Multiply Robust Federated Estimation of Targeted Average Treatment Effects

Larry Han, Zhu Shen, Jose Zubizarreta

Northeastern University, Harvard University

Promises of distributed research networks

- OHDSI
- PCORNET
- FDA Sentinel
- 4CE
- …

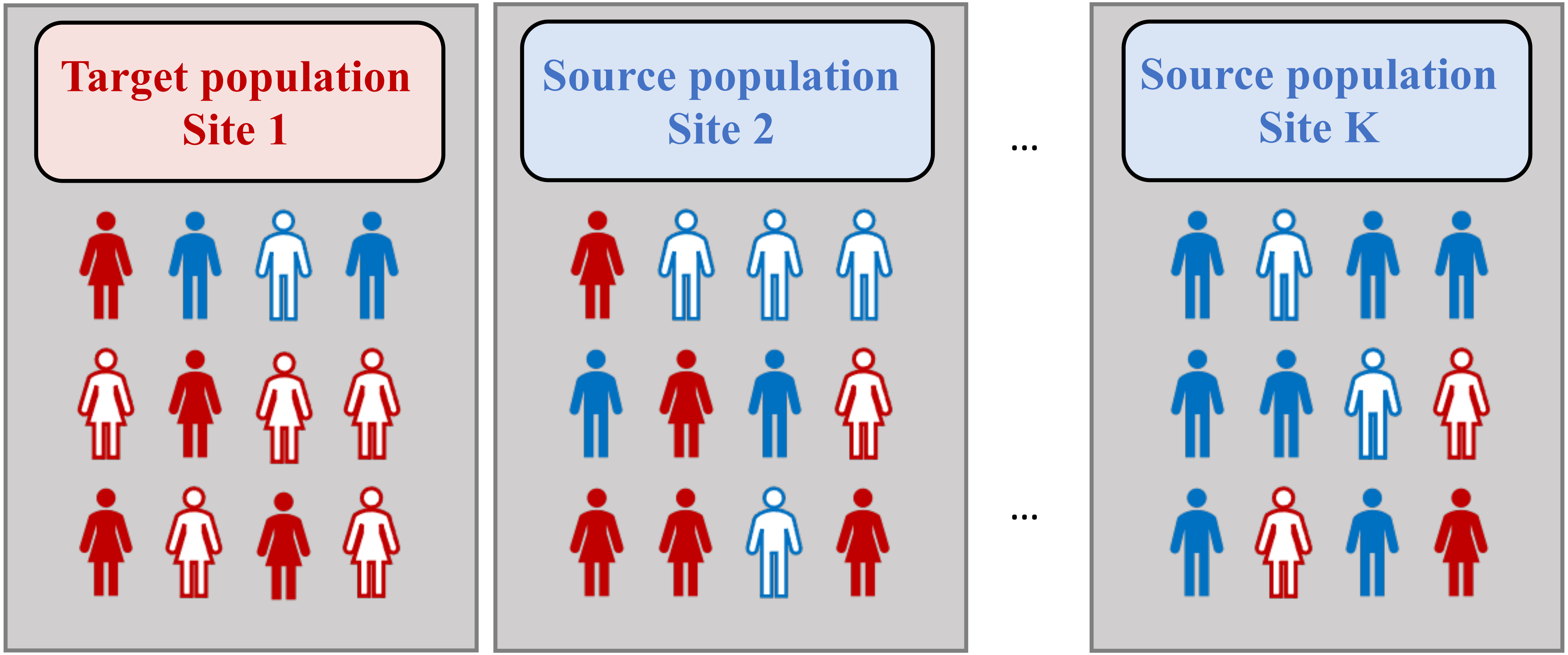
Promises of distributed research networks

