and Robins, 2020)

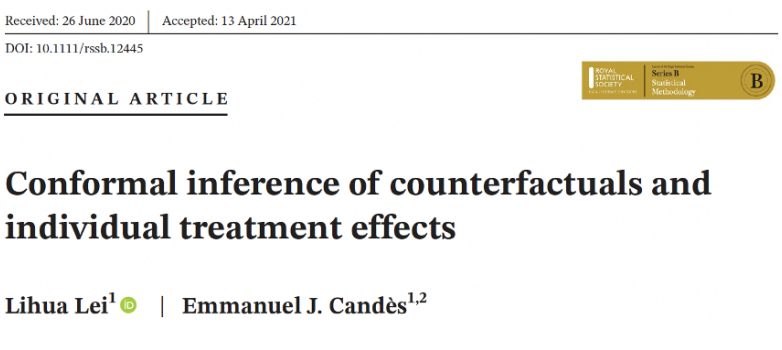
- As a result, causal inference has traditionally focused on estimands like ATE and CATE.
- However, coarsened estimates may fail to capture individual-level heterogeneity.

Conformal counterfactual inference

- Recent advances in conformal inference have shown promise in addressing the challenges of learning ITEs.

Conformal counterfactual inference

- Recent advances in conformal inference have shown promise in addressing the challenges of learning ITEs.
- Lei and Candès (2021) provided the first connection between conformal prediction and causal inference.
- Yang et al. (2024) reformulated the prediction problem under covariate shift into a missing data problem and provided doubly robust and computationally efficient methods leveraging modern semiparametric efficiency theory.



Received: 26 June 2020 | Accepted: 13 April 2021
DOI: 10.1111/rssb.12445

ORIGINAL ARTICLE

Conformal inference of counterfactuals and individual treatment effects

Lihua Lei[1] | Emmanuel J. Candès[1,2]

Journal of the Royal Statistical Society Series B: Statistical Methodology, 2024, **00**, 1–23
https://doi.org/10.1093/jrsssb/qkae009

Original Article

Doubly robust calibration of prediction sets under covariate shift

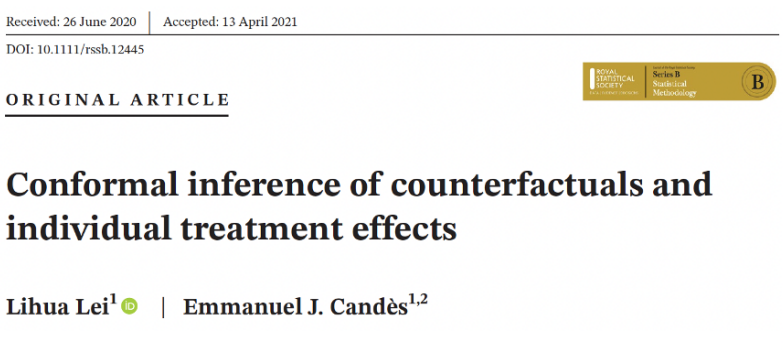
Yachong Yang[1], Arun Kumar Kuchibhotla[2] and Eric Tchetgen Tchetgen[1]

[1]Department of Statistics & Data Science, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA

Address for correspondence: Eric Tchetgen Tchetgen, Department of Statistics & Data Science, University of Pennsylvania, Academic Research Building, 265 S 37th St, Philadelphia, PA, USA. Email: ett@wharton.upenn.edu

Conformal counterfactual inference

- Recent advances in conformal inference have shown promise in addressing the challenges of learning ITEs.
- Lei and Candès (2021) provided the first connection between conformal prediction and causal inference.
- Yang et al. (2024) reformulated the prediction problem under covariate shift into a missing data problem and provided doubly robust and computationally efficient methods leveraging modern semiparametric efficiency theory.

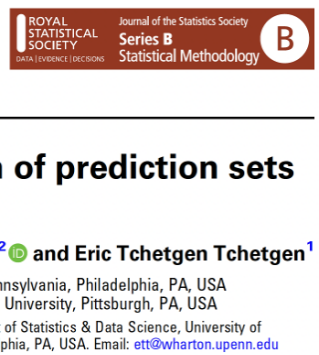


Received: 26 June 2020 | Accepted: 13 April 2021
DOI: 10.1111/rssb.12445

ORIGINAL ARTICLE

Conformal inference of counterfactuals and individual treatment effects

Lihua Lei[1] | Emmanuel J. Candès[1,2]



Journal of the Royal Statistical Society Series B: Statistical Methodology, 2024, **00**, 1–23
https://doi.org/10.1093/jrsssb/qkae009

Original Article

Doubly robust calibration of prediction sets under covariate shift

Yachong Yang[1], Arun Kumar Kuchibhotla[2] and Eric Tchetgen Tchetgen[1]

[1]Department of Statistics & Data Science, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA
Address for correspondence: Eric Tchetgen Tchetgen, Department of Statistics & Data Science, University of Pennsylvania, Academic Research Building, 265 S 37th St, Philadelphia, PA, USA. Email: ett@wharton.upenn.edu

- However, one perhaps unsurprising finding is that prediction intervals produced for ITEs are typically too wide to be useful in practice...

From Lei and Candès (2021)

"The methods with both CQR and BART are conservative."

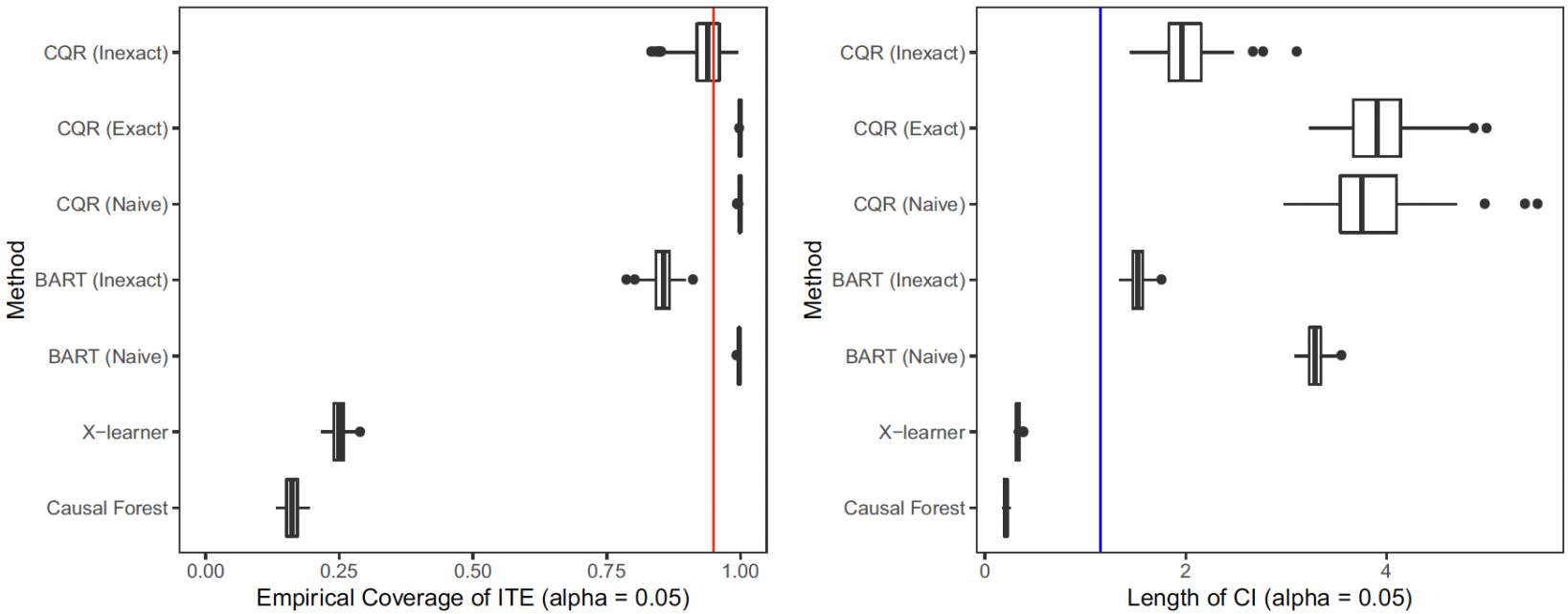


FIGURE 5 Coverage (left) and average length (right) of intervals for ITE on synthetic data generated from the NLSM data. The red vertical line corresponds to the target coverage 95 % and the blue vertical line corresponds to the average length of oracle intervals

Leveraging surrogates for efficiency gain

- Can surrogate outcomes be used to build more efficient prediction intervals for ITE?

Leveraging surrogates for efficiency gain

- Can surrogate outcomes be used to build more efficient prediction intervals for ITE?
- Similar ideas have been explored to enhance the efficiency of ATE on primary outcomes:

Leveraging surrogates for efficiency gain

- Can surrogate outcomes be used to build more efficient prediction intervals for ITE?
- Similar ideas have been explored to enhance the efficiency of ATE on primary outcomes:

Authors	**Paper**	**Year**	**Journal**
S. Athey, R. Chetty, G. W. Imbens, and H. Kang	The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely	2025	The Review of Economic Studies
S. Athey, R. Chetty, and G. Imbens	Combining experimental and observational data to estimate treatment effects on long term outcomes	2020	Arxiv
D. Cheng, A. N. Ananthakrishnan, and T. Cai	Robust and efficient semi-supervised estimation of ATEs with application to EHR data	2021	Biometrics
J. Chen and D. M. Ritzwoller	Semiparametric estimation of long-term treatment effects	2023	Journal of Econometrics
N. Kallus and X. Mao	On the role of surrogates in the efficient estimation of treatment effects with limited outcome data	2024	JRSS-B

  - Methods often require strong statistical surrogacy (Athey et al., 2025)

Related literature on "noisy surrogates"

- Growing literature on using ML predictions as "noisy surrogates" for the true outcome.

Related literature on "noisy surrogates"

- Growing literature on using ML predictions as "noisy surrogates" for the true outcome.

Author	Title	Year	Journal
S. Wang, T.H. McCormick, and J.T. Leek	Methods for correcting inference based on outcomes predicted by machine learning	2020	PNAS
A.N. Angelopoulos, S. Bates, C. Fannjiang, M.I. Jordan, and T. Zrnic	Prediction-powered inference	2023	Science
A.N. Angelopoulos, J.C. Duchi, and T. Zrnic	PPI++: Efficient prediction-powered inference	2023	Arxiv
J. Miao, X. Miao, Y. Wu, J. Zhao, and Q. Lu	Assumption-lean and data-adaptive post-prediction inference	2024	JMLR
K. Motwani & D. Witten	Revisiting inference after prediction	2024	JMLR

- PPI (Angelopoulos et al., 2023) uses a small amount of gold standard labels and a large number of ML predictions as surrogates.
- Extensions follow the general strategy of augmenting an estimator with an unbiased estimator of zero (Bickel et al., 1993, Hahn, 1998, Robins et al., 1994).

Related literature on "noisy surrogates"

- Growing literature on using ML predictions as "noisy surrogates" for the true outcome.

Author	Title	Year	Journal
S. Wang, T.H. McCormick, and J.T. Leek	Methods for correcting inference based on outcomes predicted by machine learning	2020	PNAS
A.N. Angelopoulos, S. Bates, C. Fannjiang, M.I. Jordan, and T. Zrnic	Prediction-powered inference	2023	Science
A.N. Angelopoulos, J.C. Duchi, and T. Zrnic	PPI++: Efficient prediction-powered inference	2023	Arxiv
J. Miao, X. Miao, Y. Wu, J. Zhao, and Q. Lu	Assumption-lean and data-adaptive post-prediction inference	2024	JMLR
K. Motwani & D. Witten	Revisiting inference after prediction	2024	JMLR

- PPI (Angelopoulos et al., 2023) uses a small amount of gold standard labels and a large number of ML predictions as surrogates.
- Extensions follow the general strategy of augmenting an estimator with an unbiased estimator of zero (Bickel et al., 1993, Hahn, 1998, Robins et al., 1994).
- Our work can leverage ML predictions or validated surrogates.

