Federated Causal Inference in Heterogeneous Observational Data

Ruoxuan Xiong, Allison Koenecke, Michael Powell, Zhu Shen, Joshua Vogelstein, Susan Athey

Roadmap

- How can we perform causal analysis across multiple datasets with similar structure that cannot be combined?
- Some stability of conditional treatment effects across datasets (so there is a potential benefit to combining them)

1. Motivating example (Alpha-blockers)
2. Challenges in federated causal inference
3. Federated methods for causal inference
4. Asymptotic results
5. Empirical studies

Can alpha blockers improve patient outcomes?

- Prazosin shown to prevent cytokine storms in mice [Staedtke V et al. 2018]

- Question: do α_1-adrenergic receptors (α-blocker drugs) provide a prophylactic benefit for patients at risk of respiratory distress?
  - Ideal is run an RCT, but this is not available!

Can we use observational claims data to learn about effect of taking alpha blockers?

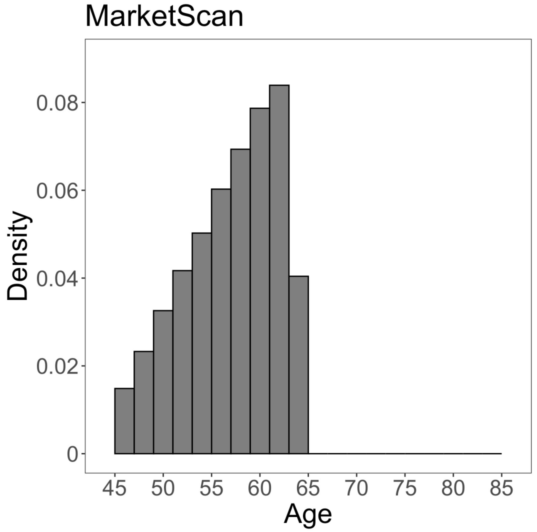
Patient ID	Patient Info	Date	Inpatient or Outpatient?	Diagnoses	Procedures	Prescribed Drugs & Duration	Expired?
	M / 58	Feb 2014	Doc's Office	BPH	Colonoscopy	tamsulosin 0.4mg / 30 days of pills	N
	M / 59	Jan 2015	Hospital	ARD, BPH	Ventilation	--	N
	F / 70	Dec 2015	Hospital	Cancer, Pneumonia	Ventilation	--	Y

More details about retrospective analysis

We have claims data from multiple sources

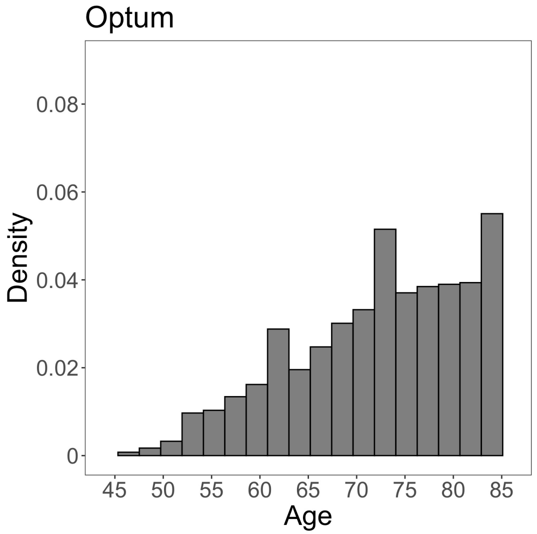
PROPRIETARY PATIENT DATA

MarketScan database



PROPRIETARY PATIENT DATA

Optum database



Challenges in federated causal inference

Challenges in causal inference using multiple datasets

- **Challenge 1**: Proprietary patient data cannot be combined at the individual level
- **Challenge 2**: Datasets are heterogeneous
  - Heterogeneity means demographics, confounders, propensity and outcome models can be different
- **Challenge 3**: Account for selection bias
- **Challenge 4**: Require both estimation and inference methods

➢Challenges 3 and 4 separate our work from the federated learning literature

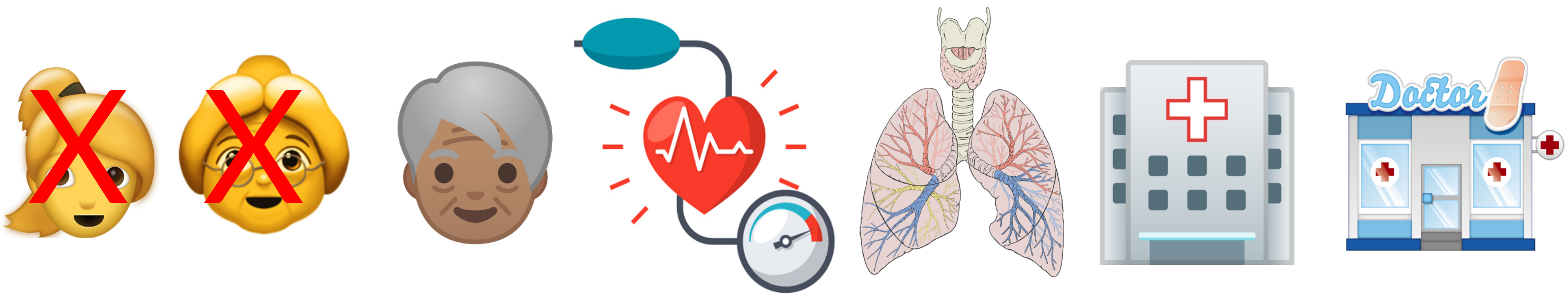
Our contribution

1. A systematic framework to federate point and variance estimates across datasets
    - Two main categories: IPW-MLE and AIPW; one supplementary category: MLE
    - Weight summary-level information cleverly depending on stability / model specification condition
    - Computationally efficient

2. Asymptotic guarantees for federated point and variance estimators
    - Federated point and variance estimators: Asymptotically the same as those using the combined individual-level data
    - Federated point estimator: Doubly robust, efficient, and asymptotic normal
    - Federated variance estimator: Consistency

3. A procedure to select federated methods on empirical datasets

Inclusion criteria & confounding



Observed Confounders
(Age, comorbidities)

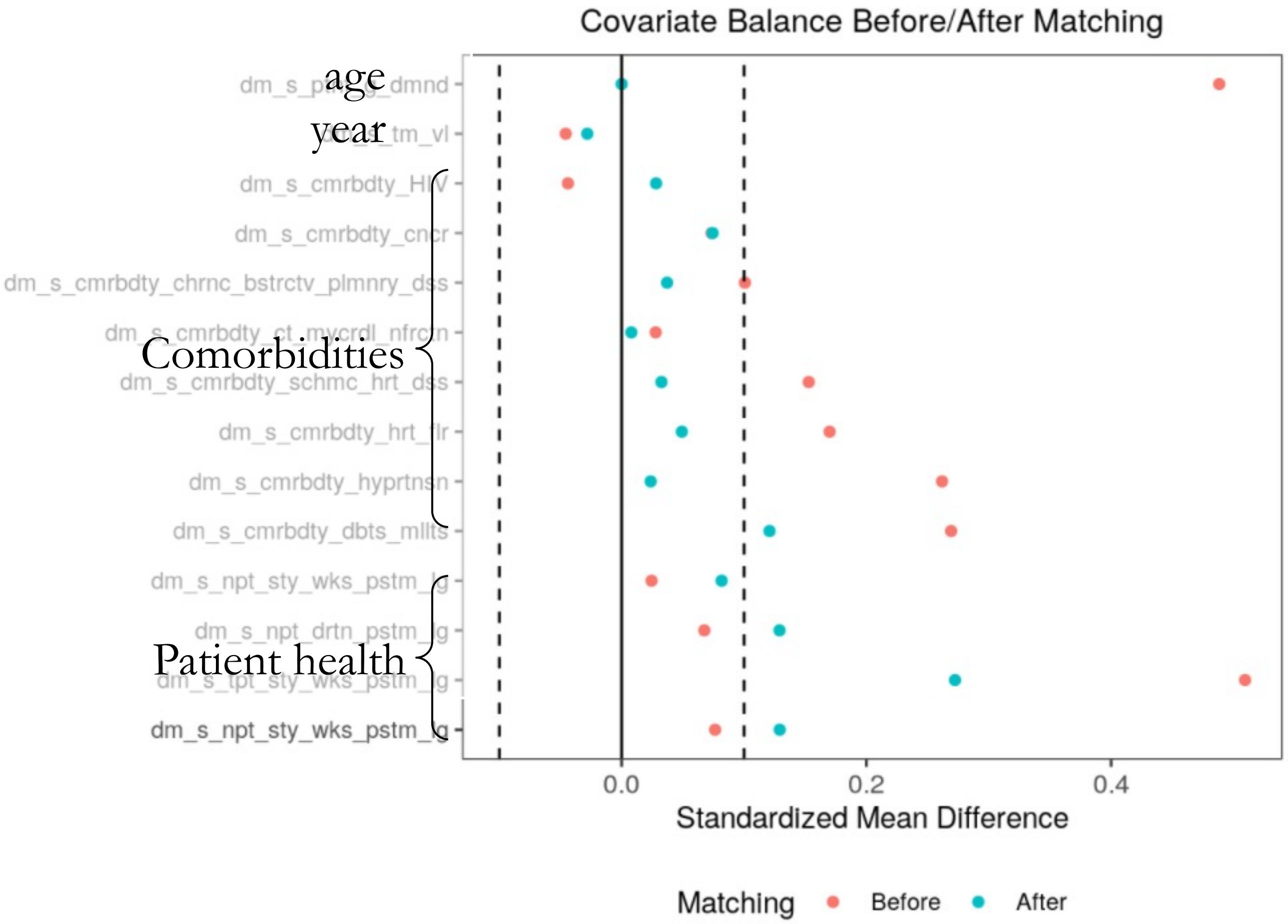
Treatment
(Taking Alpha-Blockers)

Outcome
(Ventilation / Dying)

Imbalance between the treated and control groups

- Covariates are imbalanced
  - Prostate problems tend to worsen with age
  - Treated patients are generally older
  - A larger fraction of treated patients have comorbidities and are less healthy

MarketScan



Account for confounders

- **This paper**: Make the best possible use of multiple datasets to estimate average treatment effect while adjusting for observed confounders
- Can get efficient, doubly robust estimates if we can accomplish these two goals:
  1. **Assignment model:** Estimate the relationship between treatment assignment and observed confounders, and use the resulting predictions to balance observed confounders across treatment and control groups
  2. **Outcome model:** Estimate the relationship between the outcome and observed confounders, e.g., age, comorbidities, general patients' health
- Challenges in health data
  1. Small, siloed datasets
  2. Many confounders

Assignment model

- The relationship between treatment assignment and observed confounders can be specified by
  - Parametric model
    - Model specification: e.g., $\log \frac{P(W=1|X)}{P(W=0|X)} = X^T \gamma_x$
      - W: Taking alpha−blockers
      - X: Age, comorbidities, general patients' health, …
    - Estimation: Maximum-likelihood estimator (MLE), e.g., $\hat{\gamma}_x = \arg\max_{\gamma_x} \sum_i \log \frac{P(W_i=1|X_i,\gamma_x)}{P(W_i=0|X_i,\gamma_x)}$
    - **This paper**: Leverage multiple datasets to improve the precision of $\hat{\gamma}_x$
  - Non-linear/Non-parametric model
    - Estimation: e.g., causal forests

Outcome model

- The relationship between treatment assignment and observed confounders can be specified by
  - Parametric model
    - Model specification: e.g., $\log\frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + X^T\beta_x$
      - Y: Ventilation (followed by death)
      - β_w: The effect of taking alpha-blockers in reducing the log-odds of adverse outcome
    - Estimation: Inverse-propensity weighted maximum-likelihood estimator (IPW-MLE, Wooldridge, 2002, 2007) or maximum-likelihood estimator (MLE)
    - **This paper**: Leverage multiple datasets to improve the precision of $\hat{\beta}_w$ and $\hat{\beta}_x$
  - Non-linear/Non-parametric model
    - Estimation: e.g., causal forests

Inverse-propensity weighted maximum-likelihood estimator

- Inverse-propensity weighted maximum-likelihood estimator (IPW-MLE) *balances* observed confounders across treatment and control groups

$$(\hat{\beta}_w, \hat{\beta}_x) = \arg\max_{\beta_w,\beta_x} \sum_i \varpi_i \log \frac{P(Y_i=1|X_i,W_i,\beta_x,\beta_w)}{P(Y_i=0|X_i,W_i,\beta_x,\beta_w)}$$