- *OHDSI*
- *PCORNET*
- *FDA Sentinel*
- *4CE*
- *…*

✓ Enhance generalizability

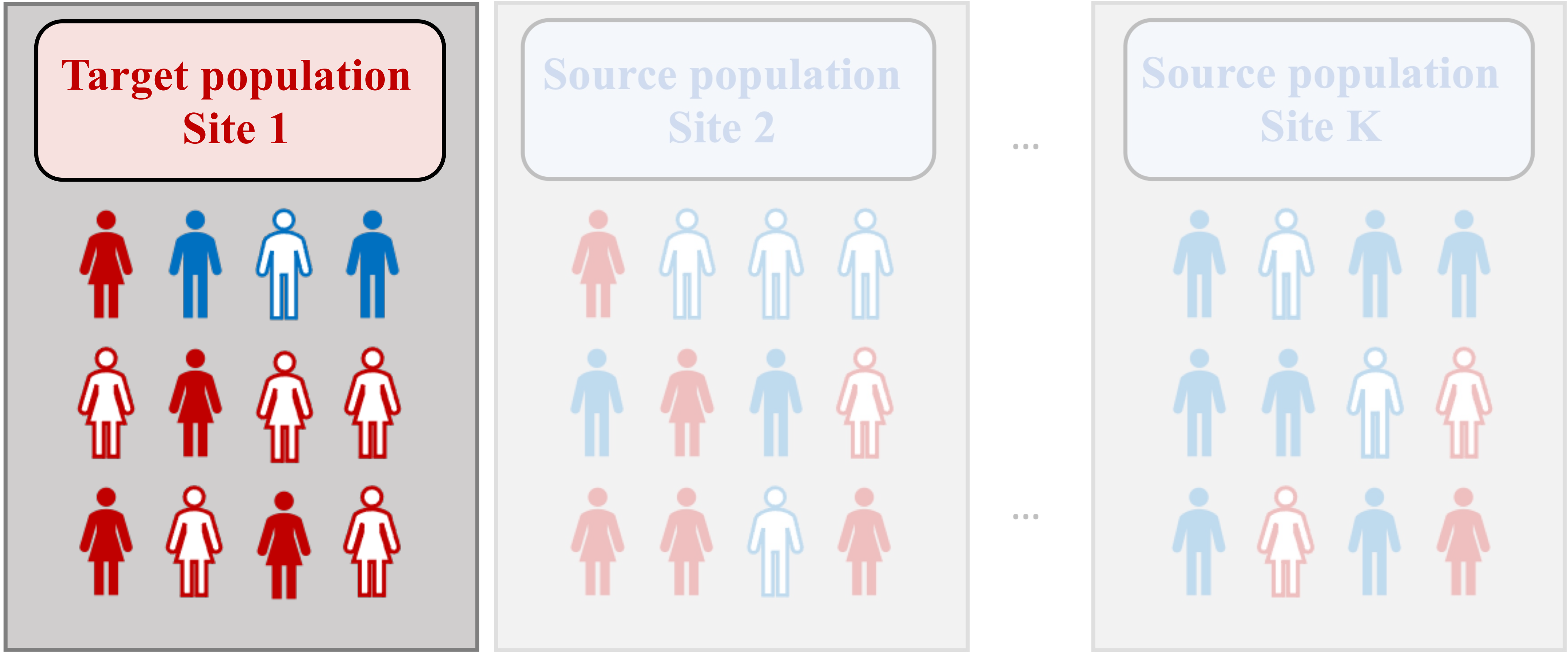
✓ Accelerate decision-making

✓ Study underrepresented populations, rare diseases, and exposures