Surrogates for COVID-19

The NEW ENGLAND JOURNAL of MEDICINE

SPECIALTIES TOPICS MULTIMEDIA CURRENT ISSUE LEARNING/CME AUTHOR CENTER PUBLICATIONS

PERSPECTIVE

A Covid-19 Milestone Attained — A Correlate of Protection for Vaccines

Authors: Peter B. Gilbert, Ph.D., Ruben O. Donis, Ph.D., Richard A. Koup, M.D., Youyi Fong, Ph.D., Stanley A. Plotkin, M.D., and Dean Follmann, Ph.D. Author Info & Affiliations

Published December 10, 2022 | N Engl J Med 2022;387:2203-2206 | DOI: 10.1056/NEJMp2211314

VOL. 387 NO. 24 | Copyright © 2022

Science

Current Issue First release papers Archive About Submit ma

HOME > SCIENCE > VOL. 375, NO. 6576 > IMMUNE CORRELATES ANALYSIS OF THE MRNA-1273 COVID-19 VACCINE EFFICACY CLINICAL TRIAL

RESEARCH ARTICLE | CORONAVIRUS

Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial

PETER B. GILBERT, DAVID C. MONTEFIORI, ADRIAN B. MCDERMOTT, YOUYI FONG, DAVID BENKESER, WEIPING DENG, HONGHONG ZHOU, CHRISTOPHER R. HOUCHENS, KAREN MARTINS, [...], AND UNITED STATES GOVERNMENT (USG)/COVPN BIOSTATISTICS TEAM +34 authors Authors Info & Affiliations

SCIENCE • 23 Nov 2021 • Vol 375, Issue 6576 • pp. 43-50 • DOI: 10.1126/science.abm3425

Phase 3 mRNA-1273 (Moderna) COVID-19 VE trial

- Analyzed 1,188 samples: 140 received placebo ($A = 0$) and 1,048 received two-dose Moderna vaccine ($A = 1$)

Phase 3 mRNA-1273 (Moderna) COVID-19 VE trial

- Analyzed 1,188 samples: 140 received placebo ($A = 0$) and 1,048 received two-dose Moderna vaccine ($A = 1$)
- Target ($D = 0$): underrepresented minorities $n_{D0} = 636$ (53.5%) and Source ($D = 1$): non-minorities $n_{D1} = 532$ (46.5%)
- Confounders X: high risk indicator for COVID-19, age-comorbidity stratification variable, and a standardized risk score

Phase 3 mRNA-1273 (Moderna) COVID-19 VE trial

- Analyzed 1,188 samples: 140 received placebo ($A = 0$) and 1,048 received two-dose Moderna vaccine ($A = 1$)
- Target ($D = 0$): underrepresented minorities $n_{D0} = 636$ (53.5%) and Source ($D = 1$): non-minorities $n_{D1} = 532$ (46.5%)
- Confounders X: high risk indicator for COVID-19, age-comorbidity stratification variable, and a standardized risk score
- 2 surrogate markers S: (i) Day 29 pseudo-neutralizing antibody titer (nAB-ID50), and (ii) Day 29 binding spike antibody titer – cheaper
- Primary outcome Y: Day 57 value of the nAB-ID50

Data setup

- Let $D \in \{0, 1\}$ be an indicator of the data origin, with $D = 1$ representing source data and $D = 0$ representing target data.

Data setup

- Let $D \in \{0, 1\}$ be an indicator of the data origin, with $D = 1$ representing source data and $D = 0$ representing target data.
- Each data point $\mathcal{O}_i = (X_i, A_i, S_i, Y_i, D_i)$ is generated by coarsening an i.i.d. draw from a population $\mathcal{O}^* = (X, A, S(0), S(1), Y(0), Y(1), D)$.

Data setup

- Let $D \in \{0, 1\}$ be an indicator of the data origin, with $D = 1$ representing source data and $D = 0$ representing target data.
- Each data point $\mathcal{O}_i = (X_i, A_i, S_i, Y_i, D_i)$ is generated by coarsening an i.i.d. draw from a population $\mathcal{O}^* = (X, A, S(0), S(1), Y(0), Y(1), D)$.

Unit	X	A	S	Y
1	✓	✓	?	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	?	✓
$n_{D1}+1$	✓	✓	?	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	?	?

Table: Semi-supervised setting

Unit	X	A	S	Y
1	✓	✓	✓	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	✓	✓
$n_{D1}+1$	✓	✓	✓	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	✓	?

Table: Surrogate-assisted SS

Unit	X	A	S	Y
1	✓	✓	✓	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	✓	✓
$n_{D1}+1$	✓	✓	?	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	?	?

Table: Intermediate setting

Conformal inference under covariate shift

- Given i.i.d. data (W_i, Y_i) for $i = 1, \cdots, N$, drawn from a source distribution $P_W \otimes P_{Y|W}$, where $W_i = (X_i, S_i) \in \mathcal{W}$ if surrogates S_i are available, or $W_i = X_i$ otherwise, construct

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(Y \in C(W)) \geq 1 - \alpha,$$

where probability is taken over the distribution of the source data and the future observation (W, Y).

Conformal inference under covariate shift

- Given i.i.d. data (W_i, Y_i) for $i = 1, \cdots, N$, drawn from a source distribution $P_W \otimes P_{Y|W}$, where $W_i = (X_i, S_i) \in \mathcal{W}$ if surrogates S_i are available, or $W_i = X_i$ otherwise, construct

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(Y \in C(W)) \geq 1 - \alpha,$$

  where probability is taken over the distribution of the source data and the future observation (W, Y).
- Prediction region: $C(W) = \{y \in \mathcal{Y} \mid R(W, y) \leq r_\alpha\}$, constructed based on a conformity score $R(w, y)$.
  - Ex 1: (Regression residual) $R(w, y) = |y - \widehat{E}(y \mid w)|$
  - Ex 2: (Conformalized quantile residual) $R(w, y) = \max\{\widehat{q}_{\alpha/2}(w) - y, y - \widehat{q}_{1-\alpha/2}(w)\}$, with estimated conditional quantiles (Romano et al., 2019)

Conformal inference under covariate shift

- Given i.i.d. data (W_i, Y_i) for $i = 1, \cdots, N$, drawn from a source distribution $P_W \otimes P_{Y|W}$, where $W_i = (X_i, S_i) \in \mathcal{W}$ if surrogates S_i are available, or $W_i = X_i$ otherwise, construct

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(Y \in C(W)) \geq 1 - \alpha,$$

  where probability is taken over the distribution of the source data and the future observation (W, Y).
- Prediction region: $C(W) = \{y \in \mathcal{Y} \mid R(W, y) \leq r_\alpha\}$, constructed based on a conformity score $R(w, y)$.
  - Ex 1: (Regression residual) $R(w, y) = |y - \widehat{E}(y \mid w)|$
  - Ex 2: (Conformalized quantile residual) $R(w, y) = \max\{\widehat{q}_{\alpha/2}(w) - y, y - \widehat{q}_{1-\alpha/2}(w)\}$, with estimated conditional quantiles (Romano et al., 2019)
- If r_α is the $(1 - \alpha)$-quantile of $R(W, Y)$ for the target distribution,

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(R(W, Y) \leq r_\alpha) \geq 1 - \alpha.$$

Conformal inference under covariate shift

- Given i.i.d. data (W_i, Y_i) for $i = 1, \cdots, N$, drawn from a source distribution $P_W \otimes P_{Y|W}$, where $W_i = (X_i, S_i) \in \mathcal{W}$ if surrogates S_i are available, or $W_i = X_i$ otherwise, construct

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(Y \in C(W)) \geq 1 - \alpha,$$

  where probability is taken over the distribution of the source data and the future observation (W, Y).
- Prediction region: $C(W) = \{y \in \mathcal{Y} \mid R(W, y) \leq r_\alpha\}$, constructed based on a conformity score $R(w, y)$.
  - Ex 1: (Regression residual) $R(w, y) = |y - \widehat{E}(y \mid w)|$
  - Ex 2: (Conformalized quantile residual) $R(w, y) = \max\{\widehat{q}_{\alpha/2}(w) - y, y - \widehat{q}_{1-\alpha/2}(w)\}$, with estimated conditional quantiles (Romano et al., 2019)
- If r_α is the $(1 - \alpha)$-quantile of $R(W, Y)$ for the target distribution,

$$P_{(W,Y)\sim Q_W \otimes P_{Y|W}}(R(W, Y) \leq r_\alpha) \geq 1 - \alpha.$$

- Goal: estimate r_α from the source data following $P_W \otimes P_{Y|W}$ while $(W, Y) \sim Q_W \otimes P_{Y|W}$.

Reformulation for individual causal effects

Assumption 1: Marginal contrast

Assume the individual causal estimand $\theta_i = f(Y_i(1), Y_i(0))$ can be completely characterized as a pre-specified transformation of the potential outcomes $Y_i(1)$ and $Y_i(0)$ separately:

$$\theta_i = u\{Y_i(1)\} - u\{Y_i(0)\}.$$

- Ex 1: ITE $Y_i(1) - Y_i(0)$
- Ex 2: Risk ratio under the logarithmic scale $\log(Y_i(1)) - \log(Y_i(0))$ for continuous outcomes
- Ex 3: For ordinal outcomes, transformations $u(\cdot)$ can correspond to utility scores

Reformulation for individual causal effects

Construction of prediction regions $C_\theta(W_i)$ for θ_i can be reformulated:

$$P\{\theta_i \in C_\theta(W_i)\} = P(D_i = 1)P(\theta_i \in C_\theta(W_i) \mid D_i = 1) \quad (1)$$
$$+ P(D_i = 0)P(\theta_i \in C_\theta(W_i) \mid D_i = 0). \quad (2)$$

Conformal inference on the source data

In source data $(D = 1)$, one of the potential primary outcomes $Y(A)$ is observed $\implies$ only need to construct the prediction regions for $u\{Y(0)\}$ when $A = 1$ and for $u\{Y(1)\}$ when $A = 0$:

Conformal inference on the source data

In source data $(D = 1)$, one of the potential primary outcomes $Y(A)$ is observed $\implies$ only need to construct the prediction regions for $u\{Y(0)\}$ when $A = 1$ and for $u\{Y(1)\}$ when $A = 0$:

$$
\begin{aligned}
&P(\theta_i \in C_\theta(W_i) \mid D_i = 1) \\
&= P(A_i = 1 \mid D_i = 1) P\{u\{Y_i(0)\} \in C_0(W_i) \mid A_i = 1, D_i = 1\} \\
&\quad + P(A_i = 0 \mid D_i = 1) P\{u\{Y_i(1)\} \in C_1(W_i) \mid A_i = 0, D_i = 1\}
\end{aligned}
$$

Reformulation for individual causal effects

In source data ($D = 1$), one of the potential primary outcomes $Y(A)$ is observed $\implies$ only need to construct the prediction regions for $u\{Y(0)\}$ when $A = 1$ and for $u\{Y(1)\}$ when $A = 0$:

$$
\begin{aligned}
&P(\theta_i \in C_\theta(W_i) \mid D_i = 1) \\
&= P(A_i = 1 \mid D_i = 1) P\{R(W, u\{Y_i(0)\}) \leq r_{\alpha,0} \mid A_i = 1, D_i = 1\} \\
&\quad + P(A_i = 0 \mid D_i = 1) P\{R(W, u\{Y_i(1)\}) \leq r_{\alpha,1} \mid A_i = 0, D_i = 1\},
\end{aligned}
$$

where $r_{\alpha,a} = \inf\{r : F_R(r \mid A = 1 - a, D = 1) \geq 1 - \alpha\}$ is the $(1 - \alpha)$-quantile of the potential conformity score $R(W, u\{Y(a)\})$ conditional on $A = 1 - a$ and $D = 1$.

Identification assumptions

Assumption 2: Overlap

There exists some constant $0 < c_0 < 1/2$ such that $c_0 \leq P(A = 1 \mid x) \leq 1 - c_0$, $c_0 \leq P(D = 1 \mid x, a) \leq 1 - c_0$ for $a = 0, 1$ and any x such that $f(x) > 0$.

Assumption 3: Unconfoundedness

$A \perp \{Y(a), S(a)\} \mid X$ for $a = 0, 1$.

Assumption 4: Surrogate exchangeability ▸ See slide 36

$D \perp S(a) \mid X, A$ for $a = 0, 1$.

Assumption 5: MAR ▸ See slide 37

$D \perp Y(a) \mid S(a), X, A$ for $a = 0, 1$.