- ϖ_i: Weight for patient i
  - ATE weighting: $\varpi_i = \frac{W_i}{e(X_i)} + \frac{1-W_i}{1-e(X_i)}$
  - ATT weighting: $\varpi_i = W_i + (1 - W_i)\frac{e(X_i)}{1-e(X_i)}$
  - $e(X_i) = P(W_i = 1|X_i)$: Propensity score for patient i

Double robustness property of IPW-MLE

- IPW-MLE is a doubly robust estimator (Wooldridge, 2007, Lumley, 2011)
  - $\hat{\beta}_w$ and $\hat{\beta}_x$ are consistent if
    - We have observed relevant covariates
    - At least one of the propensity and outcome models is correctly specified

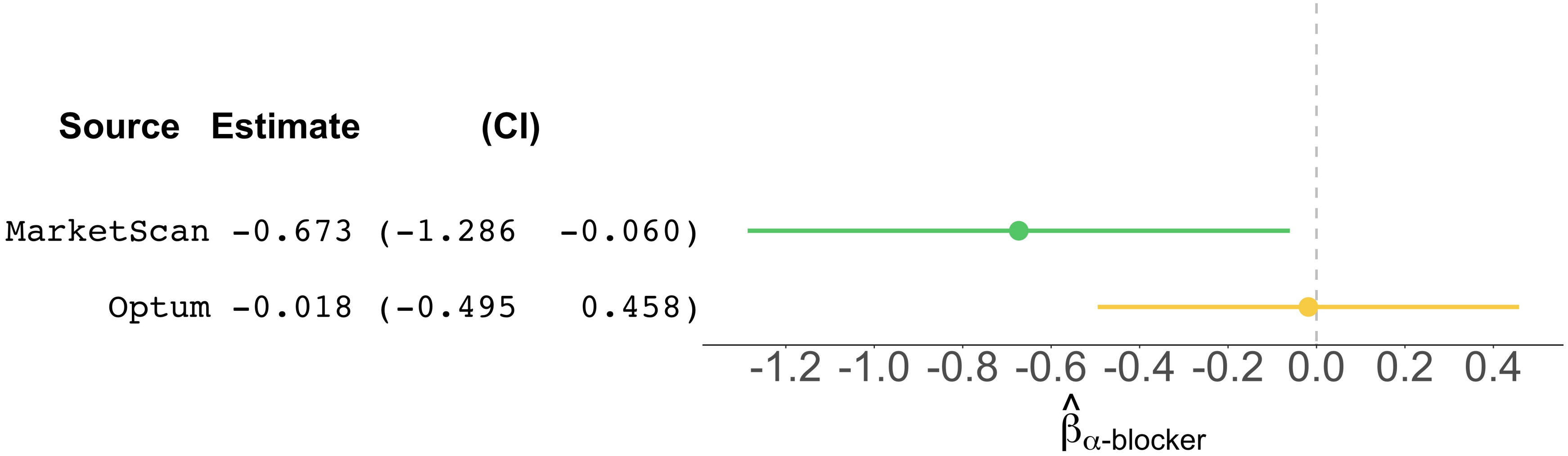
$$(\hat{\beta}_w, \hat{\beta}_x) = \arg\max_{\beta_w, \beta_x} \sum_i \varpi_i \log \frac{P(Y_i=1|X_i, W_i, \beta_x, \beta_w)}{P(Y_i=0|X_i, W_i, \beta_x, \beta_w)}$$

AIPW: Another doubly robust estimator

- Augmented inverse-propensity weighted (AIPW) estimator is also doubly robust
  - $\hat{\tau}_{ate} = \frac{1}{n}\sum_i \left(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + \frac{W_i}{\hat{e}(X_i)}\left(Y_i - \hat{\mu}_1(X_i)\right) - \frac{1-W_i}{1-\hat{e}(X_i)}\left(Y_i - \hat{\mu}_0(X_i)\right)\right)$, where $\hat{\mu}_d(X_i) = E[Y_i(d)]$
  - **This paper**: Leverage multiple datasets to improve the precision of $\hat{\mu}_d(X_i)$ and $\hat{e}(X_i)$ (that are estimated parametrically)
    - Built on the results of MLE and IPW-MLE

Results from multiple sources

- Taking alpha-blockers seems to reduce the log odds of the adverse outcome on both datasets
- $\hat{\beta}_w$ from IPW-MLE on MarketScan and Optum



- **This paper**: Can we narrow down the confidence intervals by using the information in two datasets?

Challenges in federated methods for IPW-MLE

- Recall IPW-MLE estimates β_w and β_x by maximizing the inverse propensity weighted likelihood function
  - The estimation error of the propensity model carries over to the estimation of β_w and β_x
  - The precision of $\hat{\beta}_w$ and $\hat{\beta}_x$ depend on the (weighted) gradient and Hessian of propensity and outcome models in a complex manner
  - Key challenges in federated methods for IPW-MLE: Need to account for many conditions related to model specification and heterogeneity across datasets
  - What happens if we ignore these challenges and use off-the-shelf methods, e.g., inverse variance weighting (IVW)?

Asymptotic distribution of IPW-MLE

Lemma 1. Suppose the regularity conditions for the parametric propensity and outcome models hold. As the sample size $n \to \infty$, the IPW-MLE $\hat{\beta}$ is consistent and asymptotically normal

$$n^{1/2} \cdot (\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V_\beta)$$

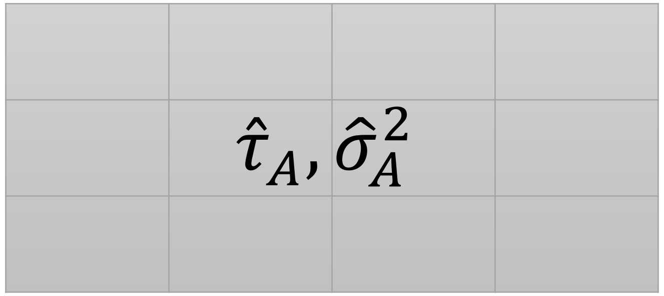
where $V_\beta = A_{\beta,\varpi}^{-1} \cdot \left(D_{\beta,\varpi} - M_{\beta,\varpi,\gamma}\right) \cdot A_{\beta,\varpi}^{-1}$ and

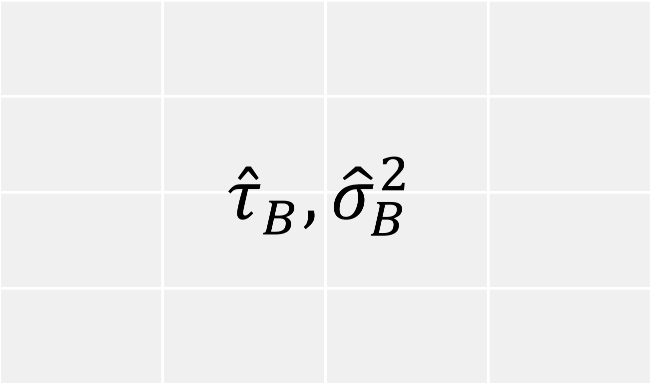
- $A_{\beta,\varpi}$: weighted Hessian of the *outcome* model
- $D_{\beta,\varpi}$: weighted outer product of the gradient of the *outcome* model
- $M_{\beta,\varpi,\gamma} = C_{\beta,\varpi,\gamma,1} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T + C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,1}^T - C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T$
  - $V_\gamma = A_\gamma^{-1} \cdot B_\gamma \cdot A_\gamma^{-1}$
    - A_γ: Hessian of the *propensity* model
    - B_γ: outer product of the gradient of the *propensity* model
  - $C_{\beta,\varpi,\gamma,1}$ and $C_{\beta,\varpi,\gamma,2}$: weighted outer products of the gradient of the *propensity* model and the gradient of the *outcome* model

➢ Matrices in V_β depend on how the *propensity* and *outcome* models are specified, whether they are correctly specified, and whether ATE or ATT weighting is used

More details

A popular approach: Inverse-variance weighting (IVW)





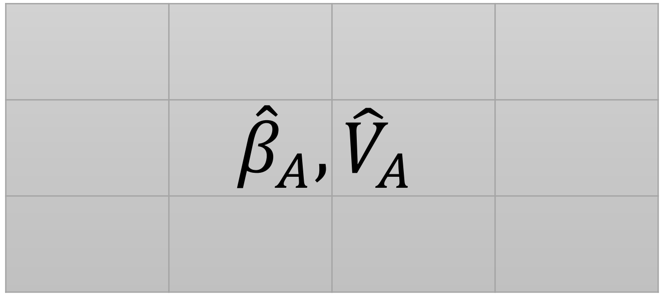
Meta-analysis (no covariates):

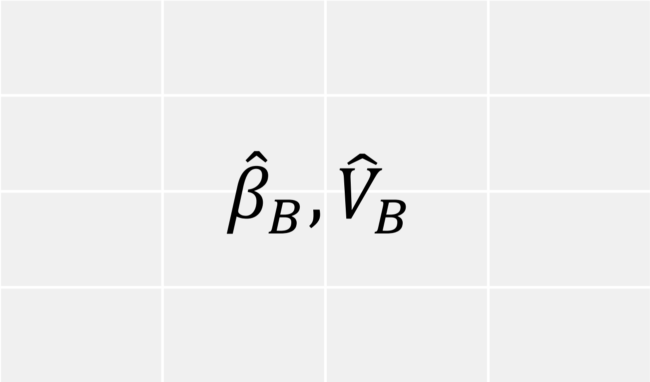
1. Estimate the treatment effect τ and variance σ_A^2 on each dataset
2. Combine coefficients by inverse variance weighting

$$\hat{\tau}_{ivw} = (\hat{\sigma}_A^{-2} + \hat{\sigma}_B^{-2})^{-1}(\hat{\sigma}_A^{-2}\hat{\tau}_A + \hat{\sigma}_B^{-2}\hat{\tau}_B)$$

DerSimonian and Laird (1986), Whitehead and Whitehead (1991), Sutton and Higgins (2008)

A popular approach: Inverse-variance weighting (IVW)





Linear regression (adjust for covariates):

1. Estimate coefficients β and variance V on each dataset
2. Combine coefficients by inverse variance weighting

$$\hat{\beta}_{ivw} = \left(\hat{V}_A^{-1} + \hat{V}_B^{-1}\right)^{-1}\left(\hat{V}_A^{-1}\hat{\beta}_A + \hat{V}_B^{-1}\hat{\beta}_B\right)$$

Du, Han, and Chen (2004), Karr, Lin, Sanil, and Reiter (2005) (machine learning and security)

Inverse-variance weighting (IVW)

$\hat{\beta}_A, \hat{V}_A$

$$\hat{\beta}_{ivw} = \left(\hat{V}_A^{-1} + \hat{V}_B^{-1}\right)^{-1}\left(\hat{V}_A^{-1}\hat{\beta}_A + \hat{V}_B^{-1}\hat{\beta}_B\right)$$

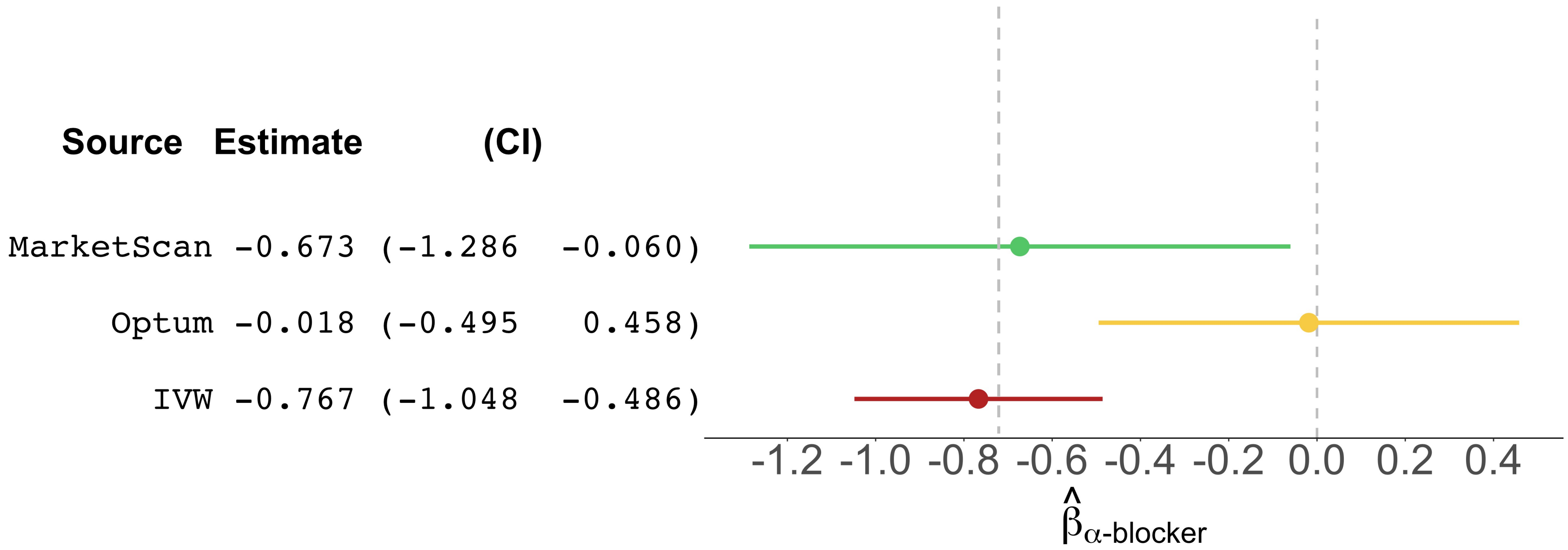
Pro: Inverse-variance weighting average has the least variance among all averages