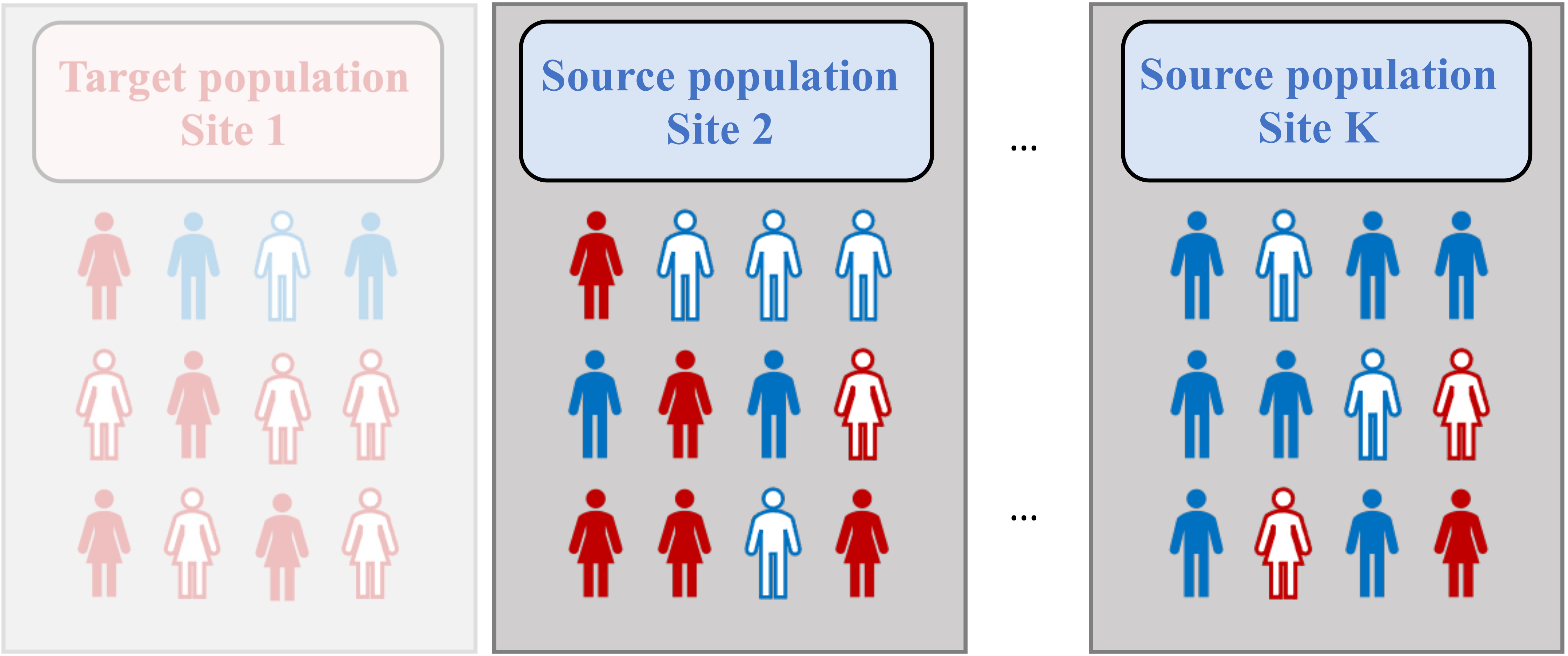
Question we seek to answer

In multi-source settings, how can we make optimal use of available data to make causal inferences for a target population of interest?