▸ See sensitivity analyses, slide 38

SWIG

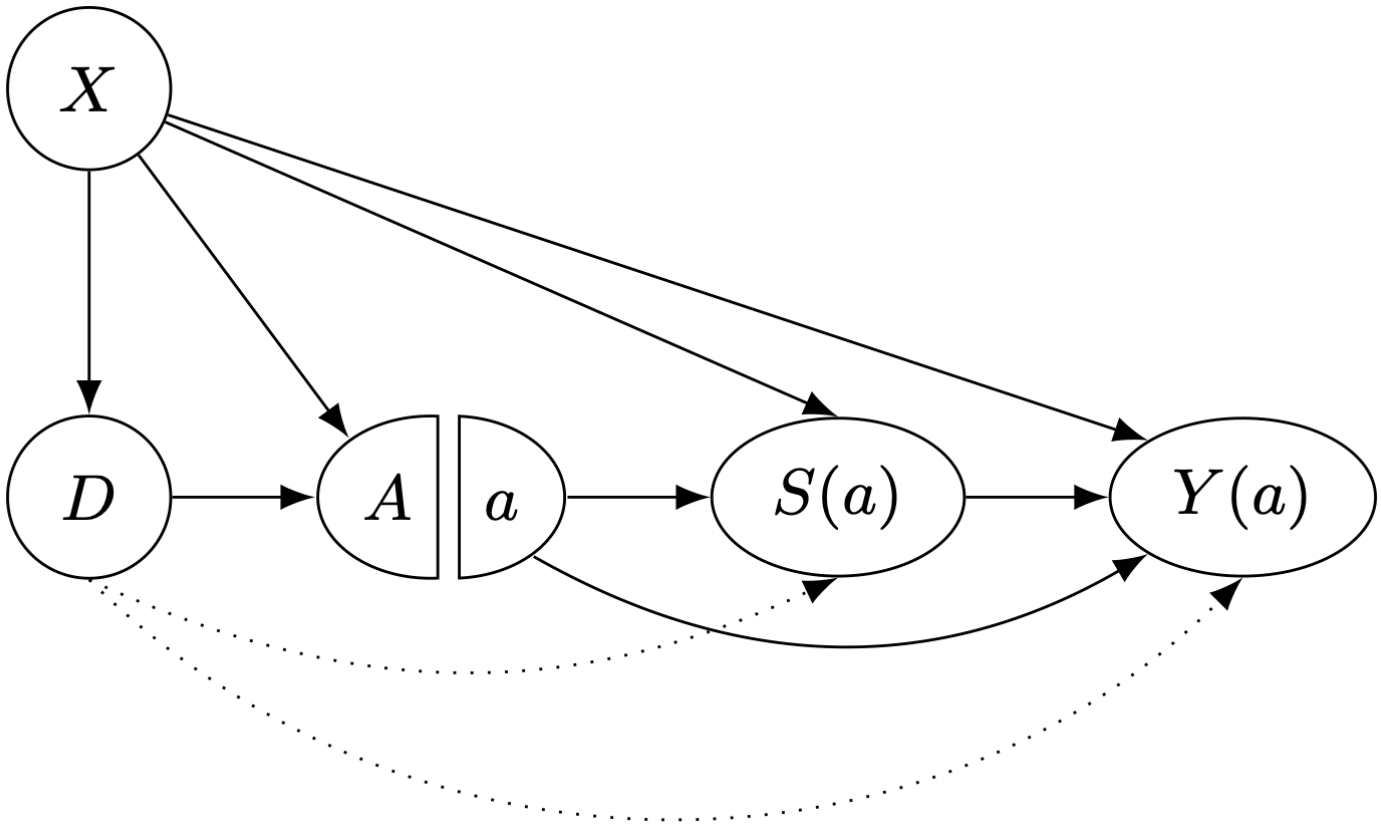


Figure: Note: Arrow from $a \to Y(a)$ means we do not assume statistical surrogacy

Summary of theoretical results

- Under Assumptions 2 to 5, the $(1-\alpha)$-quantile of the potential non-conformity scores can be identified from the observed data in all settings. ▸ See Theorem 2, slide 39
  - Identification formula for Settings 1 and 3 are the same!

Summary of theoretical results

- Under Assumptions 2 to 5, the $(1-\alpha)$-quantile of the potential non-conformity scores can be identified from the observed data in all settings. ▸ See Theorem 2, slide 39
  - Identification formula for Settings 1 and 3 are the same!
- EIFs ψ_a can be derived in all settings. ▸ See Theorem 3, slide 41
  - S provide no additional information if only observed for units that already have Y observed.

Summary of theoretical results

- Under Assumptions 2 to 5, the $(1-\alpha)$-quantile of the potential non-conformity scores can be identified from the observed data in all settings. ▸ See Theorem 2, slide 39
  - Identification formula for Settings 1 and 3 are the same!
- EIFs ψ_a can be derived in all settings. ▸ See Theorem 3, slide 41
  - S provide no additional information if only observed for units that already have Y observed.
- Slack for coverage is the sum of two parts: (i) approximating the expectation with an empirical version, scales as $O(n_{cal}^{-1/2})$; (ii) product bias from estimating nuisance functions.

▸ See Theorem 4, slide 43

Summary of theoretical results

- Under Assumptions 2 to 5, the $(1-\alpha)$-quantile of the potential non-conformity scores can be identified from the observed data in all settings. ▸ See Theorem 2, slide 39
  - Identification formula for Settings 1 and 3 are the same!
- EIFs ψ_a can be derived in all settings. ▸ See Theorem 3, slide 41
  - S provide no additional information if only observed for units that already have Y observed.
- Slack for coverage is the sum of two parts: (i) approximating the expectation with an empirical version, scales as $O(n_{cal}^{-1/2})$; (ii) product bias from estimating nuisance functions. ▸ See Theorem 4, slide 43
- Constructed prediction regions achieve the desired coverage asymptotically up to a negligible term vanishing in probability (PAC guarantee) ▸ See Cor 1, slide 45

Theorem 1: Efficiency Gain

Under Assumptions 2-5, the efficiency gain from observing surrogates in both datasets (Setting 2) compared to only in the source data (Setting 1) in estimating $r_{\alpha,1}$ is given by

$$V_1^{(S2)} - V_1^{(S1)} = E\left[\frac{1 - e_D(X,1)}{e_D(X,1)} \frac{e_D^2(X,0)\{1 - e_A(X)\}^2}{e_A(X)} \mathrm{var}\{\tilde{m}_1(X,S) \mid X\}\right],$$

where $V_1^{(S1)} = \mathrm{var}\{\psi_1^{(S1)}\}$ and $V_1^{(S2)} = \mathrm{var}\{\psi_1^{(S2)}\}$.

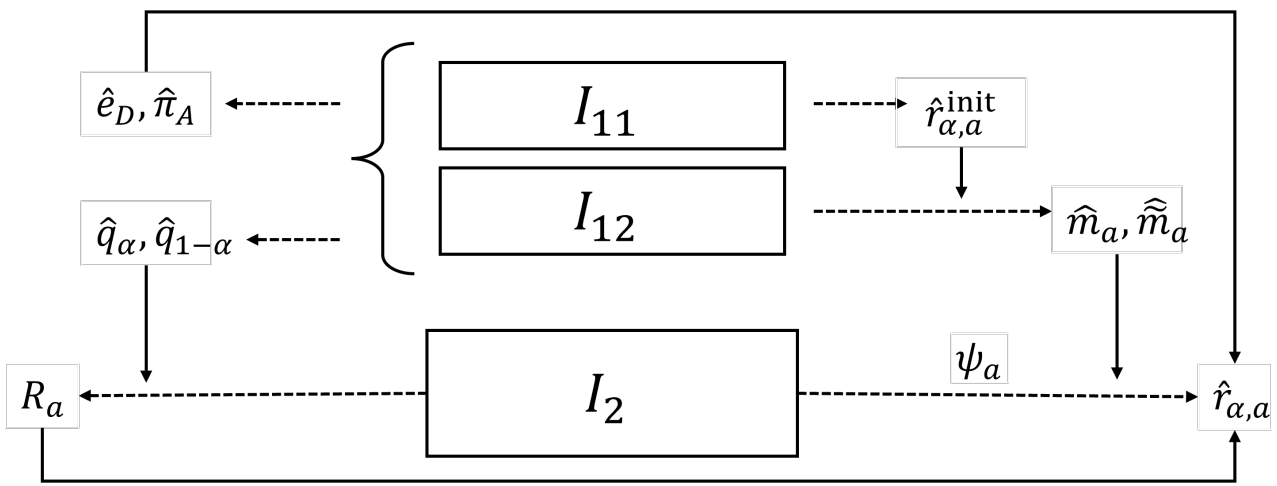
- $\uparrow \mathrm{var}\{\tilde{m}_1(X,S) \mid X\} = P(R_1 < r_{\alpha,1} \mid X, S, A = 1, D = 1$ Predictiveness of surrogates: Additional variability in predicting R_1 that is captured by the surrogates S beyond the covariates X
- $\downarrow e_D(X,1) = P(D = 1 \mid X, A = 1)$ and $\uparrow e_D(X,0) = P(D = 1 \mid X, A = 0)$: Proportion of primary outcomes observed is large for the control group but small for the treatment group: Intuitively, less missingness in the primary outcome leads to less room for efficiency gain from leveraging surrogates.

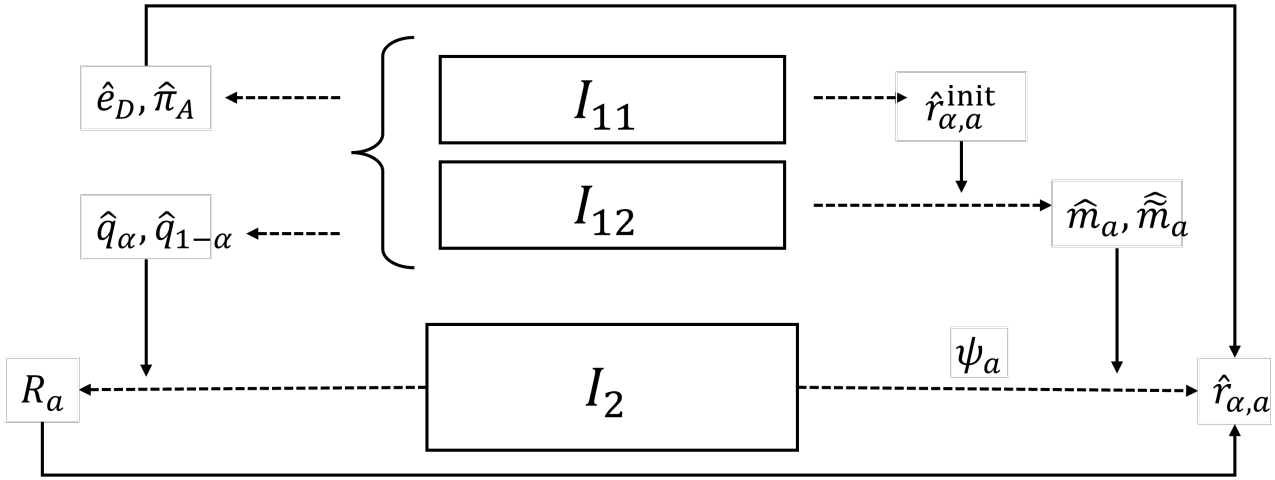
Algorithmic implementation

- Split conformal to construct efficient estimators for $r_{\alpha,a}$ (Lei and Wasserman, 2014).
- To avoid learning the entire conditional distribution for infinitely many r, we use localized debiased ML (Kallus et al., 2024), where $\mathcal{I}_1$ is further divided into $\mathcal{I}_{11}$ and $\mathcal{I}_{12}$.