$\hat{\beta}_B, \hat{V}_B$

Con: Inverse-variance weighting does not account for selection bias

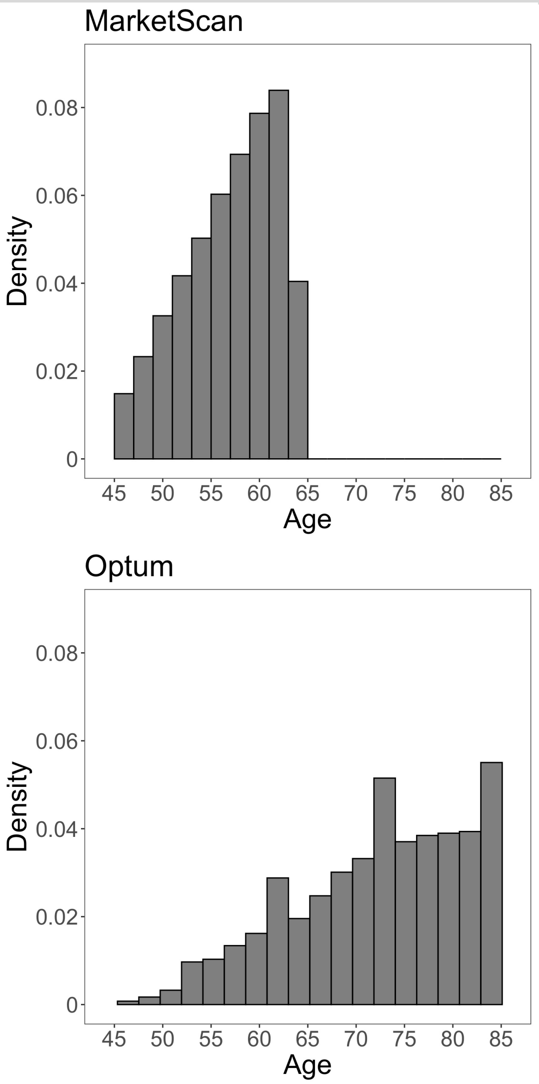
Combining heterogeneous patient data by IVW

- Concern: The IVW pooled IPW-MLE estimate lies outside of those on MarketScan and Optum



What is wrong with IVW?

- MarketScan and Optum have different age populations
- Coefficients and variance-covariance matrices across datasets are heterogeneous



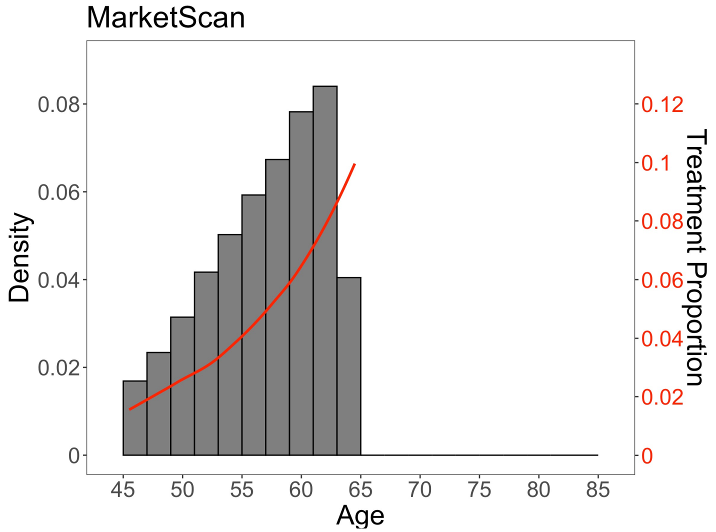
What is wrong with IVW?

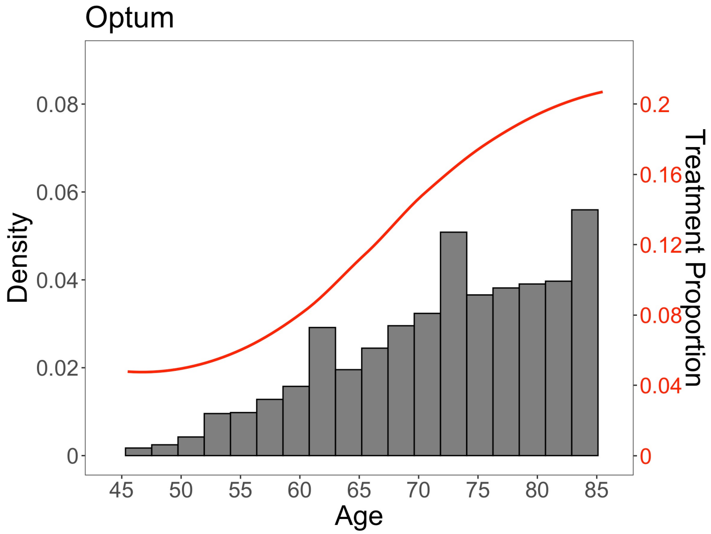
- MarketScan and Optum have different age populations
- Coefficients and variance-covariance matrices across datasets are heterogeneous
  - Coefficient of age in outcome model switches sign (controlling for other covariates)
  - Covariance between treatment and age switches sign

$$\hat{\beta}_{ivw} = \left(\hat{V}_M^{-1} + \hat{V}_O^{-1}\right)^{-1}\left(\hat{V}_M^{-1}\hat{\beta}_M + \hat{V}_O^{-1}\hat{\beta}_O\right) = \begin{bmatrix} -0.71 \\ 1.42 \end{bmatrix}$$

$$\hat{\beta}_M = \begin{bmatrix} \hat{\beta}_{M,w} \\ \hat{\beta}_{M,age} \end{bmatrix} = \begin{bmatrix} -0.67 \\ 2.03 \end{bmatrix} \qquad \hat{V}_M^{-1} = \begin{bmatrix} 51.6 & -28.6 \\ -28.6 & 474.02 \end{bmatrix}$$

$$\hat{\beta}_O = \begin{bmatrix} \hat{\beta}_{O,w} \\ \hat{\beta}_{O,age} \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.15 \end{bmatrix} \qquad \hat{V}_O^{-1} = \begin{bmatrix} 55.34 & 14.61 \\ 14.61 & 187.08 \end{bmatrix}$$





What is wrong with IVW?

- Coefficients on two data sets are $\beta_M = \begin{bmatrix} \beta_{M,w} \\ \beta_{M,age} \end{bmatrix}$ and $\beta_O = \begin{bmatrix} \beta_{O,w} \\ \beta_{O,age} \end{bmatrix}$
  - $\beta_{M,age} > 0$ and $\beta_{O,age} < 0$ (in our application, $\beta_{M,age} = 2.03$ and $\beta_{O,age} = -0.15$)
- Inverse variance-covariance matrix is $V_M^{-1} = \begin{bmatrix} v_{M,11} & v_{M,12} \\ v_{M,12} & v_{M,22} \end{bmatrix}$ and $V_O^{-1} = \begin{bmatrix} v_{O,11} & v_{O,12} \\ v_{O,12} & v_{O,22} \end{bmatrix}$
  - $v_{M,12} < 0$ and $v_{O,12} > 0$ (in our application, $v_{M,12} = -28.6$ and $v_{O,12} = 14.61$)
- Without loss of generality, assume $\beta_{M,w} < \beta_{O,w}$ (in our application, $\beta_{M,w} = -0.67$ and $\beta_{O,w} = -0.02$)

$$\beta_{ivw} = (V_M^{-1} + V_O^{-1})^{-1}(V_M^{-1}\beta_M + V_O^{-1}\beta_O)$$
$$= \beta_{M,w} + \frac{1}{c}\cdot\left(\underbrace{\left((v_{M,22} + v_{O,22})\cdot v_{O,11} - (v_{M,12} + v_{O,12})\cdot v_{O,12}\right)}_{>0}\cdot\underbrace{(\beta_{O,w} - \beta_{M,w})}_{>0} + \underbrace{(v_{O,22}v_{M,12} - v_{M,22}v_{O,12})}_{<0}\cdot\underbrace{(\beta_{M,age} - \beta_{O,age})}_{>0}\right)$$

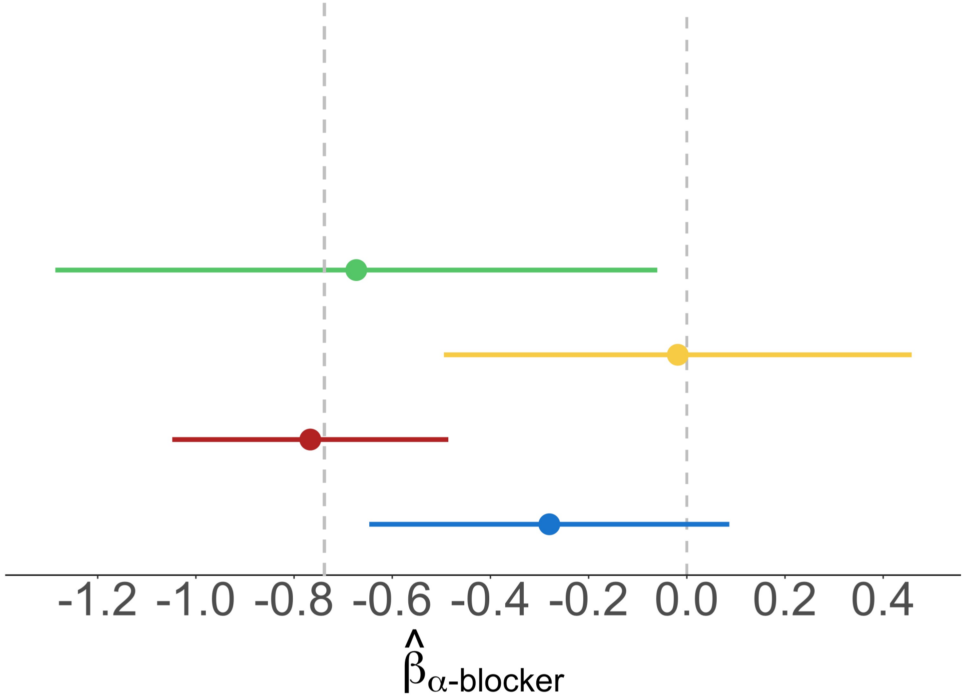
➢ $\beta_{ivw} < \beta_{M,w}$ when $\beta_{M,age} - \beta_{O,age} \gg \beta_{O,w} - \beta_{M,w}$
  ➢ In our application, $\beta_{M,age} - \beta_{O,age} = 2.18$ and $\beta_{O,w} - \beta_{M,w} = 0.65$

$c = (v_{M,11} + v_{O,11})\cdot(v_{M,22} + v_{O,22}) - (v_{M,12} + v_{O,12})^2 > 0$

Results from our federated IPW-MLE

- The federated coefficient from our approach lies between those from MarketScan and Optum
- The confidence interval is narrower than those from MarketScan and Optum

Source	Estimate	(CI)	
MarketScan	-0.673	(-1.286	-0.060)
Optum	-0.018	(-0.495	0.458)
IVW	-0.767	(-1.048	-0.486)
Federated IPW-MLE	-0.280	(-0.647	0.087)



Federated methods for causal inference

Framework for federated causal inference

1. Input

Dataset A

Dataset B

2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Determine which conditions hold