Question we seek to answer

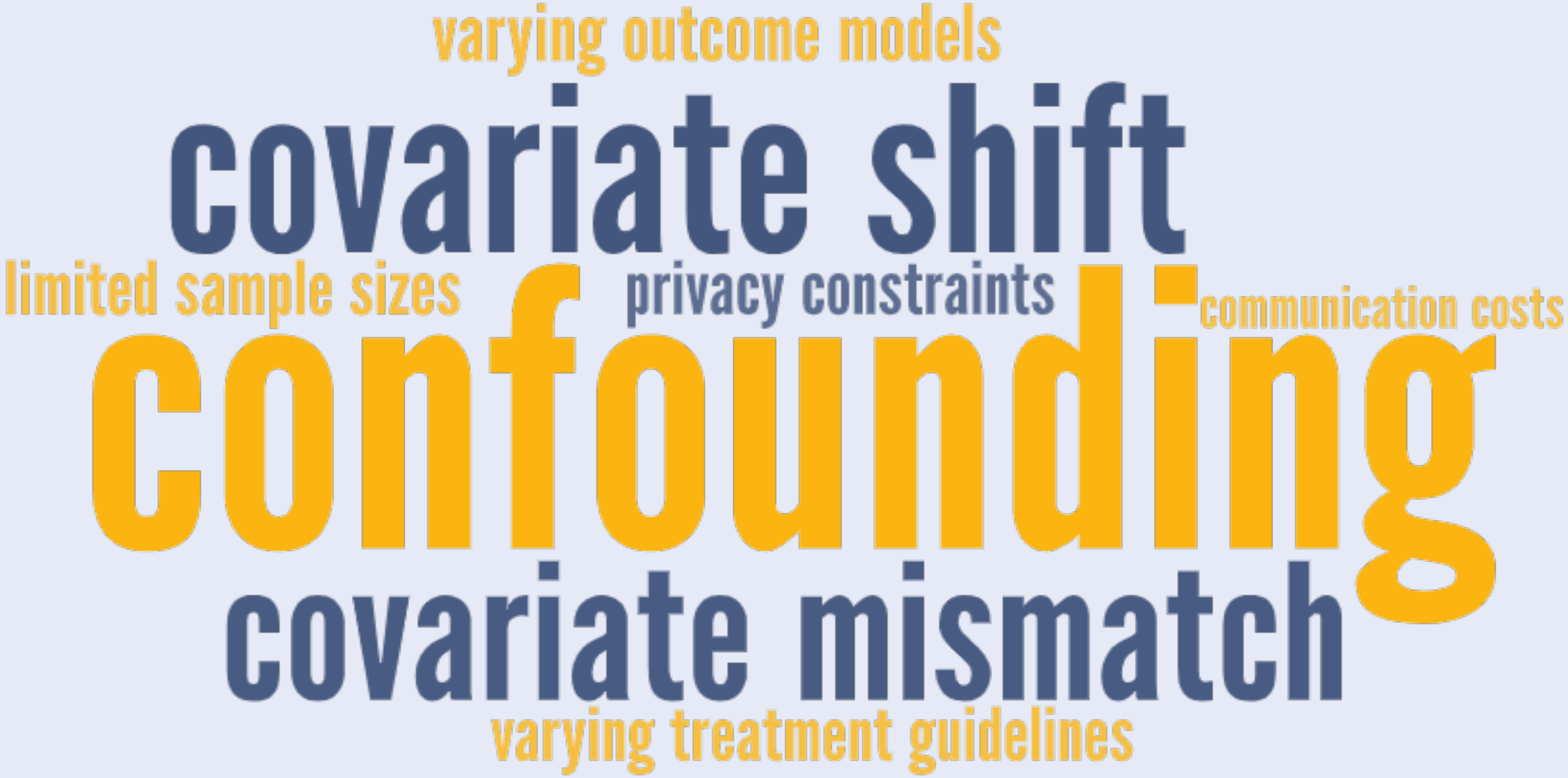
In multi-source settings, how can we make optimal use of available data to make causal inferences for a target population of interest?



Question we seek to answer

In multi-source settings, how can we make optimal use of available data to make causal inferences for a target population of interest?

Under multiple real-world constraints:



Workflow

Identification

- Estimand
- Assumption

Site-specific Estimation

- Density Ratio Weighting
- Multiply Robust Estimation
- Covariate Mismatch

Federated Estimation

- Adaptive Ensemble

Workflow

Identification

- Estimand
- Assumption

Site-specific Estimation

- Density Ratio Weighting
- Multiply Robust Estimation
- Covariate Mismatch

Federated Estimation

- Adaptive Ensemble

Target Estimand

$$\Delta_T = \mu_{1,T} - \mu_{0,T}, \text{ where } \mu_{a,T} = E\{Y_i(a) | R_i = T\} \text{ for } a \in \{0,1\}$$