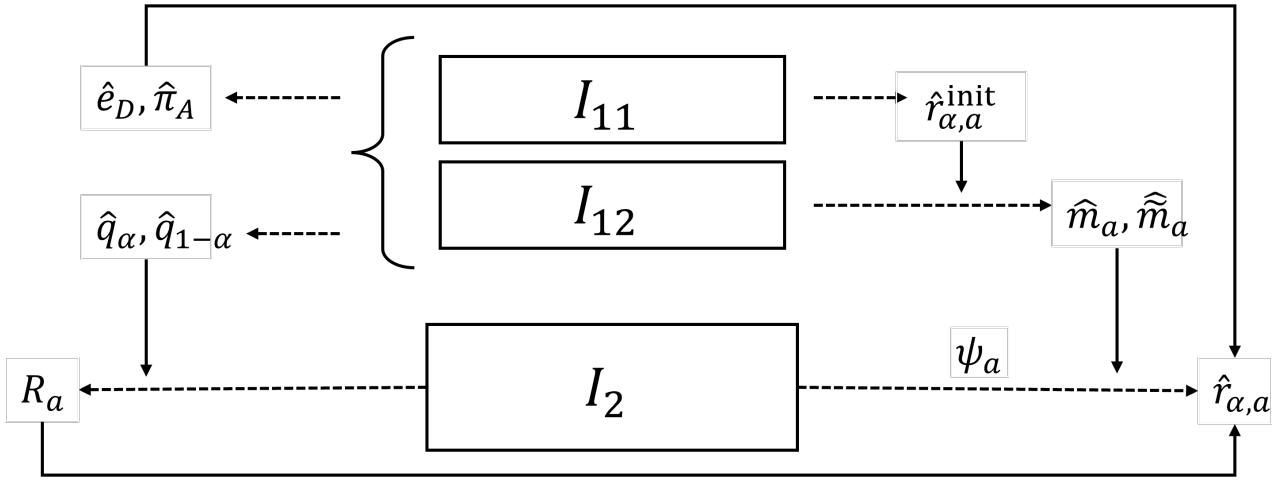
Algorithmic implementation

- Split conformal to construct efficient estimators for $r_{\alpha,a}$ (Lei and Wasserman, 2014).
- To avoid learning the entire conditional distribution for infinitely many r, we use localized debiased ML (Kallus et al., 2024), where $\mathcal{I}_1$ is further divided into $\mathcal{I}_{11}$ and $\mathcal{I}_{12}$.
  - $\mathcal{I}_{11}$: construct an initial $\widehat{r}^{\text{init}}_{\alpha,a}$ e.g., WCQR.
  - $\mathcal{I}_{12}$: estimate nuisances with any ML algorithm.

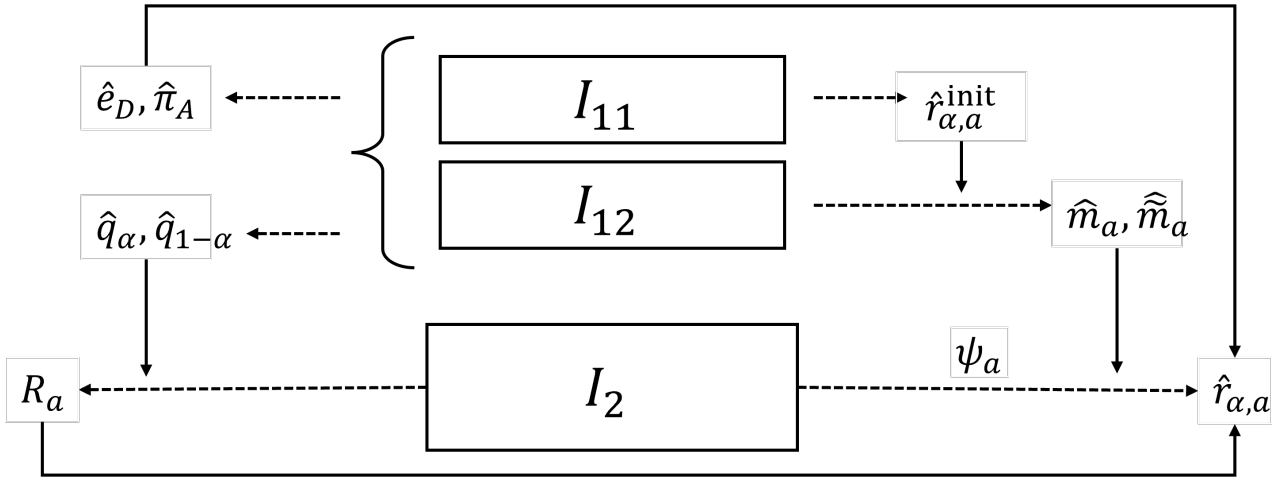




- $\mathcal{I}_2$: evaluate conformity scores
  $R_{a,i} = \mathbf{1}(A_i = a) \max\{\widehat{q}_{\alpha/2,a}(W_i) - Y_i, Y_i - \widehat{q}_{1-\alpha/2,a}(W_i)\}$,
  with estimated quantile models $\widehat{q}(\cdot)$.



- $\mathcal{I}_2$: evaluate conformity scores
  $R_{a,i} = \mathbf{1}(A_i = a) \max\{\widehat{q}_{\alpha/2,a}(W_i) - Y_i, Y_i - \widehat{q}_{1-\alpha/2,a}(W_i)\}$,
  with estimated quantile models $\widehat{q}(\cdot)$.
- $\mathcal{I}_2$: solve for the semi-parametric efficient estimator $\widehat{r}_{\alpha,a}$ with estimated nuisances such that
  $\sum_{i \in \mathcal{I}_2} \psi_a(\widehat{r}_{\alpha,a}, W_i; \widehat{m}, \widehat{\tilde{m}}, \widehat{e}_D, \widehat{\pi}_A) \geq 0$.



- $\mathcal{I}_2$: evaluate conformity scores
  $R_{a,i} = \mathbf{1}(A_i = a) \max\{\widehat{q}_{\alpha/2,a}(W_i) - Y_i, Y_i - \widehat{q}_{1-\alpha/2,a}(W_i)\}$,
  with estimated quantile models $\widehat{q}(\cdot)$.
- $\mathcal{I}_2$: solve for the semi-parametric efficient estimator $\widehat{r}_{\alpha,a}$ with estimated nuisances such that
  $\sum_{i \in \mathcal{I}_2} \psi_a(\widehat{r}_{\alpha,a}, W_i; \widehat{m}, \widehat{\tilde{m}}, \widehat{e}_D, \widehat{\pi}_A) \geq 0$.
- Construct prediction regions $C_a(W_i; \widehat{r}_{\alpha,a}) = \{y : R_{a,i} \leq \widehat{r}_{\alpha,a}\}$.

Simulation studies

- $\mathbf{X} \in \mathcal{N}(0, I_2)$ for populations of sizes $N = 3000, 5000, 10000$.
- A is generated via a logistic regression model to achieve an average treatment rate of 0.5.
- $S(a) \sim \mathcal{N}((-1)^{a+1}, \sigma_S^2 I_2)$, with σ_S ranging from 1 to 50, indicating their predictive strength for primary outcomes $Y(a)$
- $Y(a) = (-1)^{a+1} + \frac{(-1)^a}{2} \frac{\sum_{j=1}^{2} S_j(a)}{5} + \sum_{j=1}^{2} \beta_j X_j + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1)$
- Conformity scores constructed using CQR
- Nuisance functions fit using `SuperLearner`, with RandomForest and GLM as base learners
- Interest in the prediction intervals for the ITEs

Simulation results

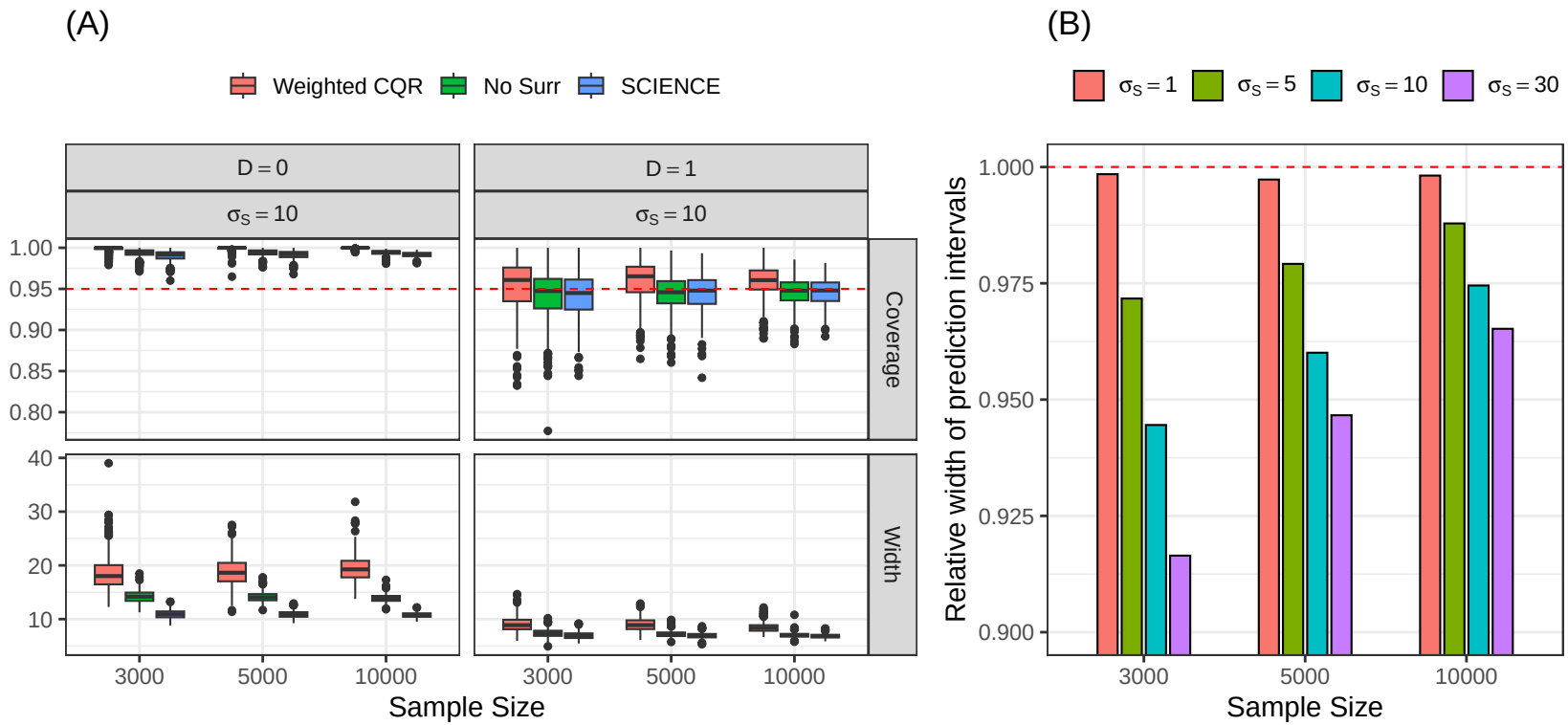


Figure: (A) Empirical coverage and width of the 95% prediction intervals, when $\sigma_S = 10$ and (B) average relative width of the 95% prediction intervals ('SCIENCE' versus 'No Surr') when $\sigma_S = 1, 5, 10, 30$ for ITE across 500 replicates.

Real data analysis

- Recall that interest is in analyzing the phase 3 mRNA-1273 (Moderna) COVID-19 VE trial
- 2 surrogate markers: (i) Day 29 pseudo-neutralizing antibody titer (nAB-ID50), and (ii) Day 29 binding spike antibody titer
- Primary outcome: Day 57 value of the nAB-ID50
- 75-25 split for training and testing, repeated 100 times, and summarized the average coverage and widths of prediction intervals for different methods and $\alpha \in \{0.05, 0.1, 0.2, 0.3, 0.4\}$

α	Coverage $\times 100\%$					Prediction Interval Width				
	WCQR	**No Surr**	**SCIENCE (Binding)**	**SCIENCE (nAb-50)**	**SCIENCE (Both)**	**WCQR**	**No Surr**	**SCIENCE (Binding)**	**SCIENCE (nAb-50)**	**SCIENCE (Both)**
0.05	95.5	95.0	95.2	94.8	95.3	2.07	1.96	1.88	1.85	1.72
0.10	91.3	89.0	90.1	89.9	90.1	1.69	1.55	1.48	1.47	1.44
0.20	81.5	80.3	80.9	79.9	80.9	1.28	1.23	1.19	1.18	1.15
0.30	66.1	70.0	72.0	70.3	71.7	1.05	1.02	1.02	1.00	0.98
0.40	59.4	59.2	61.3	60.5	60.9	0.89	0.88	0.86	0.85	0.83