- Stable 🚫
- Correct Spec ✅

4. Federate propensity models across datasets

$$\begin{pmatrix} \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}}} & \mathbf{0}_{|\mathcal{S}|\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}^\complement}} & \mathbf{0}_{|\mathcal{S}|\times S_{k+1}^{D}} \\ \mathbf{0}_{S_1^{k-1}\times|\mathcal{S}|} & \mathbf{0}_{S_1^{k-1}\times S_1^{k-1}} & \mathbf{0}_{S_1^{k-1}\times S_k^{k}} & \mathbf{0}_{S_1^{k-1}\times S_{k+1}^{D}} \\ \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}^\complement},\boldsymbol{\beta}_{\mathcal{S}}} & \mathbf{0}_{S_k^{k}\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}^\complement},\boldsymbol{\beta}_{\mathcal{S}^\complement}} & \mathbf{0}_{S_k^{k}\times S_{k+1}^{D}} \\ \mathbf{0}_{S_{k+1}^{D}\times|\mathcal{S}|} & \mathbf{0}_{S_{k+1}^{D}\times S_1^{k-1}} & \mathbf{0}_{S_{k+1}^{D}\times S_k^{k}} & \mathbf{0}_{S_{k+1}^{D}\times S_{k+1}^{D}} \end{pmatrix}$$

6. Output

$\hat{\beta}_{fed}, \hat{V}_{fed}$

5. Federate outcome models across datasets

Estimate propensity and outcome models for each dataset

1. Estimate propensity model $P(W = 1|X)$ for each dataset
   - Parametric model: $P(W = 1|X) = e(X, \gamma)$
   - Estimate γ by maximizing the likelihood function (MLE)

2. Estimate outcome model $f(Y|X, W)$ for each dataset
   - Parametric model: $f(Y|X, W, \beta)$
   - Estimate β by maximizing the inverse-propensity weighted likelihood function (IPW-MLE)

➢ Estimated propensity and outcome models can be used as the input of AIPW to estimate ATE/ATT for each dataset

- AIPW is consistent even if one of the propensity and outcome models is misspecified

Framework for federated causal inference

1. Input

Dataset A

Dataset B

2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Determine which conditions hold

- Stable 🚫
- Correct Spec ✅

4. Federate propensity models across datasets

$$\begin{pmatrix} \hat{\mathrm{H}}^{(k)}_{\beta_S,\beta_S} & 0_{|S|\times S_1^{k-1}} & \hat{\mathrm{H}}^{(k)}_{\beta_S,\beta_{S^C}} & 0_{|S|\times S_{k+1}^D} \\ 0_{S_1^{k-1}\times|S|} & 0_{S_1^{k-1}\times S_1^{k-1}} & 0_{S_1^{k-1}\times S_k^k} & 0_{S_1^{k-1}\times S_{k+1}^D} \\ \hat{\mathrm{H}}^{(k)}_{\beta_{S^C},\beta_S} & 0_{S_k^k\times S_1^{k-1}} & \hat{\mathrm{H}}^{(k)}_{\beta_{S^C},\beta_{S^C}} & 0_{S_k^k\times S_{k+1}^D} \\ 0_{S_{k+1}^D\times|S|} & 0_{S_{k+1}^D\times S_1^{k-1}} & 0_{S_{k+1}^D\times S_k^k} & 0_{S_{k+1}^D\times S_{k+1}^D} \end{pmatrix}$$

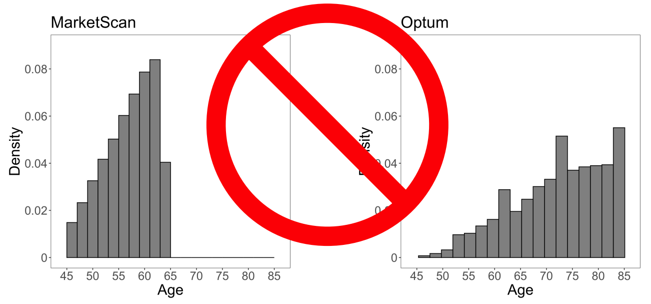
5. Federate outcome models across datasets

6. Output

$\hat{\beta}_{fed}, \hat{V}_{fed}$

Select conditions to impose (will return to see how)

- Stability conditions across datasets



  - Whether the set of covariates and their joint distribution are the same
  - Whether parameters in the propensity/outcome model are the same

$$\beta_M = \begin{bmatrix} \beta_{M,w} \\ \beta_{M,age} \end{bmatrix} = \begin{bmatrix} -0.57 \\ 2.03 \end{bmatrix} \quad \beta_O = \begin{bmatrix} \beta_{O,w} \\ \beta_{O,age} \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.15 \end{bmatrix}$$

- Model specification conditions
  - Whether the propensity/outcome model is correctly specified

True model: $\log \frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W \cdot \beta_w + age \cdot \beta_{age}$

Framework for federated causal inference

1. Input

Dataset A

Dataset B

2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Determine which conditions hold

- Stable 🚫
- Correct Spec ✅

4. Federate propensity models across datasets

$$\begin{pmatrix} \hat{\mathbf{H}}^{(k)}_{\beta_S,\beta_S} & \mathbf{0}_{|S|\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\beta_S,\beta_{S^\complement}} & \mathbf{0}_{|S|\times S_{k+1}^D} \\ \mathbf{0}_{S_1^{k-1}\times|S|} & \mathbf{0}_{S_1^{k-1}\times S_1^{k-1}} & \mathbf{0}_{S_1^{k-1}\times S_k^k} & \mathbf{0}_{S_1^{k-1}\times S_{k+1}^D} \\ \hat{\mathbf{H}}^{(k)}_{\beta_{S^\complement},\beta_S} & \mathbf{0}_{S_k^k\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\beta_{S^\complement},\beta_{S^\complement}} & \mathbf{0}_{S_k^k\times S_{k+1}^D} \\ \mathbf{0}_{S_{k+1}^D\times|S|} & \mathbf{0}_{S_{k+1}^D\times S_1^{k-1}} & \mathbf{0}_{S_{k+1}^D\times S_k^k} & \mathbf{0}_{S_{k+1}^D\times S_{k+1}^D} \end{pmatrix}$$

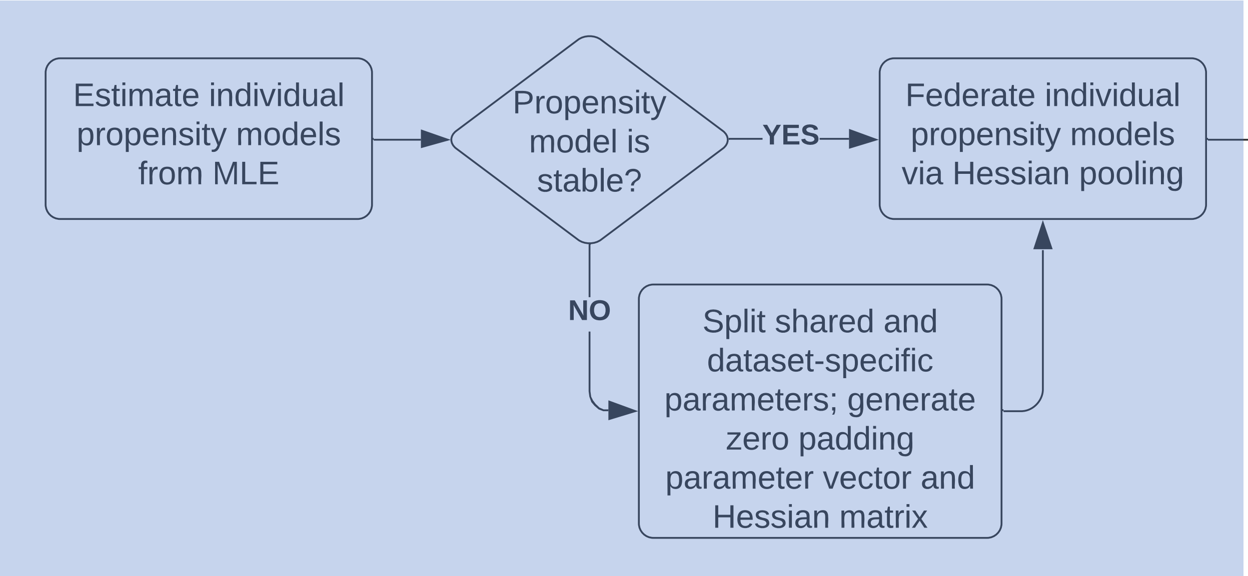
5. Federate outcome models across datasets

6. Output

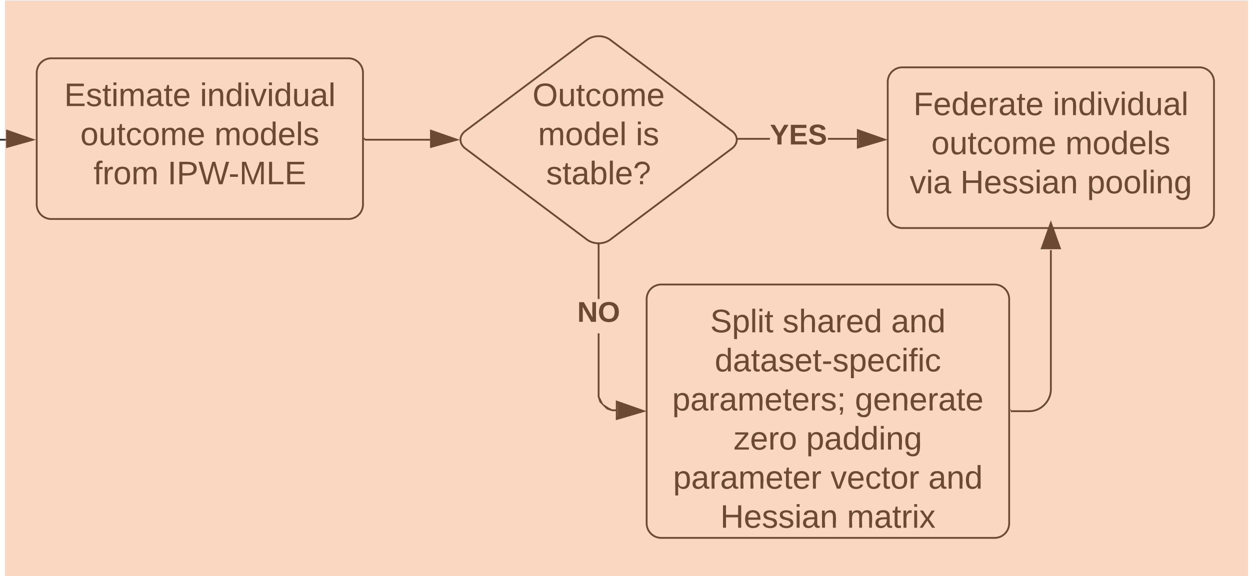
$\hat{\beta}_{fed}, \hat{V}_{fed}$

Flowchart for coefficient federation in IPW-MLE

Propensity model



Outcome model



$$\hat{\gamma}_{fed} = \left(\hat{H}^*_{\gamma,A} + \hat{H}^*_{\gamma,B}\right)^{-1}\left(\hat{H}^*_{\gamma,A}\hat{\gamma}^*_A + \hat{H}^*_{\gamma,B}\hat{\gamma}^*_B\right)$$

- $\hat{\gamma}^*_k$ and $\hat{H}^*_{\gamma,k}$ are (zero-padded) coefficient and Hessian in the propensity model on dataset k
- Conceptually we only federate the coefficients that are shared/stable across datasets

1. Use federated propensity model $\hat{e}_{fed}(x)$ to estimate federated IPW weight ϖ_{fed}
2. Use ϖ_{fed} to estimate outcome model for each dataset

$$\hat{\beta}_{fed} = \left(\hat{H}^*_{\beta,A} + \hat{H}^*_{\beta,B}\right)^{-1}\left(\hat{H}^*_{\beta,A}\hat{\beta}^*_A + \hat{H}^*_{\beta,B}\hat{\beta}^*_B\right)$$

- $\hat{\beta}^*_k$ and $\hat{H}^*_{\beta,k}$ are (zero-padded) coefficient and Hessian in the outcome model on dataset k
- Conceptually we only federate the coefficients that are shared/stable across datasets

Variance federation in IPW-MLE

- Recall the asymptotic variance of $\hat{\beta}$ is $V_\beta = A_{\beta,\varpi}^{-1} \cdot \left(D_{\beta,\varpi} - M_{\beta,\varpi,\gamma}\right) \cdot A_{\beta,\varpi}^{-1}$,

where $M_{\beta,\varpi,\gamma} = C_{\beta,\varpi,\gamma,1} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T + C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,1}^T - C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T$