Target Population

Outcome

Site Indicator

Binary Treatment

Identification Assumptions

- Consistency
- Mean exchangeability over treatment in target population
- Mean exchangeability over treatment in source population
- Positivity of treatment in target population
- Positivity of treatment in source population
- Mean exchangeability over site selection
- Positivity of site selection

Target Estimand

$$\Delta_T = \mu_{1,T} - \mu_{0,T}, \text{ where } \mu_{a,T} = E\{Y_i(a) | R_i = T\} \text{ for } a \in \{0,1\}$$

Target Population

Outcome

Site Indicator

Binary Treatment

Identification Assumptions

- **Consistency**
- **Mean exchangeability over treatment in target population**
- Mean exchangeability over treatment in source population
- **Positivity of treatment in target population**
- Positivity of treatment in source population
- Mean exchangeability over site selection
- Positivity of site selection

Workflow

Identification

- Estimand
- Assumption

Site-specific Estimation

- Density Ratio Weighting
- Multiply Robust Estimation
- Covariate Mismatch

Federated Estimation

- Adaptive Ensemble

Density Ratio Weighting

for adjusting covariate shift across sites

Existing methods
adjust for heterogeneity in covariate distributions (covariate shift) across
sites by the **inverse probability of selection weighting (IPSW)**

But
IPSW requires pooling target and source samples, which is often not possible due to **data privacy regulations**.

Density Ratio Weighting
for adjusting covariate shift across sites

Existing methods
adjust for heterogeneity in covariate distributions (covariate shift) across sites by the **inverse probability of selection weighting (IPSW)**

But
IPSW requires pooling target and source samples, which is often not possible due to **data privacy regulations**.

We
consider a **density ratio weighting** approach, which offers equivalent estimation without the need for direct data pooling.

Multiply Robust Estimation

for multiple, different models across sites

Existing methods
require **common models** to be specified across sites.

But
it is beneficial for investigators at different sites to incorporate **site-specific knowledge** when specifying candidate models.

Multiply Robust Estimation
for multiple, different models across sites

Existing methods
require **common models** to be specified across sites.

But
it is beneficial for investigators at different sites to incorporate **site-specific knowledge** when specifying candidate models.

We
We relax this requirement by adopting a **multiply robust** estimator, allowing investigators in each site to propose **multiple, different outcome and treatment models**.

Covariate Mismatch

adjusted by a new nuisance function $\tau_{a,k}$

Existing methods
assume **a common set of observed covariates**.

But
assumption rarely met due to **variations in local practices**, e.g., differing data collection standards and coding practices.

Covariate Mismatch

adjusted by a new nuisance function $\tau_{a,k}$

Existing methods
assume **a common set of observed covariates**.

But
assumption rarely met due to **variations in local practices**, e.g., differing data collection standards and coding practices.

We
We introduce a **new nuisance function $\tau_{a,k}$** which projects all site-specific estimates of conditional outcomes to a common hyperplane defined by shared effect modifiers across sites.

Workflow

Identification
- Estimand
- Assumption

Site-specific Estimation
- Density Ratio Weighting
- Multiply Robust Estimation
- Covariate Mismatch

Federated Estimation
- Adaptive Ensemble

Federated Estimation
by an adaptive ensemble method

Existing methods
- Target only
- Sample size weighting (SS)
- Inverse variance weighting (IVW)
- …

But
preventing **negative transfer** is critical when there are multiple, potentially biased source sites.

Federated Estimation
by an adaptive ensemble method

Existing methods
- Target only
- Sample size weighting (SS)
- Inverse variance weighting (IVW)
- …

But
preventing **negative transfer** is critical when there are multiple, potentially biased source sites.

We
We combine all site-specific estimates by an **adaptive ensemble** method; control for bias due to non-transportable site estimates while achieving **optimal efficiency**.