Visualizing efficiency gain across a discrete surrogate

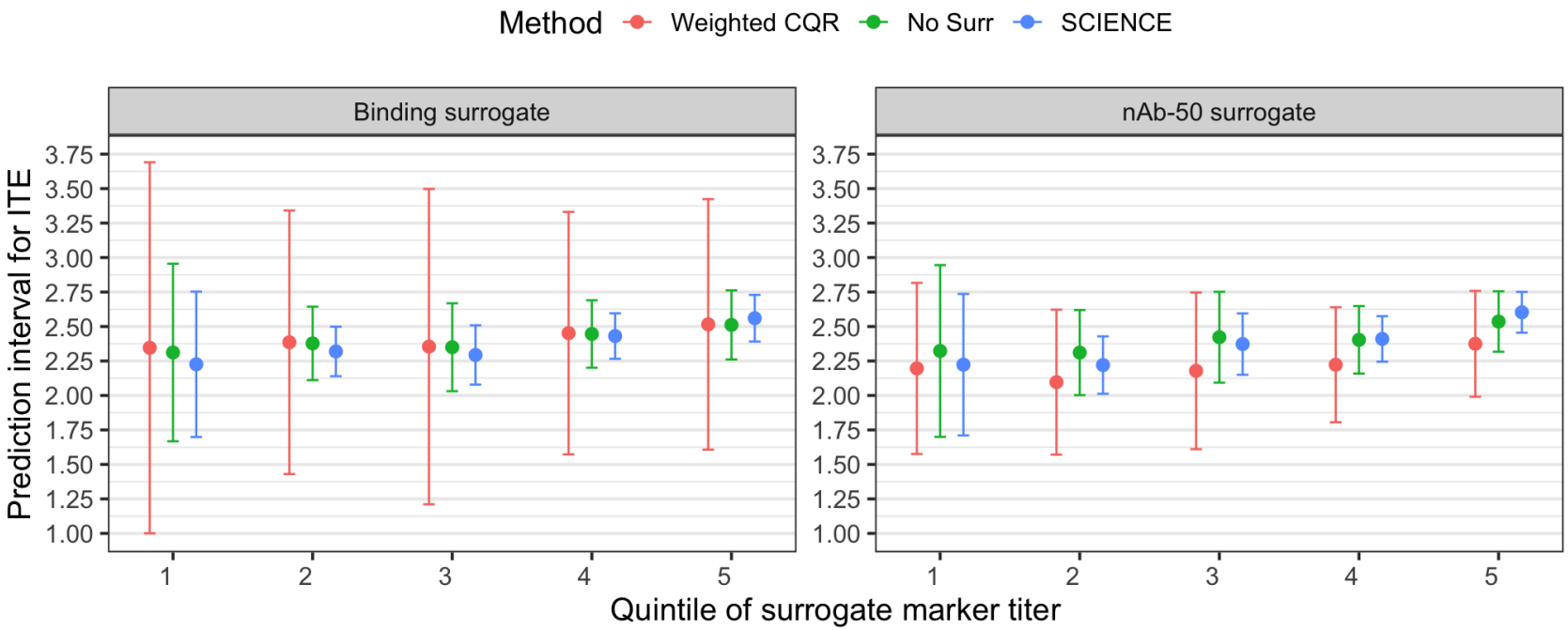


Figure: 90% prediction intervals for the ITE on the Day 57 neutralizing antibody titer primary outcome across quintiles of Day 29 surrogate marker titers

R package SurrConformalDR

- github.com/Gaochenyin/SurrConformalDR

SurrConformalDR

An R package for Surrogate-assisted Conformal Inference for Efficient Individual Causal Effect Estimation

Installation

To access the vignette, you can install the development version of SurrConformalDR from GitHub with:

```
install.packages("devtools")
devtools::install_github("Gaochenyin/SurrConformalDR",
                         build_vignettes = TRUE)
```

Discussion

- We proposed a novel Surrogate-assisted Conformal Inference for Efficient INdividualized Causal Effects (SCIENCE) for constructing more efficient prediction intervals.
- SCIENCE is applicable to scenarios with or without observed primary outcomes and handles distributional shifts between source and target populations.
- Extensions to balancing fairness and efficiency See slide 53-on
  - Gao, C., Gilbert, P., and **Han, L.** Balancing Fairness and Efficiency in Conformal Inference. ICML 2025.

Appendix Slides

Sensitivity analysis: surrogate shift & outcome shift

- Goal: assess robustness of SCIENCE to violations of transportability assumptions.
- We evaluate empirical coverage under a family of perturbed DGPs indexed by

$$(\Delta_S, \Delta_Y),$$

  where Δ_S controls **surrogate shift** and Δ_Y controls **outcome shift**.
- Let $P(S(a) \mid X, D)$ denote conditional surrogate densities and

$$m_{Y(a)}(S, X, D) := \mathbb{E}\{Y(a) \mid S, X, D\}$$

  denote conditional mean functions of potential outcomes.

Interpretation

- $\Delta_S = 0$ and $\Delta_Y = 0$ correspond to the no-shift benchmark.
- Increasing $\|\Delta_S\|$ or $\|\Delta_Y\|$ for more shift.

◂ Back

Surrogate shift via exponential tilting

- Violate surrogate transportability (Assumption 4) by modifying the target surrogate distribution ($D = 0$) relative to source ($D = 1$).
- Model the target as an exponentially-tilted version of the source:

$$P(S(a) \mid X, D = 0; \Delta_S) \;\propto\; P(S(a) \mid X, D = 1) \exp\left\{\Delta_S^\top g_S(s, a, x)\right\},$$

- $g_S(\cdot)$ is user-specified (e.g., moments/features of S); Δ_S quantifies shift magnitude.
- In simulations (Gaussian surrogates), tilting induces a mean shift:

$$S(a) \mid D = 0 \sim \mathcal{N}\big((-1)^{a+1} + \sigma_S^2 \Delta_S,\; \sigma_S^2 I_2\big).$$

◂ Back

Outcome shift model

- Violate MAR (Assumption 5) by allowing target outcome regression to deviate from the source regression.

Sensitivity models for $m_{Y(a)}$

Continuous outcomes (additive shift):

$$m_{Y(a)}(S, X, D = 0; \Delta_Y) = m_{Y(a)}(S, X, D = 1) + \Delta_Y^\top g_Y(a, x, s).$$

Binary outcomes (log-odds shift):

$$\text{logit}\{m_{Y(a)}(S, X, D = 0)\} = \text{logit}\{m_{Y(a)}(S, X, D = 1)\} + \Delta_Y^\top g_Y(a, x, s).$$

◂ Back

Sensitivity analysis results

- Empirical coverage of ITE prediction intervals under induced distribution shifts:

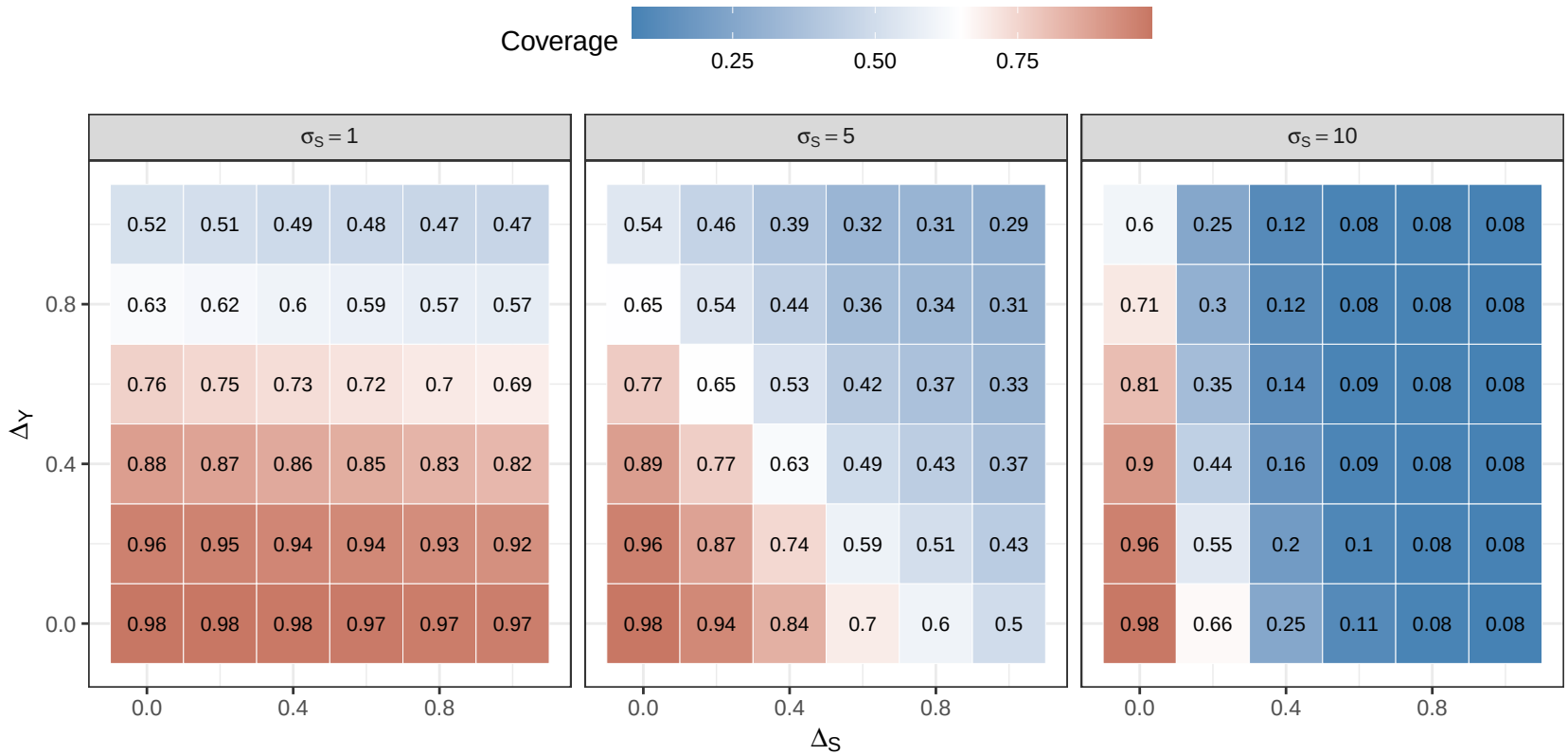


Figure: Empirical coverage of 95% prediction intervals for target ITEs under (Δ_S, Δ_Y) and $\sigma_S \in \{1, 5, 10\}$ (500 replicates).

◂ Back

Theorem 2: Identification

Under Assumptions 2 to 5, we have

$$\begin{aligned} 1-\alpha &= P(R_a < r_{\alpha,a} \mid A = 1-a, D = 1) \\ &= E_W\{P_Y(R_a < r_{\alpha,a} \mid A = a, D = 1, W) \mid A = 1-a, D = 1\}. \end{aligned}$$

Thus, the identification formulas for $r_{\alpha,a}$ under Settings 1 and 3 are

$$1-\alpha = E_X\{P_Y(R_a < r_{\alpha,a} \mid A = a, D = 1, X) \mid A = 1-a, D = 1\},$$

and under Setting 2 is

$$\begin{aligned} 1-\alpha &= E_W\{P_Y(R_a < r_{\alpha,a} | A = a, D = 1, W) | A = 1-a, D = 1\} \\ &= E_X[E_S\{P_Y(R_a < r_{\alpha,a} | A = a, D = 1, S, X) | A = a, X\} | A = 1-a, D = 1]. \end{aligned}$$

Unit	X	A	S	Y
1	✓	✓	?	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	?	✓
$n_{D1}+1$	✓	✓	?	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	?	?

Table: Semi-supervised setting

Unit	X	A	S	Y
1	✓	✓	✓	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	✓	✓
$n_{D1}+1$	✓	✓	✓	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	✓	?

Table: Surrogate-assisted SS

Unit	X	A	S	Y
1	✓	✓	✓	✓
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
n_{D1}	✓	✓	✓	✓
$n_{D1}+1$	✓	✓	?	?
⋮	⋮	⋮	⋮	⋮
$n_{D1}+n_{D0}$	✓	✓	?	?