- We seek to estimate the federated variance $V_{\beta,fed}$

1. Use $\hat{\gamma}_{fed}$ to estimate matrices in V_γ for each dataset
2. Use sample-size weighting to combine matrices across datasets and obtain $\hat{V}_{\gamma,fed}$

→

1. Use $\hat{\beta}_{fed}$ and $\hat{\gamma}_{fed}$ to estimate $A_{\beta,\varpi}$, $C_{\beta,\varpi,\gamma,1}$, $C_{\beta,\varpi,\gamma,2}$, $D_{\beta,\varpi}$ for each dataset
2. Use sample-size weighting to combine these matrices across datasets and obtain $\hat{V}_{\beta,fed}$

Federate individual propensity/outcome models

<table>
<tr><th>Description</th><th>Assume Stable Known Propensity and Stable Outcome Model (IPW-MLE #1)</th><th>Assume Stable Propensity and Stable Outcome Model (IPW-MLE #2)</th><th>Assume Stable Misspecified Propensity and Stable Outcome Model (IPW-MLE #3)</th><th>Assume Unstable Propensity or Unstable Outcome Model (IPW-MLE #4)</th></tr>
<tr><td>Stable Covariate Distribution</td><td>yes or no</td><td>yes or no</td><td>yes or no</td><td>yes or no</td></tr>
<tr><td>Known Propensity</td><td>yes</td><td>no</td><td>no</td><td>yes or no</td></tr>
<tr><td>Stable Propensity Model</td><td>yes</td><td>yes</td><td>yes</td><td>yes or no</td></tr>
<tr><td>Stable Outcome Model</td><td>yes</td><td>yes</td><td>yes</td><td>yes or no</td></tr>
<tr><td>Correct Propensity Model Specification</td><td>yes</td><td>yes</td><td>no</td><td>yes or no</td></tr>
<tr><td>Correct Outcome Model Specification</td><td>yes or no</td><td>yes or no</td><td>yes or no</td><td>yes or no</td></tr>
<tr><td>Sample Size Assumption</td><td>yes or no</td><td>yes or no</td><td>yes or no</td><td>yes</td></tr>
<tr><td>Coefficient $\boldsymbol{\beta}$ federation</td><td>(1) Estimate $\boldsymbol{\beta}^{(k)}$ using $\boldsymbol{\gamma}_0$; (2) Federate $\hat{\boldsymbol{\beta}}^{(k)}$ by Hessian weighting.</td><td colspan="2">(1) Federate $\hat{\boldsymbol{\gamma}}^{(k)}$ by Hessian weighting; (2) Estimate $\boldsymbol{\beta}^{(k)}$ using $\hat{\boldsymbol{\gamma}}^{\text{fed}}$; (3) Federate $\hat{\boldsymbol{\beta}}^{(k)}$ by Hessian weighting.</td><td>Same federation procedure with $\hat{\boldsymbol{\gamma}}^{\text{pad},(k)}$ and $\hat{\mathbf{H}}_{\boldsymbol{\gamma}}^{\text{pad},(k)}$ if propensity models are unstable and estimated, and with $\hat{\boldsymbol{\beta}}^{\text{pad},(k)}$ and $\hat{\mathbf{H}}_{\boldsymbol{\beta}}^{\text{pad},(k)}$ if outcomes models are unstable</td></tr>
<tr><td rowspan="3">Variance $\mathbf{V}_{\boldsymbol{\beta}}$ federation</td><td rowspan="2">$\mathbf{V}_{\boldsymbol{\beta}} = \mathbf{A}_{\boldsymbol{\beta},\varpi}^{-1}\mathbf{D}_{\boldsymbol{\beta},\varpi}\mathbf{A}_{\boldsymbol{\beta},\varpi}^{-1}$</td><td colspan="2">$\mathbf{V}_{\boldsymbol{\beta}} = \mathbf{A}_{\boldsymbol{\beta},\varpi}^{-1}(\mathbf{D}_{\boldsymbol{\beta},\varpi} - \mathbf{M}_{\boldsymbol{\beta},\varpi,\boldsymbol{\gamma}})\mathbf{A}_{\boldsymbol{\beta},\varpi}^{-1}$
$\mathbf{M}_{\boldsymbol{\beta},\varpi,\boldsymbol{\gamma}} = \mathbf{C}_{\boldsymbol{\beta},\varpi}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{C}_{\boldsymbol{\beta},\varpi}^{\top}$ for ATE weighting; $\mathbf{M}_{\boldsymbol{\beta},\varpi,\boldsymbol{\gamma}} = \mathbf{C}_{\boldsymbol{\beta},\varpi,1}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{C}_{\boldsymbol{\beta},\varpi,2}^{\top} + \mathbf{C}_{\boldsymbol{\beta},\varpi,2}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{C}_{\boldsymbol{\beta},\varpi,1}^{\top} - \mathbf{C}_{\boldsymbol{\beta},\varpi,2}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{C}_{\boldsymbol{\beta},\varpi,2}^{\top}$ for ATT weighting</td><td rowspan="3">Same federation procedure with $\hat{\boldsymbol{\gamma}}^{\text{pad},(k)}$, estimated $\mathbf{A}_{\boldsymbol{\gamma}}^{\text{pad},(k)}$, $\mathbf{C}_{\boldsymbol{\beta},\varpi}^{\text{pad},(k)}$ (and $\mathbf{B}_{\boldsymbol{\gamma}}^{\text{pad},(k)}$ if needed) if propensity models are unstable and estimated, and with $\hat{\boldsymbol{\beta}}^{\text{pad},(k)}$, estimated $\mathbf{A}_{\boldsymbol{\beta},\varpi}^{\text{pad},(k)}$, $\mathbf{D}_{\boldsymbol{\beta},\varpi}^{\text{pad},(k)}$, $\mathbf{C}_{\boldsymbol{\beta},\varpi}^{\text{pad},(k)}$ if outcomes models are unstable</td></tr>
<tr><td>$\mathbf{V}_{\boldsymbol{\gamma}} = \mathbf{A}_{\boldsymbol{\gamma}}^{-1}$</td><td>$\mathbf{V}_{\boldsymbol{\gamma}} = \mathbf{A}_{\boldsymbol{\gamma}}^{-1}\mathbf{B}_{\boldsymbol{\gamma}}\mathbf{A}_{\boldsymbol{\gamma}}^{-1}$</td></tr>
<tr><td colspan="3">(1) Federate $\hat{\boldsymbol{\gamma}}^{(k)}$ by Hessian weighting (skip for known propensity); (2) Estimate $\boldsymbol{\beta}^{(k)}$ using $\hat{\boldsymbol{\gamma}}^{\text{fed}}$; (3) Estimate $\mathbf{A}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{C}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{D}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{A}_{\boldsymbol{\gamma}}^{(k)}$, and $\mathbf{B}_{\boldsymbol{\gamma}}^{(k)}$ using $\hat{\boldsymbol{\gamma}}^{\text{fed}}$ and $\hat{\boldsymbol{\beta}}^{\text{fed}}$; (4) Federate estimated $\mathbf{A}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{C}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{D}_{\boldsymbol{\beta},\varpi}^{(k)}, \mathbf{A}_{\boldsymbol{\gamma}}^{(k)}$, and $\mathbf{B}_{\boldsymbol{\gamma}}^{(k)}$ by sample size weighting.</td></tr>
<tr><td>Results</td><td colspan="3">Theorem 2</td><td>Proposition 4</td></tr>
</table>

Federate individual propensity/outcome models

- Key components
  - Multiple matrices involved: Hessian, outer product of gradient, …
  - Multiple weighting methods involved: Hessian weighting, sample size weighting, …
  - Unrestricted federated method with a flexible specification when propensity/outcome models are unstable

Description	Assume Stable Known Propensity and Stable Outcome Model (IPW-MLE #1)	Assume Stable Propensity and Stable Outcome Model (IPW-MLE #2)	Assume Stable Misspecified Propensity and Stable Outcome Model (IPW-MLE #3)	Assume Unstable Propensity or Unstable Outcome Model (IPW-MLE #4)
Stable Covariate Distribution	yes or no	yes or no	yes or no	yes or no
Known Propensity	yes	no	no	yes or no
Stable Propensity Model	yes	yes	yes	yes or no
Stable Outcome Model	yes	yes	yes	yes or no
Correct Propensity Model Specification	yes	yes	no	yes or no
Correct Outcome Model Specification	yes or no	yes or no	yes or no	yes or no
Sample Size Assumption	yes or no	yes or no	yes or no	yes
Coefficient β federation	(1) Estimate $\beta^{(k)}$ using γ_0; (2) Federate $\hat{\beta}^{(k)}$ by Hessian weighting.	(1) Federate $\hat{\gamma}^{(k)}$ by Hessian weighting; (2) Estimate $\beta^{(k)}$ using $\hat{\gamma}^{\text{fed}}$; (3) Federate $\hat{\beta}^{(k)}$ by Hessian weighting.		Same federation procedure with $\hat{\gamma}^{\text{pad},(k)}$ and $\hat{\mathbf{H}}_{\gamma}^{\text{pad},(k)}$ if propensity models are unstable and estimated, and with $\hat{\beta}^{\text{pad},(k)}$ and $\hat{\mathbf{H}}_{\beta}^{\text{pad},(k)}$ if outcomes models are unstable.
Variance $\mathbf{V}$ federation	$\mathbf{V}_\beta = \mathbf{A}_{\beta,\varpi}^{-1}\mathbf{D}_{\beta,\varpi}\mathbf{A}_{\beta,\varpi}^{-1}$	$\mathbf{V}_\beta = \mathbf{A}_{\beta,\varpi}^{-1}(\mathbf{D}_{\beta,\varpi} - \mathbf{M}_{\beta,\varpi,\gamma})\mathbf{A}_{\beta,\varpi}^{-1}$ $\mathbf{M}_{\beta,\varpi,\gamma} = \mathbf{C}_{\beta,\varpi}\mathbf{V}_\gamma\mathbf{C}_{\beta,\varpi}^\top$ for ATE weighting; $\mathbf{M}_{\beta,\varpi,\gamma} = \mathbf{C}_{\beta,\varpi,1}\mathbf{V}_\gamma\mathbf{C}_{\beta,\varpi,2}^\top + \mathbf{C}_{\beta,\varpi,2}\mathbf{V}_\gamma\mathbf{C}_{\beta,\varpi,1}^\top - \mathbf{C}_{\beta,\varpi,2}\mathbf{V}_\gamma\mathbf{C}_{\beta,\varpi,2}^\top$ for ATT weighting; $\mathbf{V}_\gamma = \mathbf{A}_\gamma^{-1}$	$\mathbf{V}_\gamma = \mathbf{A}_\gamma^{-1}\mathbf{B}_\gamma\mathbf{A}_\gamma^{-1}$	Same federation procedure with $\hat{\gamma}^{\text{pad},(k)}$, estimated $\mathbf{A}_{\gamma}^{\text{pad},(k)}$, $\mathbf{C}_{\beta,\varpi}^{\text{pad},(k)}$ (and $\mathbf{B}_{\gamma}^{\text{pad},(k)}$ if needed) if propensity models are unstable and estimated, and with $\hat{\beta}^{\text{pad},(k)}$, estimated $\mathbf{A}_{\beta,\varpi}^{\text{pad},(k)}$, $\mathbf{D}_{\beta,\varpi}^{\text{pad},(k)}$, $\mathbf{C}_{\beta,\varpi}^{\text{pad},(k)}$ if outcomes models are unstable
	(1) Federate $\hat{\gamma}^{(k)}$ by Hessian weighting (skip for known propensity); (2) Estimate $\beta^{(k)}$ using $\hat{\gamma}^{\text{fed}}$; (3) Estimate $\mathbf{A}_{\beta,\varpi}^{(k)}, \mathbf{C}_{\beta,\varpi}^{(k)}, \mathbf{D}_{\beta,\varpi}^{(k)}, \mathbf{A}_{\gamma}^{(k)}$, and $\mathbf{B}_{\gamma}^{(k)}$ using $\hat{\gamma}^{\text{fed}}$ and $\hat{\beta}^{\text{fed}}$; (4) Federate estimated $\mathbf{A}_{\beta,\varpi}^{(k)}, \mathbf{C}_{\beta,\varpi}^{(k)}, \mathbf{D}_{\beta,\varpi}^{(k)}, \mathbf{A}_{\gamma}^{(k)}$, and $\mathbf{B}_{\gamma}^{(k)}$ by sample size weighting.			
Results		Theorem 2		Proposition 4

Federation with treatment effect heterogeneity

- An interactive outcome model specification: e.g.,
  $\log \frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + W \cdot X^T\beta_{wx} + X^T\beta_x$