Treatment effect of percutaneous coronary intervention (PCI) on length of hospital stay for acute myocardial infarction (AMI) patients

- Target state: Maine
- Source states: 48 other continental states
- Coarsened covariates: Demographics
- Additional covariates: Comorbidities

Treatment effect of percutaneous coronary intervention (PCI) on length of hospital stay for acute myocardial infarction (AMI) patients

- Target state: Maine
- Source states: 48 other continental states
- Coarsened covariates: Demographics
- Additional covariates: Comorbidities

Estimator	Est.	(CI)
Target	−7.63	(−11.45 −3.81)
SS	−9.93	(−15.29 −4.56)
IVW	−8.94	(−9.47 −8.41)
AIPW−L1	−7.84	(−9.60 −6.09)
MR−L1	−7.49	(−9.82 −5.16)

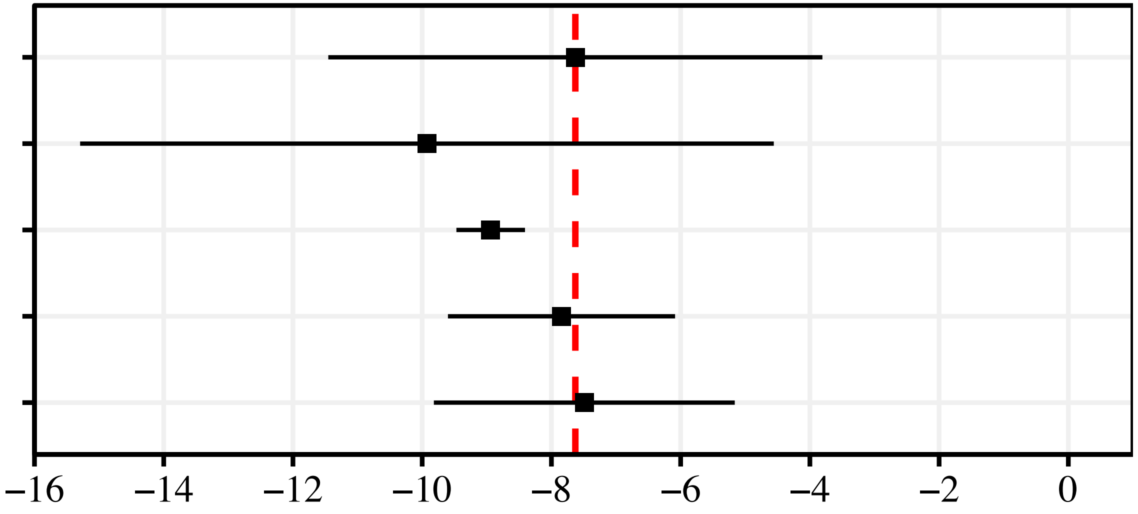


Figure: Estimates of PCI treatment effect in Maine with covariate mismatch in patient comorbidities