Table: Intermediate setting

◂ Back

Defining nuisance functions

Model	**Definition**
$e_A(X) = P(A = 1 \mid X)$	PS for treatment
$\pi_A(X) = \{1 - e_A(X)\}/e_A(X)$	inverse odds of treatment assignment
$e_D(X) = P(D = 1 \mid X)$	PS for observing primary outcomes
$e_D(X, A) = P(D = 1 \mid X, A)$	PS for observing primary outcomes within each treatment group
$\pi_D(X) = \{1 - e_D(X)\}/e_D(X)$	inverse odds of observing primary outcomes
$\tilde{m}_a(r, X, S)$ $= P(R_a < r \mid X, S, A = a, D = 1)$	conditional CDF of R_a at r w/surrogates
$m_a(r, X)$ $= P(R_a < r \mid X, A = a, D = 1)$	conditional CDF of R_a at r w/out surrogates
$\tilde{m}_C(r, X, S)$ $= P(R_C < r \mid X, S, D = 1)$	conditional CDF of R_C at r w/surrogates
$m_C(r, X)$ $= P(R_C < r \mid X, D = 1)$	conditional CDF of R_C at r w/out surrogates

Theorem 3: Efficient Influence Functions

Under Assumptions 2-5, the EIF $\psi_1^{(j)}$ for $r_{\alpha,1}$ under Setting j, for $j = S1, S2, S3$, is given, up to a proportionality constant, by

$$\begin{aligned}
&\psi_1^{(S1)}(r_{\alpha,1}, W; m, e_D, \pi_A) = \psi_1^{(S3)}(r_{\alpha,1}, W; m, e_D, \pi_A) \\
&= D(1-A)\{m_1(r_{\alpha,1}, X) - (1-\alpha)\} \\
&+ \frac{AD\pi_A(X)e_D(X,0)}{e_D(X,1)}\{\mathbf{1}(R_1 < r_{\alpha,1}) - m_1(r_{\alpha,1}, X)\},
\end{aligned}$$

$$\begin{aligned}
\psi_1^{(S2)}(r_{\alpha,1}, W; m, \tilde{m}, e_D, \pi_A) &= D(1-A)\{m_1(r_{\alpha,1}, X) - (1-\alpha)\} \\
&+ A\pi_A(X)e_D(X,0)\{\tilde{m}_1(r_{\alpha,1}, X, S) - m_1(r_{\alpha,1}, X)\} \\
&+ \frac{AD\pi_A(X)e_D(X,0)}{e_D(X,1)}\{\mathbf{1}(R_1 < r_{\alpha,1}) - \tilde{m}_1(r_{\alpha,1}, X, S)\}.
\end{aligned}$$

Intuition: Surrogates provide no extra information if they are only observed for the units that already have primary outcomes observed.

Back

Asymptotic properties

Establish asymptotic properties for the estimators of $r_{\alpha,1}$ and $r_{\alpha,0}$, which are important for constructing prediction intervals with the desired coverage level $1-\alpha$.

Lemma 1: Deviation of asymptotic coverage

For any EIF $\psi_a(r_{\alpha,a}, W)$ under Settings 1 – 3, the following holds

$$P(R_a < r_{\alpha,a} \mid A = 1-a, D = 1) = 1 - \alpha + \frac{E\{\psi_a(r_{\alpha,a}, W)\}}{P(A = 1-a, D = 1)}.$$

Lemma 42 states that the asymptotic coverage probability deviates from the nominal level $1-\alpha$ by a term proportional to the expected value of the EIF. Back

Theorem 4

Under Assumptions 2 to 5 and monotonicity and boundedness conditions, there exist some constants C_0 and C_1 for any $\delta > 0$, such that

$$P(R_a \leq \widehat{r}_{\alpha,a}^{(S1)} \mid A_i = 1-a, D_i = 1) \geq (1-\alpha)$$
$$- C_0 \frac{\pi_0 \bar{e}_0 \underline{e}_0^{-1}(1+m_0)}{P(A=1-a, D=1)} \sqrt{\frac{\log(1/\delta)+1}{|\mathcal{I}_2|}}$$
$$- C_1 \left\{ \frac{\|\widehat{\pi}_A(X) - \pi_A(X)\|}{P(A=1-a, D=1)} \cdot \sup_r \|\widehat{m}_a(r,X) - m_a(r,X)\| \right.$$
$$\left. + \frac{\|\widehat{e}_D(X,A) - e_D(X,A)\|}{P(A=1-a, D=1)} \cdot \sup_r \|\widehat{m}_a(r,X) - m_a(r,X)\| \right\}$$

holds with probability at least $1-\delta$.

◂ Back

Theorem for asymptotic properties for coverage (Informal)

Coverage guarantee of $1-\alpha$ for the prediction intervals of $u\{Y(1)\}$ and $u\{Y(0)\}$: slack for the coverage guarantee is the sum of two parts:

- first part is due to approximating the expectation with its empirical version, which scales as $O(N^{-1/2})$, by splitting the data into two folds $\mathcal{I}_1$ and $\mathcal{I}_2$ of similar size
- (2) second part is by the product bias from the estimation of the nuisance functions, which is negligible if either:
  - $\|\widehat{\pi}_A(X) - \pi_A(X)\| + \|\widehat{e}_D(X,A) - e_D(X,A)\| = o_p(1)$; or
  - $\sup_r \|\widehat{m}_a(r,X) - m_a(r,X)\| + \sup_r \|\widehat{\tilde{m}}_a(r,X,S) - \tilde{m}_a(r,X,S)\| = o_p(1)$.

◂ Back

Corollary 1: Probably Approximately Correct (PAC) coverage

Under Settings 1-3, the overall coverage in the source data satisfies

$$
\begin{aligned}
&P(\theta_i \in C_\theta(W_i; \widehat{r}_\alpha) \mid D_i = 1) \\
&= P(A_i = 1 \mid D_i = 1) P(R_0 \leq \widehat{r}_{\alpha,0} \mid A_i = 1, D_i = 1) \\
&+ P(A_i = 0 \mid D_i = 1) P(R_1 \leq \widehat{r}_{\alpha,1} \mid A_i = 0, D_i = 1) \\
&\geq (1 - \alpha) + o_p(1).
\end{aligned}
$$

This confirms that the constructed prediction regions achieve the desired coverage level asymptotically in the source data, up to a negligible term vanishing in probability (Krishnamoorthy and Mathew, 2009).

◂ Back

Group-conditional performance for continuous outcomes

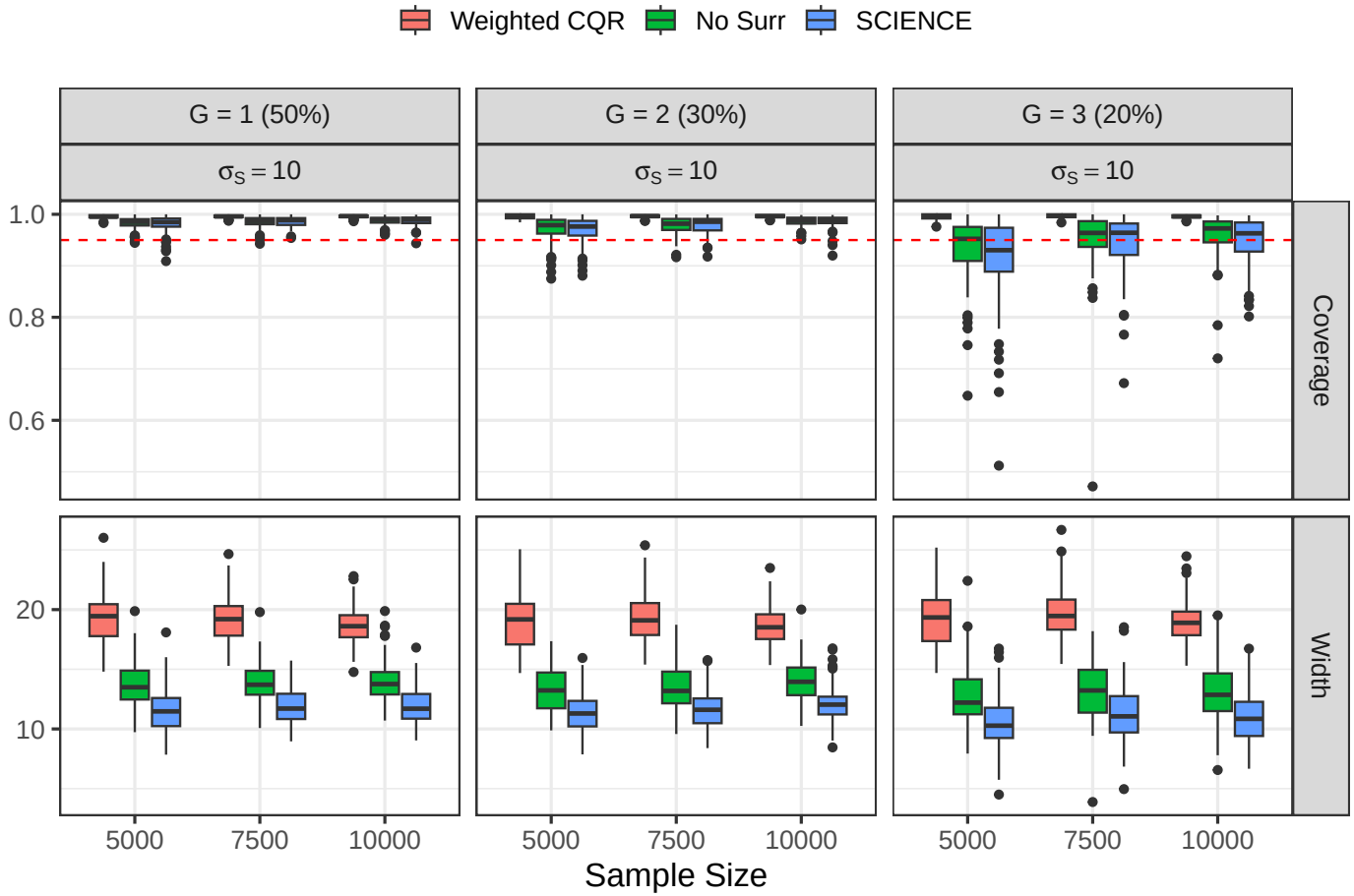


Figure: Empirical coverage and width of the 95% prediction intervals for the ITE of the combined data when $\sigma_S = 10$, conditional on the group variable G, across 500 replicates.

Real data results

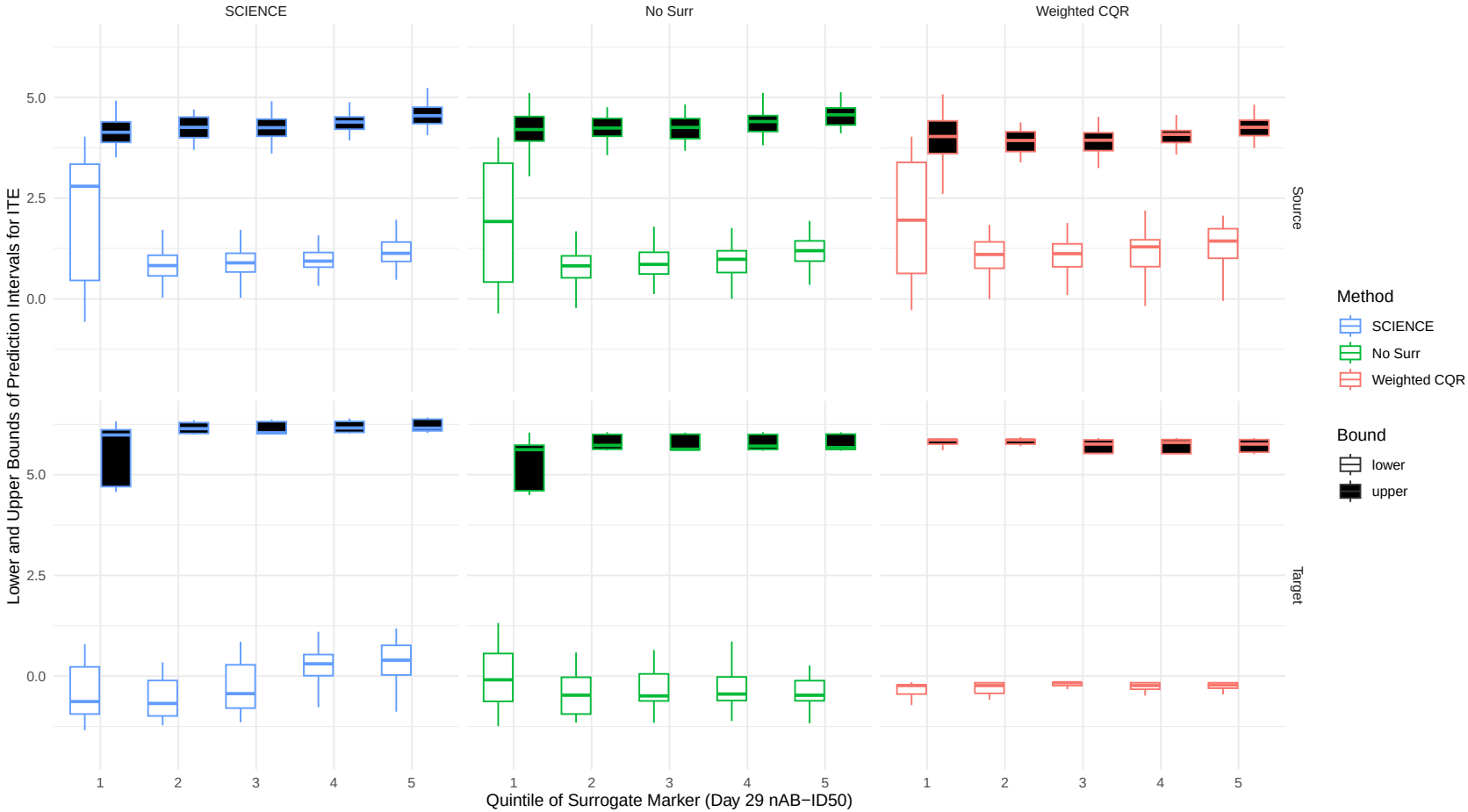


Figure: Lower and upper bounds of the 90% prediction intervals for test points on the source and target populations on the Day 57 nAb-ID50 primary outcome across quintiles of Day 29 nAb-ID50 surrogate marker titer; Quintile 1: $(0.0828, 0.501]$, Q2: $(0.501, 0.944]$, Q3: $(0.944, 1.291]$, Q4: $(1.291, 1.627]$, Q5: $(1.627, 4.068]$ in units IU50/ml