  - Heterogeneous treatment effect: The treatment effect on the log-odds ratio is $\beta_w + X^T\beta_{wx}$

  - Our federation procedure continues to work
    - If β_{wx} is stable across some datasets, federation increases the precision of $\hat{\beta}_{wx}$

Framework for federated causal inference

1. Input

Dataset A

Dataset B

2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Determine which conditions hold

- Stable 🚫
- Correct Spec ✅

4. Federate propensity models across datasets

$$\begin{pmatrix} \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}}} & \mathbf{0}_{|\mathcal{S}|\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}^\complement}} & \mathbf{0}_{|\mathcal{S}|\times S_{k+1}^{D}} \\ \mathbf{0}_{S_1^{k-1}\times|\mathcal{S}|} & \mathbf{0}_{S_1^{k-1}\times S_1^{k-1}} & \mathbf{0}_{S_1^{k-1}\times S_k^{k}} & \mathbf{0}_{S_1^{k-1}\times S_{k+1}^{D}} \\ \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}^\complement},\boldsymbol{\beta}_{\mathcal{S}}} & \mathbf{0}_{S_k^{k}\times S_1^{k-1}} & \hat{\mathbf{H}}^{(k)}_{\boldsymbol{\beta}_{\mathcal{S}^\complement},\boldsymbol{\beta}_{\mathcal{S}^\complement}} & \mathbf{0}_{S_k^{k}\times S_{k+1}^{D}} \\ \mathbf{0}_{S_{k+1}^{D}\times|\mathcal{S}|} & \mathbf{0}_{S_{k+1}^{D}\times S_1^{k-1}} & \mathbf{0}_{S_{k+1}^{D}\times S_k^{k}} & \mathbf{0}_{S_{k+1}^{D}\times S_{k+1}^{D}} \end{pmatrix}$$

5. Federate outcome models across datasets

6. Output

$\hat{\beta}_{fed}, \hat{V}_{fed}$

Asymptotic results

Asymptotic distribution of federated IPW-MLE

Theorem 1. Suppose the regularity conditions for the parametric propensity and outcome models hold. $\hat{\beta}_{fed}$ and $\hat{V}_{\beta,fed}$ are the federated coefficients and variance from our federated IPW-MLE. $\hat{\beta}_{pooled}$ and $\hat{V}_{\beta,pooled}$ are the estimated coefficients and variance from IPW-MLE on the combined individual-level data. As the sample size of each dataset grows to infinity, we have

$$n_{pooled}^{1/2} \cdot \hat{V}_{\beta,pooled}^{-1/2} \cdot \left(\hat{\beta}_{fed} - \beta_0\right) \xrightarrow{d} N(0,1) \qquad \text{Eq. (1)}$$

where n_{pooled} is the total sample size. If we replace $\hat{\beta}_{fed}$ by $\hat{\beta}_{pooled}$ and/or replace $\hat{V}_{\beta,pooled}$ by $\hat{V}_{\beta,fed}$, Eq. (1) continues to hold.

- Theorem 1 implies
  1. $\hat{\beta}_{fed}$ is doubly robust and asymptotically normal
  2. $\hat{\beta}_{fed}$ is as efficient as $\hat{\beta}_{pooled}$
  3. $\hat{V}_{\beta,fed}$ is consistent

➢ Our federated IPW-MLE provides valid confidence intervals for β_0

Asymptotic distribution of federated AIPW

Theorem 2. Suppose at least one of the propensity and outcome models are correctly specified. $\hat{\tau}_{fed}$ and $\hat{V}_{\tau,fed}$ are the federated coefficients and variance from our federated AIPW. $\hat{\tau}_{pooled}$ and $\hat{V}_{\tau,pooled}$ are the estimated treatment effect and its variance from AIPW on the combined individual-level data. As the sample size of each dataset grows to infinity, we have

$$n_{pooled}^{1/2} \cdot \hat{V}_{\tau,pooled}^{-1/2} \cdot \left(\hat{\tau}_{fed} - \tau_0\right) \xrightarrow{d} N(0,1) \qquad \text{Eq. (2)}$$

where n_{pooled} is the total sample size. If we replace $\hat{\tau}_{fed}$ by $\hat{\tau}_{pooled}$ and/or replace $\hat{V}_{\tau,pooled}$ by $\hat{V}_{\tau,fed}$, Eq. (2) continues to hold.

- Theorem 2 implies
  1. $\hat{\tau}_{fed}$ is doubly robust and asymptotically normal
  2. $\hat{\tau}_{fed}$ is as efficient as $\hat{\tau}_{pooled}$
  3. $\hat{V}_{\tau,fed}$ is consistent

➢Our federated AIPW provides valid confidence intervals for τ_0 and the treatment coefficient

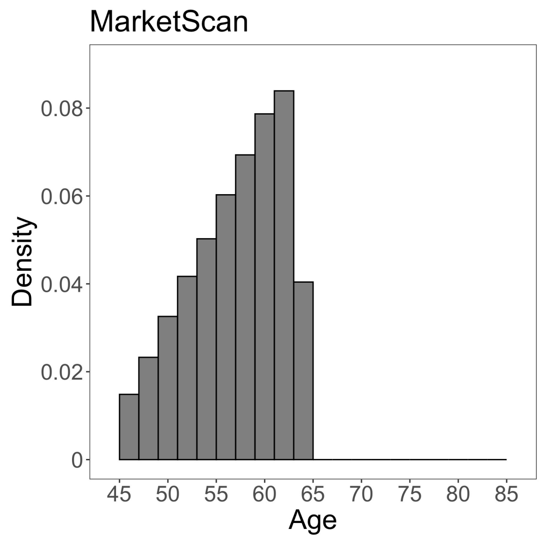
Empirical results

Empirical applications: Selecting a federation method

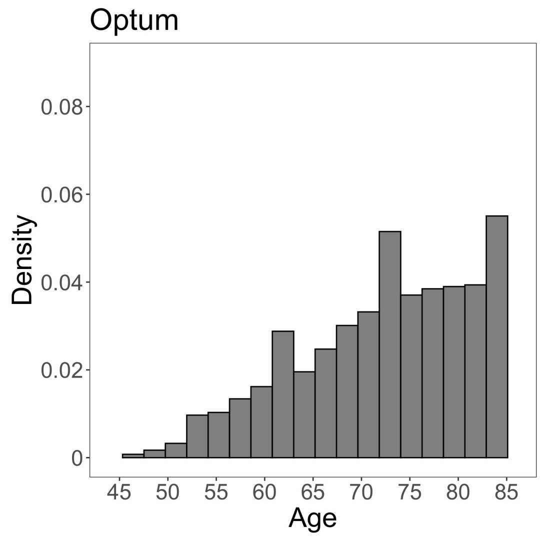
- Which method should we use to federate MarketScan and Optum?
  - We do not know the ground truth of the result on the combined data

- Select a method based on sampling from one dataset and and federation of subsamples

MarketScan

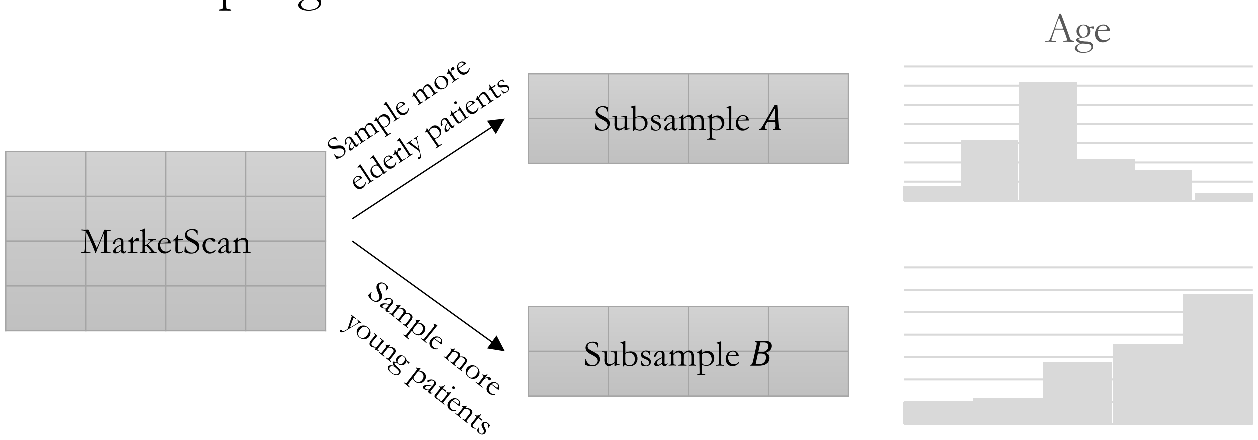


Optum



Procedure of sampling and federation

1. Construct subsamples by sampling from one dataset



2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Federate propensity and outcome models across datasets by various methods

$\hat{\beta}_{ivw}, \hat{V}_{ivw}$

$\hat{\beta}^{r}_{fed}, \hat{V}^{r}_{fed}$

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

⋮

4. Compare with that from the combined data

$\hat{\beta}_M, \hat{V}_M$

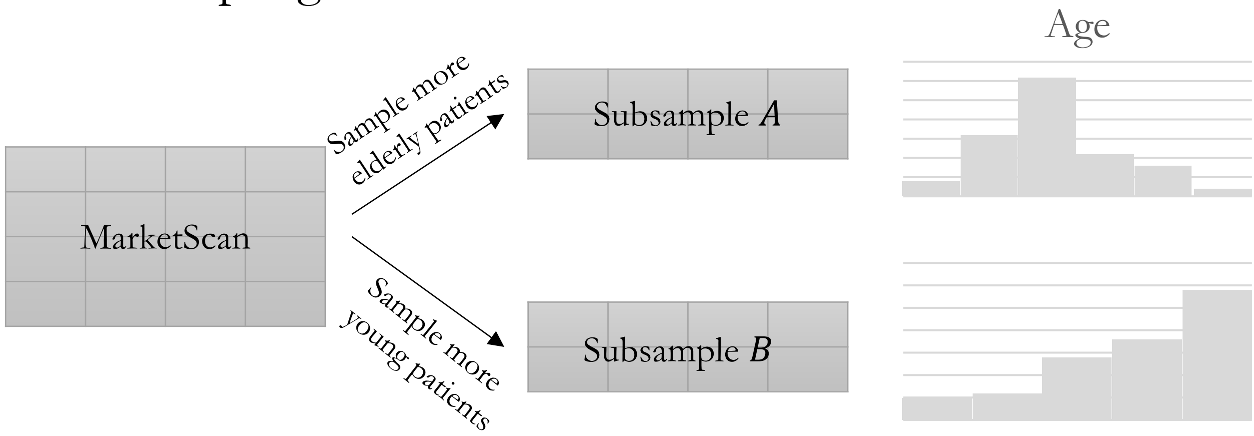
5. Output the method with the min MAE

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

6. Apply this method to federate two datasets

Procedure of sampling and federation

1. Construct subsamples by sampling from one dataset



2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Federate propensity and outcome models across datasets by various methods

$\hat{\beta}_{ivw}, \hat{V}_{ivw}$

$\hat{\beta}^{r}_{fed}, \hat{V}^{r}_{fed}$

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

⋮

4. Compare with that from the combined data

$\hat{\beta}_M, \hat{V}_M$

5. Output the method with the min MAE

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

6. Apply this method to federate two datasets

Results based sampling from one dataset

- *Restricted model*: Parameters in the propensity and outcome models are stable across subsamples
  - On the combined data: $\log\frac{P(W=1|X)}{P(W=0|X)} = X^T\gamma_x$ and $\log\frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + X^T\beta_x$
- *Unrestricted model*: Parameters of covariates in the propensity and outcome models are unstable
  - On the combined data: $\log\frac{P(W=1|X)}{P(W=0|X)} = X^T\gamma_x^{(k)}$ and $\log\frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + X^T\beta_x^{(k)}$, k indicates the subsample a patient belongs to

	$\hat{\beta}^{\mathrm{r}}_{w,\mathrm{bm}}$ mean	$\hat{\beta}_{w,\mathrm{ivw}}$ MAE	$\hat{\beta}^{\mathrm{r.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{\beta}^{\mathrm{unr.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*-0.7096*	0.9128	0.0526	0.0780
MS PNA	*-0.3019*	0.3883	0.0094	0.0115
Optum PNA	*-0.1832*	0.0536	0.0011	0.0043
	$\hat{V}^{\mathrm{r}}_{w,\mathrm{bm}}$ mean	$\hat{V}_{w,\mathrm{ivw}}$ MAE	$\hat{V}^{\mathrm{r.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{V}^{\mathrm{unr.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*0.0966*	0.0690	0.0282	0.0192
MS PNA	*0.0244*	0.0103	0.0024	0.0006
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