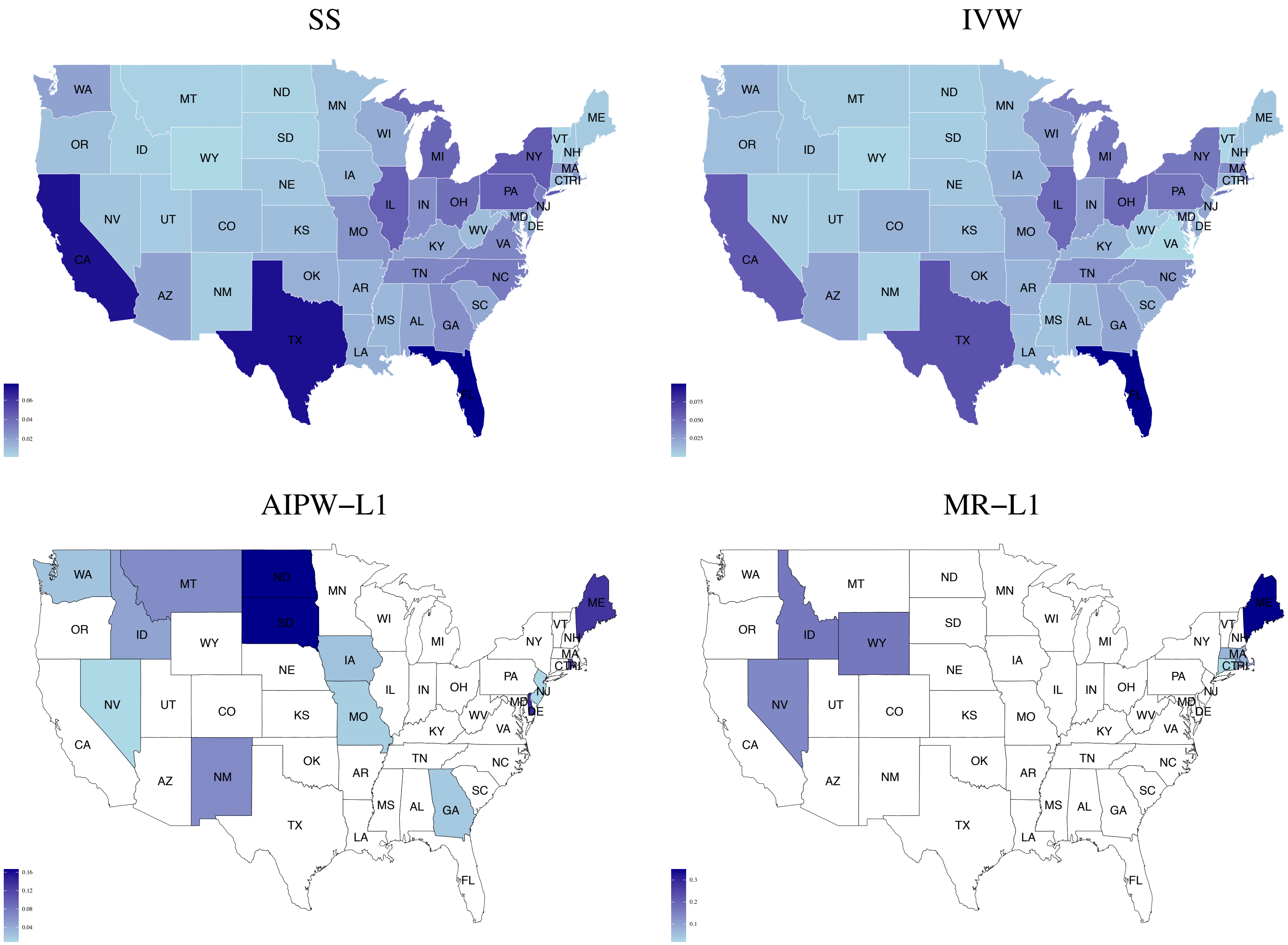


Figure: Federation weights across states for the PCI treatment effect in Maine with four federated estimators

Multiply Robust Federated Estimation of Targeted Average Treatment Effects

Larry Han
Department of Health Sciences
Northeastern University
Boston, MA 02115
lar.han@northeastern.edu

Zhu Shen
Department of Biostatistics
Harvard University
Boston, MA 02115
zhushen@g.harvard.edu

Jose Zubizarreta
Departments of Health Care Policy, Biostatistics, and Statistics
Harvard University
Boston, MA 02115
zubizarreta@hcp.med.harvard.edu

Paper: https://arxiv.org/abs/2309.12600

We thank Yige Li, Bijan Niknam, Zhenghao Zeng, Wei Li, feedback received at the 2023 American Causal Inference Conference and the 2023 Institute for Operations Research and the Management Sciences Annual Meeting, and the five anonymous reviewers for their helpful comments and input that greatly enhanced the paper.

Larry Han

Zhu Shen

Jose Zubizarreta