Recall conformal inference on the target data ($D = 0$)

$$P\{\theta_i \in C_\theta(W_i; r_\alpha)\} = P(D_i = 1)P(\theta_i \in C_\theta(W_i; r_\alpha) \mid D_i = 1) + P(D_i = 0)P(\theta_i \in C_\theta(W_i; r_\alpha) \mid D_i = 0).$$

Building upon the nested conformal inference framework proposed by Lei and Candès (2021), we develop a two-step procedure to achieve this goal.

Conformal inference on the target data

For term (2), where both the potential primary outcomes are missing, our plan is to first create the data (W_i, C_i) by

$$C_i = \begin{cases} u(Y_i) - C_0(W_i; r_{\alpha,0}), & A_i = 1, D_i = 1 \\ C_1(W_i; r_{\alpha,1}) - u(Y_i), & A_i = 0, D_i = 1, \end{cases}$$

where C_i is the prediction regions for θ_i, which can be perceived as the pseudo-outcomes. These intervals C_i are designed to cover the causal estimand θ_i with probability at least $1 - \alpha$ conditional on the source data

Next, a secondary conformal inference is performed on the left- and the right-end point of C_i to create the prediction regions $C(W; r_{\gamma,C})$ that nests C_i with high probability, satisfying:

$$P(C_i \subset C(W; r_{\gamma,C}) \mid D_i = 0) \geq 1 - \gamma,$$

and we will have $P(\theta_i \in C(W_i; r_{\gamma,C}) \mid D_i = 0) \geq 1 - \alpha - \gamma$.

Specifically, we aim to determine $r_{\gamma,C}$ such that

$$P\big(C_i \subset C(W_i; r_{\gamma,C}) \mid D_i = 0\big) = P\big(R_{C,i} \leq r_{\gamma,C} \mid D_i = 0\big) = 1 - \gamma,$$

where $R_{C,i} = R(W_i, C_i)$ is the conformity score measuring the discrepancy between the pseudo-outcome C and its prediction based on W.

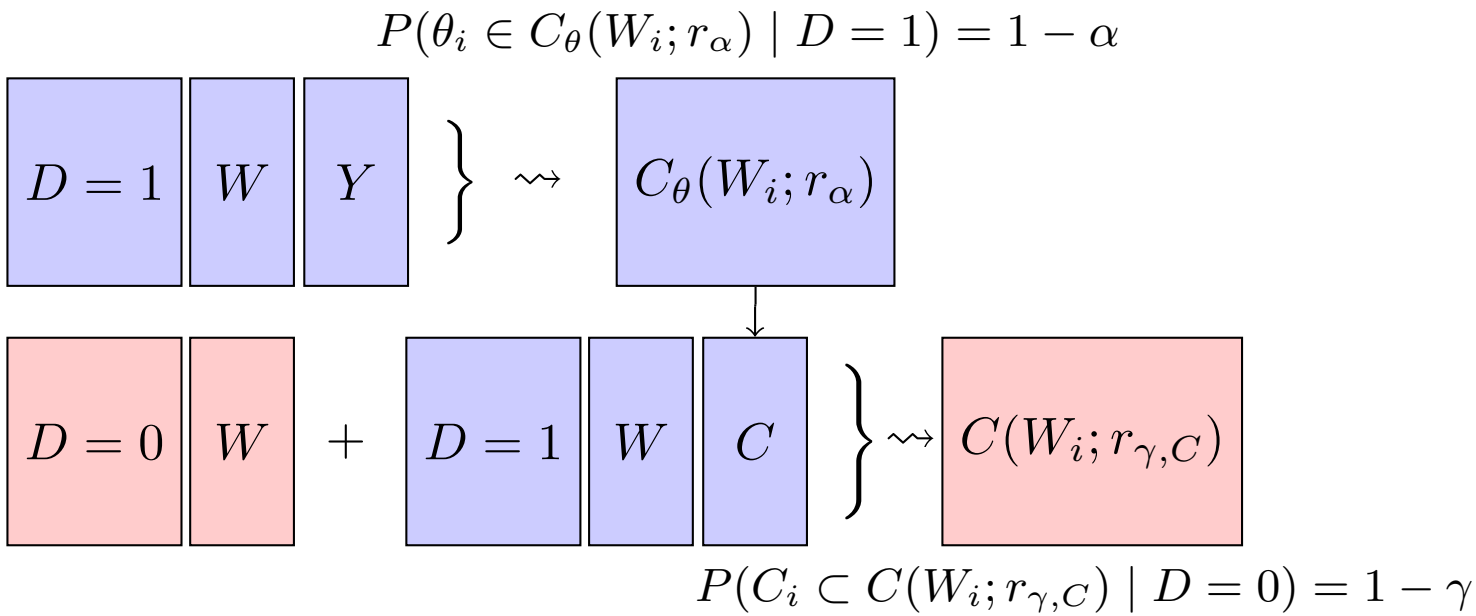


Figure: Schematic illustration for constructing the nested prediction regions $C(W; r_{\alpha,C})$ for target data $D = 0$

Theorem 5

Under Assumptions 2-5, let $r_{\gamma,C}$ be the $(1-\gamma)$-th quantile of the conformity scores R_C. The EIF for $r_{\gamma,C}$ is, up to a proportionality constant, for Setting 2,

$$\begin{aligned}\psi_C^{(S2)}(r_{\gamma,C}, W; m_C, \tilde{m}_C, e_D, \pi_A) &= (1-D)\{m_C(r_{\gamma,C}, X) - (1-\gamma)\} \\ &+ \{1 - e_D(X)\}\{\tilde{m}_C(r_{\gamma,C}, X, S) - m_C(r_{\gamma,C}, X)\} \\ &+ D\pi_D(X)\{\mathbf{1}(R_C < r_{\gamma,C}) - \tilde{m}_C(r_{\gamma,C}, X, S)\},\end{aligned}$$

and for Settings 1 and 3,

$$\begin{aligned}&\psi_C^{(S1)}(r_{\gamma,C}, W; m_C, e_D, \pi_A) = \psi_C^{(S3)}(r_{\gamma,C}, W; m_C, e_D, \pi_A) \\ &= (1-D)\{m_C(r_{\gamma,C}, X) - (1-\gamma)\} + D\pi_D(X)\{\mathbf{1}(R_C < r_{\gamma,C}) - m_C(r_{\gamma,C}, X)\}.\end{aligned}$$

Analogously, we could quantify the optimal efficiency gain from leveraging surrogates for estimating $r_{\gamma,C}$.

Corollary 2

Under Assumptions 2 to 5, the efficiency gain from using surrogates in estimating $r_{\gamma,C}$ in the target data is

$$V_C^{(S1)} - V_C^{(S2)} = E\left[\pi_D(X)\{1 - e_D(X)\}^2 \mathrm{var}\{\tilde{m}_C(X,S) \mid X\}\right],$$

where

$$V_C^{(S1)} = \mathrm{var}\{\psi_C^{(S1)}(r_{\gamma,C}, W; m_C, e_D, \pi_A)\},$$

$$V_C^{(S2)} = \mathrm{var}\{\psi_C^{(S2)}(r_{\gamma,C}, W; m_C, \tilde{m}_C, e_D, \pi_A)\}.$$

Corollary 2 shows that the efficiency gain from incorporating surrogates depends positively on the variance of $\tilde{m}_C(X,S)$ given X. This variance captures the additional predictiveness of the surrogates for the conformity scores R_C.

Fairness and protected groups

- Protected group Z
  - Age, race or other sensitive information

$$Pr\{y \in C_\alpha(X; \boldsymbol{r_\alpha})\} = 1 - \alpha \quad \not\Rightarrow \quad Pr\{y \in C_\alpha(X; \boldsymbol{r_\alpha}) \mid Z = z\} = 1 - \alpha$$

Group-specific thresholds (Vovk 2012)

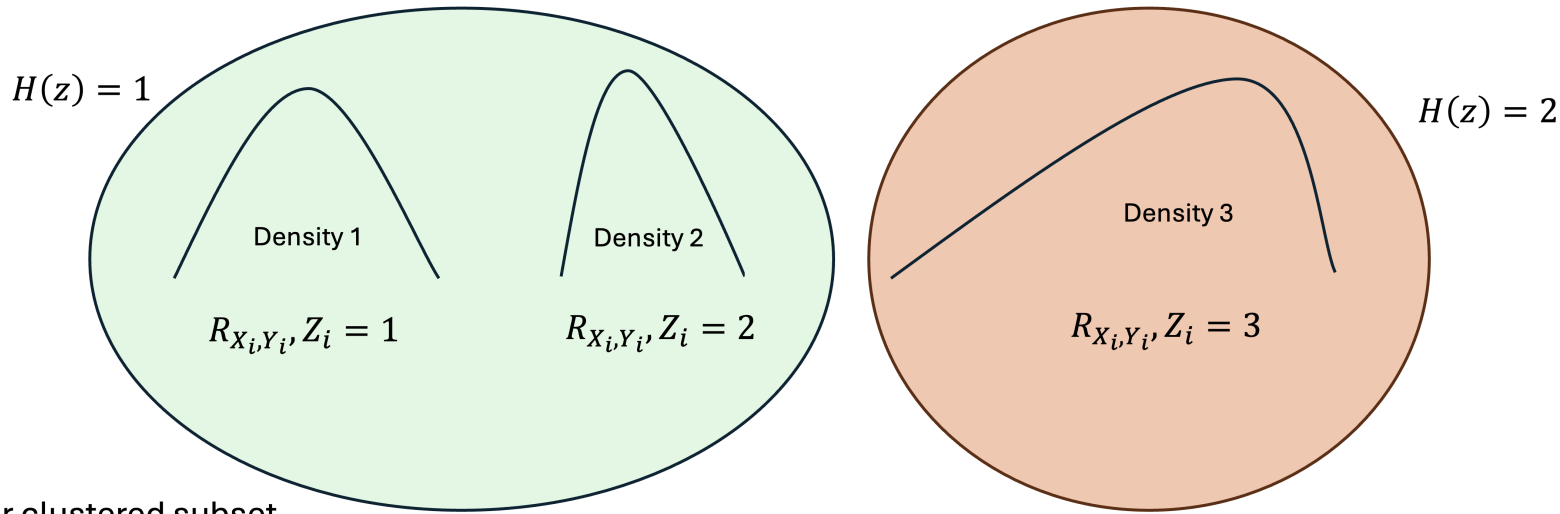
$$Pr\{y \in C_\alpha(X; \boldsymbol{r_\alpha^z}) \mid Z = z\} = 1 - \alpha, \boldsymbol{r_\alpha^z} \text{ is obtained on the } \textbf{subset } I^z = \{i : Z_i = z\}$$

Less efficient

Back

Clustered conformal prediction

- Cluster the subsets $I^z = \{i: Z_i = z\}$ if the **scores** $R_{X,Y}$ are similar



A larger clustered subset

$I^k = \{i: H(z_i) = k\}$

A clustering algorithm $H(z)$, e.g. K-means

Back

Surrogate-assisted group-clusted conformal prediction

- Cluster the subsets $I^z = \{i: Z_i = z\}$ if the **scores** $R_{X,Y}$ are similar
- **Surrogate-assisted Doubly robust** for **cluster k**

$$\sum_{i \in I^k} \left\{ D_i \frac{P(D=0 \mid X_i)}{P(D=1 \mid X_i)} \{1(R_i \le r_\alpha^k) - m(r_\alpha^k, X)\} - (1-D_i)\{m(r_\alpha^k, X) - (1-\alpha)\} \right\}$$
$$+ \sum_{i \in I^k} \left\{ D_i \frac{P(D=0 \mid X_i)}{P(D=1 \mid X_i)} - P(D=0 \mid X_i) \right\} \{m(r_\alpha^k, X) - \tilde{m}(r_\alpha^k, X, S)\} = 0$$

where $I^k = \{i: H(z_i) = k\}$

◂ Back

Surrogate-assisted group-clusted conformal prediction

- Cluster the subsets $I^z = \{i: Z_i = z\}$ if the **scores** $R_{X,Y}$ are similar
- Clustered conformal prediction is a *tradeoff* between **efficiency** and **fairness**.