	$\hat{\beta}^{\mathrm{unr}}_{w,\mathrm{bm}}$ mean	$\hat{\beta}_{w,\mathrm{ivw}}$ MAE	$\hat{\beta}^{\mathrm{r.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{\beta}^{\mathrm{unr.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*-0.7495*	0.8728	0.1089	0.0302
MS PNA	*-0.3034*	0.3869	0.0144	0.0027
Optum PNA	*-0.1852*	0.0517	0.0043	0.0002
	$\hat{V}^{\mathrm{unr}}_{w,\mathrm{bm}}$ mean	$\hat{V}_{w,\mathrm{ivw}}$ MAE	$\hat{V}^{\mathrm{r.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{V}^{\mathrm{unr.fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*0.0835*	0.0559	0.0159	0.0054
MS PNA	*0.0242*	0.0102	0.0022	0.0002
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

Results based sampling from one dataset

- On the combined individual-level data
  - *Restricted benchmark* ($\hat{\beta}^{r}_{w,bm}$ *and* $\hat{V}^{r}_{w,bm}$): Estimates of β_x and its variance in $\log \frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + X^T\beta_x$
  - *Unrestricted benchmark* ($\hat{\beta}^{unr}_{w,bm}$ *and* $\hat{V}^{unr}_{w,bm}$): Estimates of β_x and its variance in $\log \frac{P(Y=1|X,W)}{P(Y=0|X,W)} = W\beta_w + X^T\beta_x^{(k)}$

	$\hat{\beta}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ MAE	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*-0.7096*	0.9128	0.0526	0.0780
MS PNA	*-0.3019*	0.3883	0.0094	0.0115
Optum PNA	*-0.1832*	0.0536	0.0011	0.0043
	$\hat{V}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ MAE	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*0.0966*	0.0690	0.0282	0.0192
MS PNA	*0.0244*	0.0103	0.0024	0.0006
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

	$\hat{\beta}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ MAE	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*-0.7495*	0.8728	0.1089	0.0302
MS PNA	*-0.3034*	0.3869	0.0144	0.0027
Optum PNA	*-0.1852*	0.0517	0.0043	0.0002
	$\hat{V}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ MAE	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ MAE
MS ARD	*0.0835*	0.0559	0.0159	0.0054
MS PNA	*0.0242*	0.0102	0.0022	0.0002
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

Results based sampling from one dataset (restricted benchmark)

Restricted benchmark

IVW

Our federated method for IPW-MLE assuming a restricted model on the combined data

Our federated method for IPW-MLE assuming an unrestricted model on the combined data

	$\hat{\beta}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ **MAE**	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*-0.7096*	0.9128	0.0526	0.0780
MS PNA	*-0.3019*	0.3883	0.0094	0.0115
Optum PNA	*-0.1832*	0.0536	0.0011	0.0043
	$\hat{V}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ **MAE**	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*0.0966*	0.0690	0.0282	0.0192
MS PNA	*0.0244*	0.0103	0.0024	0.0006
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

	$\hat{\beta}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ **MAE**	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*-0.7495*	0.8728	0.1089	0.0302
MS PNA	*-0.3034*	0.3869	0.0144	0.0027
Optum PNA	*-0.1852*	0.0517	0.0043	0.0002
	$\hat{V}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ **MAE**	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*0.0835*	0.0559	0.0159	0.0054
MS PNA	*0.0242*	0.0102	0.0022	0.0002
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

Results based sampling from one dataset (unrestricted benchmark)

Unrestricted benchmark

IVW

Our federated method for IPW-MLE assuming a restricted model on the combined data

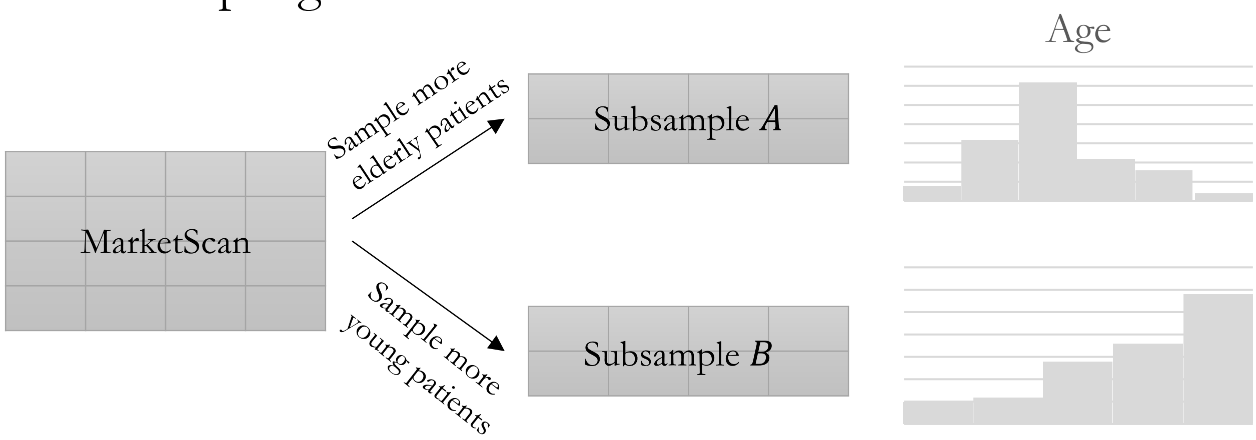
Our federated method for IPW-MLE assuming an unrestricted model on the combined data

	$\hat{\beta}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ **MAE**	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*-0.7096*	0.9128	0.0526	0.0780
MS PNA	*-0.3019*	0.3883	0.0094	0.0115
Optum PNA	*-0.1832*	0.0536	0.0011	0.0043
	$\hat{V}^{\mathbf{r}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ **MAE**	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*0.0966*	0.0690	0.0282	0.0192
MS PNA	*0.0244*	0.0103	0.0024	0.0006
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

	$\hat{\beta}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{\beta}_{w,\mathrm{ivw}}$ **MAE**	$\hat{\beta}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{\beta}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*-0.7495*	0.8728	0.1089	0.0302
MS PNA	*-0.3034*	0.3869	0.0144	0.0027
Optum PNA	*-0.1852*	0.0517	0.0043	0.0002
	$\hat{V}^{\mathbf{unr}}_{w,\mathrm{bm}}$ **mean**	$\hat{V}_{w,\mathrm{ivw}}$ **MAE**	$\hat{V}^{\mathbf{r}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**	$\hat{V}^{\mathbf{unr}.\mathrm{fed}}_{w,\mathrm{ipw\text{-}mle}}$ **MAE**
MS ARD	*0.0835*	0.0559	0.0159	0.0054
MS PNA	*0.0242*	0.0102	0.0022	0.0002
Optum PNA	*0.0031*	0.0003	0.0001	0.0000

Procedure of sampling and federation

1. Construct subsamples by sampling from one dataset



2. Estimate propensity and outcome models for each dataset

$\hat{\beta}_A, \hat{V}_A$

$\hat{\beta}_B, \hat{V}_B$

3. Federate propensity and outcome models across datasets by various methods

$\hat{\beta}_{ivw}, \hat{V}_{ivw}$

$\hat{\beta}^{r}_{fed}, \hat{V}^{r}_{fed}$

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

$\vdots$

4. Compare with that from the combined data

$\hat{\beta}_M, \hat{V}_M$

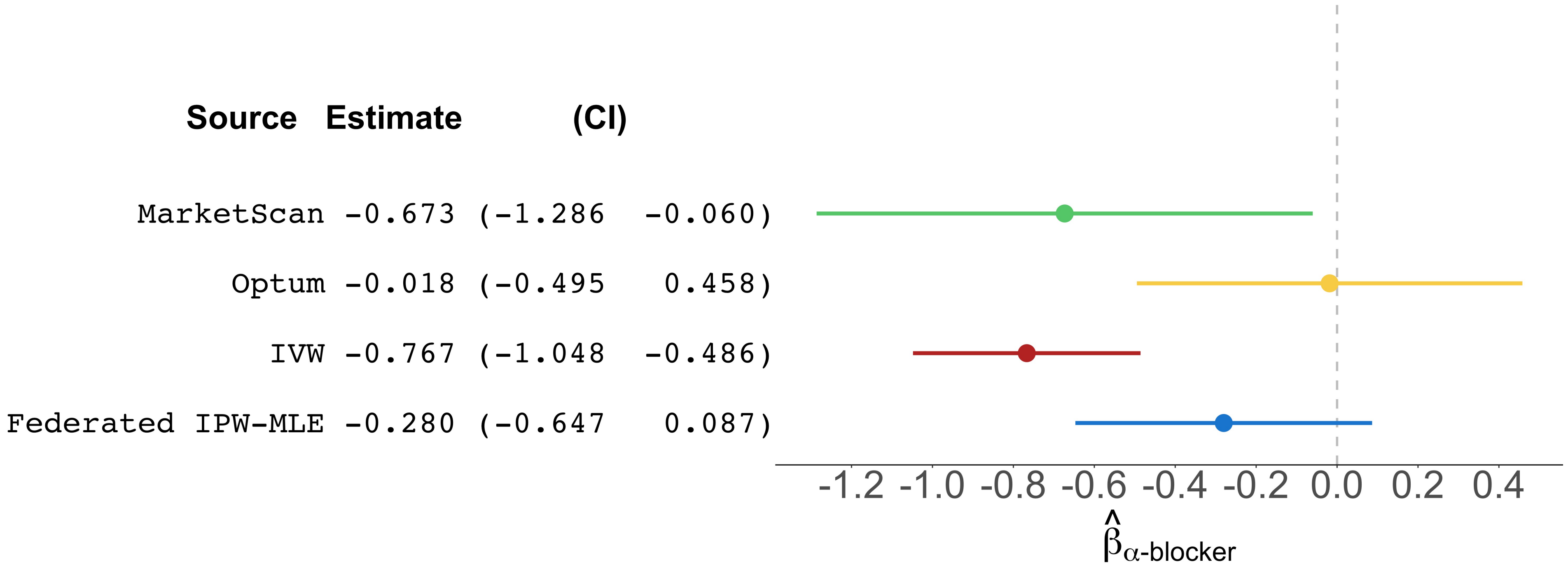
5. Output the method with the min MAE

$\hat{\beta}^{unr}_{fed}, \hat{V}^{unr}_{fed}$

6. Apply this method to federate two datasets

Applying the selected method to combine two datasets

- The federated estimate from the unrestricted federated method lies between the estimates on MarketScan and Optum
- Based on information from both datasets, we find that taking alpha-blockers reduces the log odds of the adverse outcome