  (**Efficiency**): $\left|I^{\mathrm{k}} = \{i:\ H(z_i) = k\}\right| \geq |I^z = \{i: Z_i = z\}|$

  (**Fairness**) $Pr\{y \in C_\alpha(X; \boldsymbol{r_\alpha^k}) \mid Z = z\} \geq 1 - \alpha - o(1)$ where $H(z) = k$ if clustering is sufficiently good.

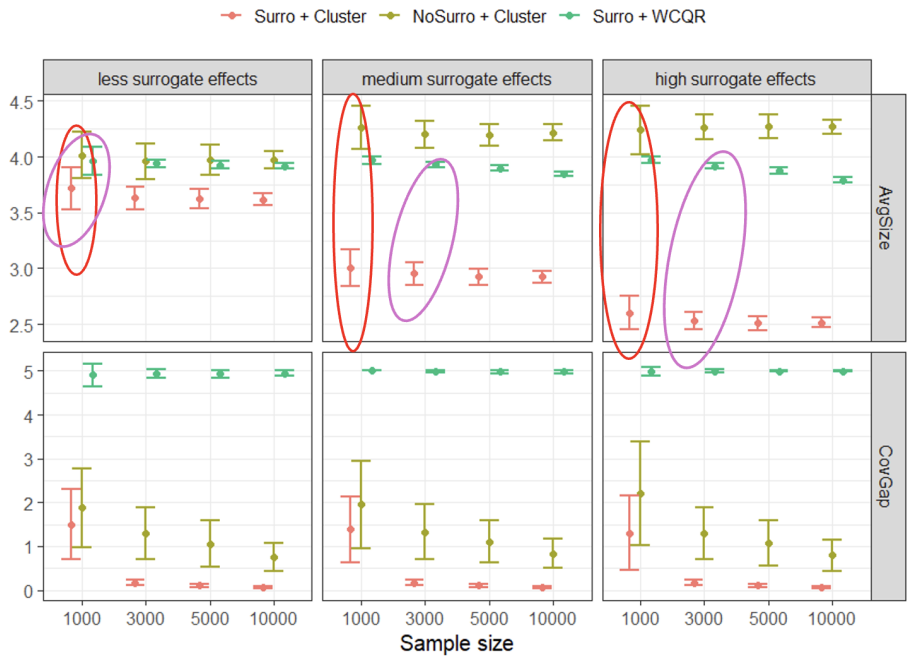
◂ Back

Experiments

- Number of group 20 and number of cluster 5.
- $R_{X,Y}$ is the *conformalized quantile residuals* (Romano et al.,2019)
- $H(Z)$ is the *k-means* (guaranteed to be good in Zhu et al. (2021))
- Two metrics
  - **Efficiency**: average of length of $C_\alpha\left(X; \boldsymbol{r}_{\alpha}^{\boldsymbol{k}}\right)$
  - **Fairness**: The gap in group-coverage $\sum_z |\hat{c}_z - (1-\alpha)|$, where $\hat{c}_z$ is the empirical coverage for group z.

◂ Back

Results



- **AvgSize -> Efficiency (lower better)**

Surro+Cluster **(proposed)**

Improve in **Eff** by EIF

Surro+WCQR

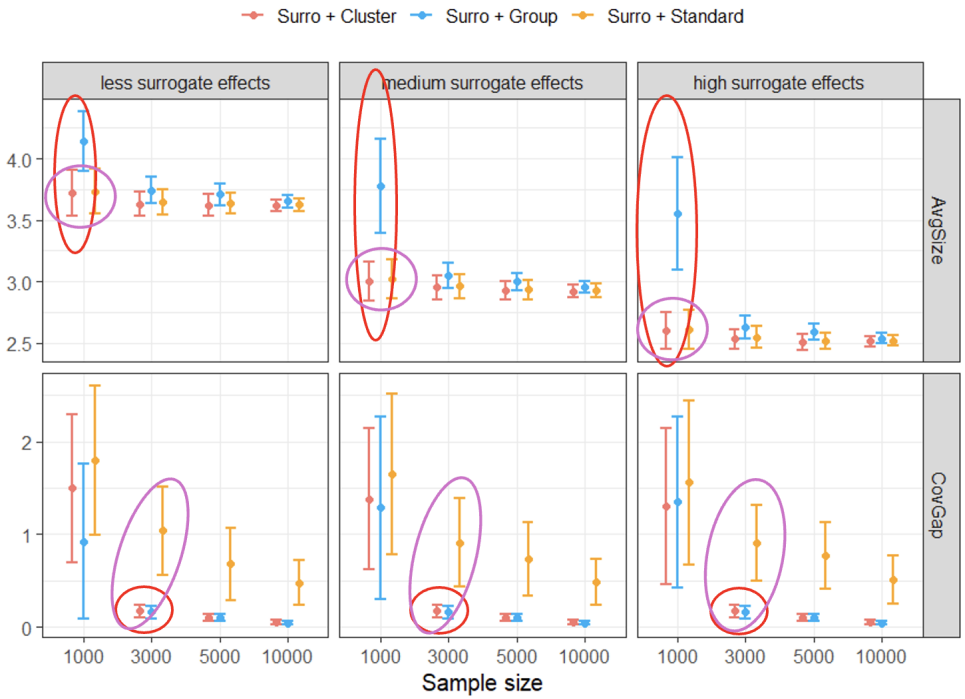
Improve in **Eff** by Surrogates

NoSurro+Cluster

- **CovGap -> Fairness (lower better)**

Back

Results



Tradeoff ☺

Surro+Cluster (proposed)

1) Improvement in **Efficiency** by clustering

2) Similar Fairness

1) Improvement in **Fairness**

2) Similar Efficiency

Surro+Group

Estimated within each group

Surro+Standard

Estimated with the full data

◂ Back

Real-data

Data: Moderna phase3 COVID 19 vaccine efficacy trial

- Source data (with labels): non-minorities (e.g., White, Asian, $n = 766$)
- Target data (without labels): minorities (African American, Latinx, $n = 652$)

- Outcome Y: antibody level at Day **57**
- Surrogate S: nAb-ID 50 titer at Day **29**
- Protected groups Z: age and comorbidities (**three** groups)
- Covariates X: sex, BMI, enrollment time, ...

◂ Back

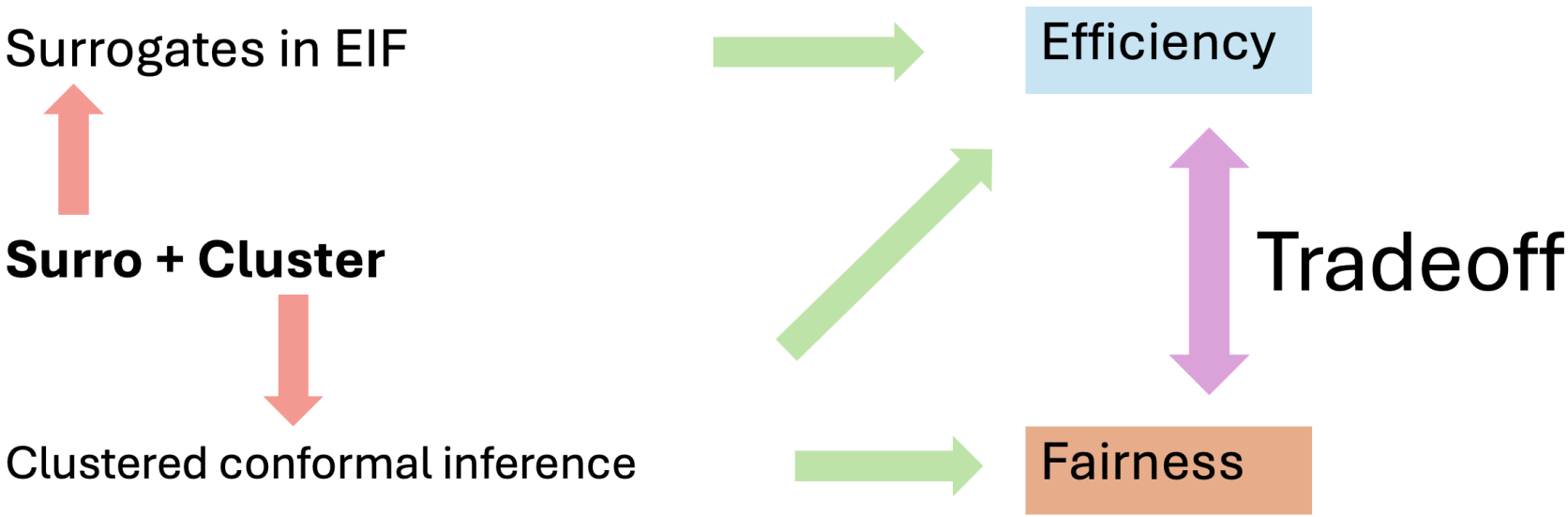
Proposed method performs well!

Method	AvgSize	CovGap	Coverage	Width
Fair but less eff				
Surro+Group	3.93	**0.14**	**(0.95, 0.95, 0.95)**	(3.80, 4.06, 3.94)
Eff but less fair				
Surro + WCQR	4.05	5.24	(0.98, 0.82, 0.88)	(4.21, 3.82, 4.24)
Surro + Standard	3.83	1.09	(0.95. 0.93, 0.96)	(3.82, 3.91, 3.76)
Tradeoff				
Surro + Cluster (proposed)	**3.61**	0.18	**(0.95, 0.95, 0.95)**	(3.63, 3.55, 3.66)

Meet our goal! 

◂ Back

Takeaways


Surrogates in EIF
Efficiency
Surro + Cluster
Tradeoff
Clustered conformal inference
Fairness

References I

A. N. Angelopoulos, S. Bates, C. Fannjiang, M. I. Jordan, and T. Zrnic. Prediction-powered inference. *Science*, 382:669–674, 2023. doi: 10.1126/science.adi6000. URL https://www.science.org/doi/abs/10.1126/science.adi6000.

A. N. Angelopoulos, R. F. Barber, and S. Bates. Theoretical foundations of conformal prediction. *arXiv preprint arXiv:2411.11824*, 2024.

S. Athey, R. Chetty, G. W. Imbens, and H. Kang. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, The Review of Economic Studies, 2025.

P. J. Bickel, C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov. *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer, 1993.

J. Hahn. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pages 315–331, 1998.

References II

M. Hernán and J. Robins. Causal inference: What if. boca rat chapman hill/crc; 2020, 2020.

P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–960, 1986.

N. Kallus, X. Mao, and M. Uehara. Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. *Journal of Machine Learning Research*, 25:1–59, 2024.

K. Krishnamoorthy and T. Mathew. *Statistical tolerance regions: theory, applications, and computation*. John Wiley & Sons, 2009.

J. Lei and L. Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76:71–96, 2014.

L. Lei and E. J. Candès. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83:911–938, 2021.

References III

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866, 1994.

Y. Romano, E. Patterson, and E. Candes. Conformalized quantile regression. *Advances in Neural Information Processing Systems*, 32, 2019.

D. Whicher, M. Ahmed, J. Paulus, and D. Kent. Caring for the individual patient: Understanding heterogeneous treatment effects. 2019.

Y. Yang, A. K. Kuchibhotla, and E. Tchetgen Tchetgen. Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae009, 2024.