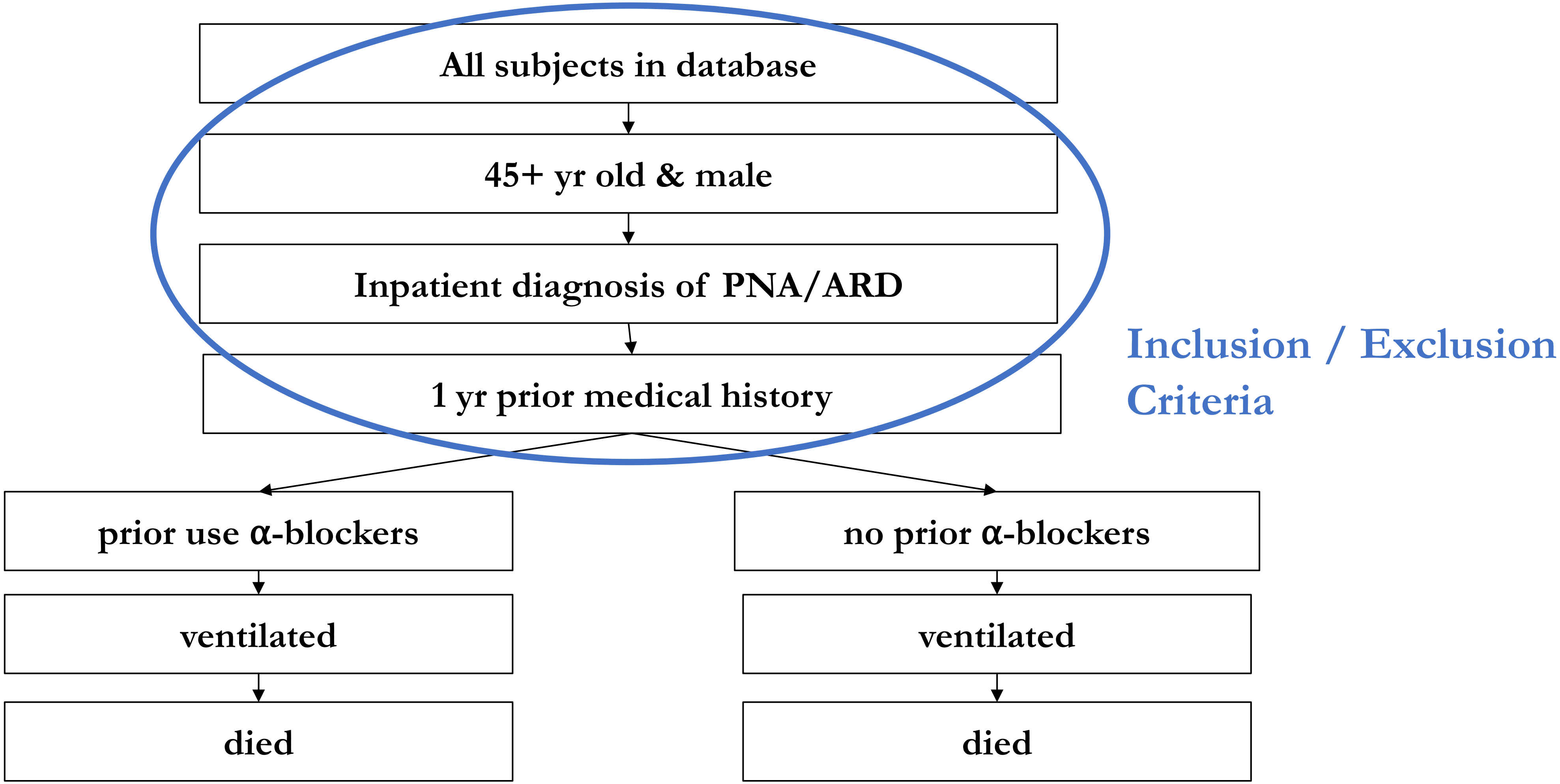


Conclusion

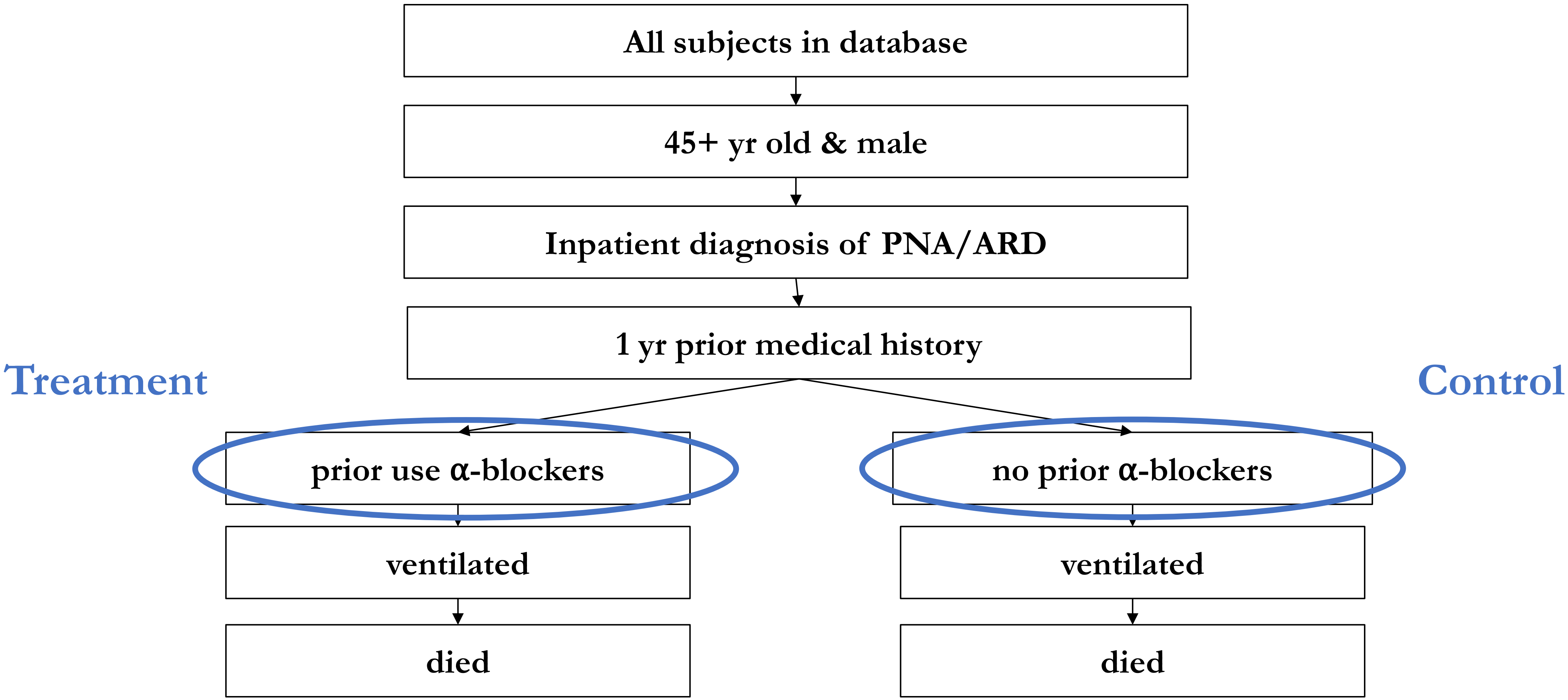
1. Federated methods that only use summary-level information from heterogeneous datasets
   - Depend on the stability and model specification conditions of propensity and outcome models
   - Two main categories: IPW-MLE and AIPW; one supplementary category: MLE
2. Asymptotic guarantees for federated point and variance estimators
   - Doubly robust, efficient, and asymptotic normal
3. A procedure to select federated methods on empirical datasets

Supplementary slides

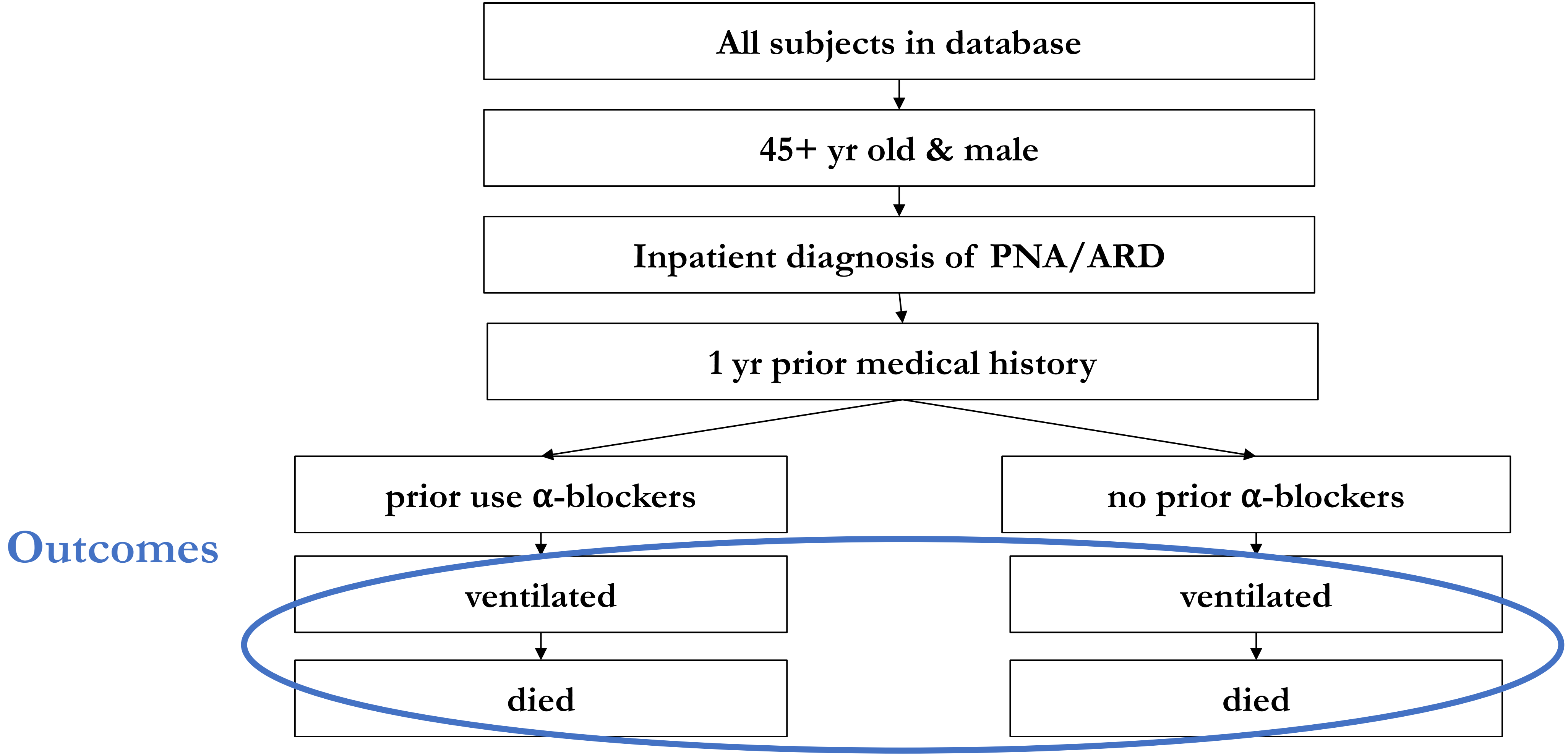
Ideal: an RCT


All subjects in database
45+ yr old & male
Inpatient diagnosis of PNA/ARD
1 yr prior medical history
Inclusion / Exclusion Criteria
prior use α-blockers
ventilated
died
no prior α-blockers
ventilated
died

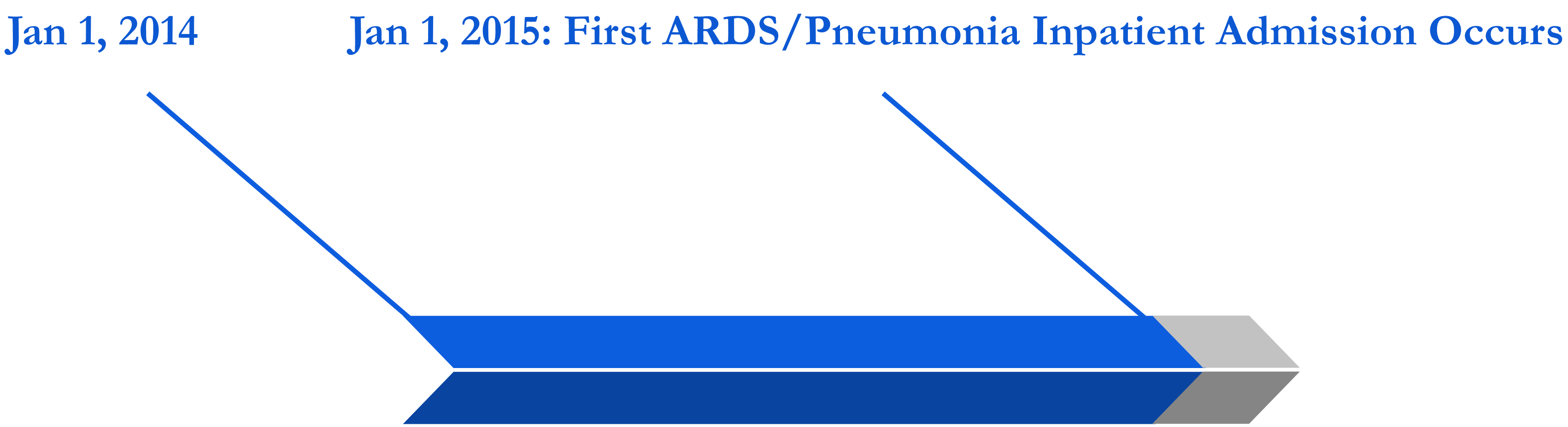
Ideal: an RCT


All subjects in database
45+ yr old & male
Inpatient diagnosis of PNA/ARD
1 yr prior medical history
Treatment
Control
prior use α-blockers
no prior α-blockers
ventilated
ventilated
died
died

Ideal: an RCT


All subjects in database
45+ yr old & male
Inpatient diagnosis of PNA/ARD
1 yr prior medical history
prior use α-blockers
no prior α-blockers
Outcomes
ventilated
ventilated
died
died

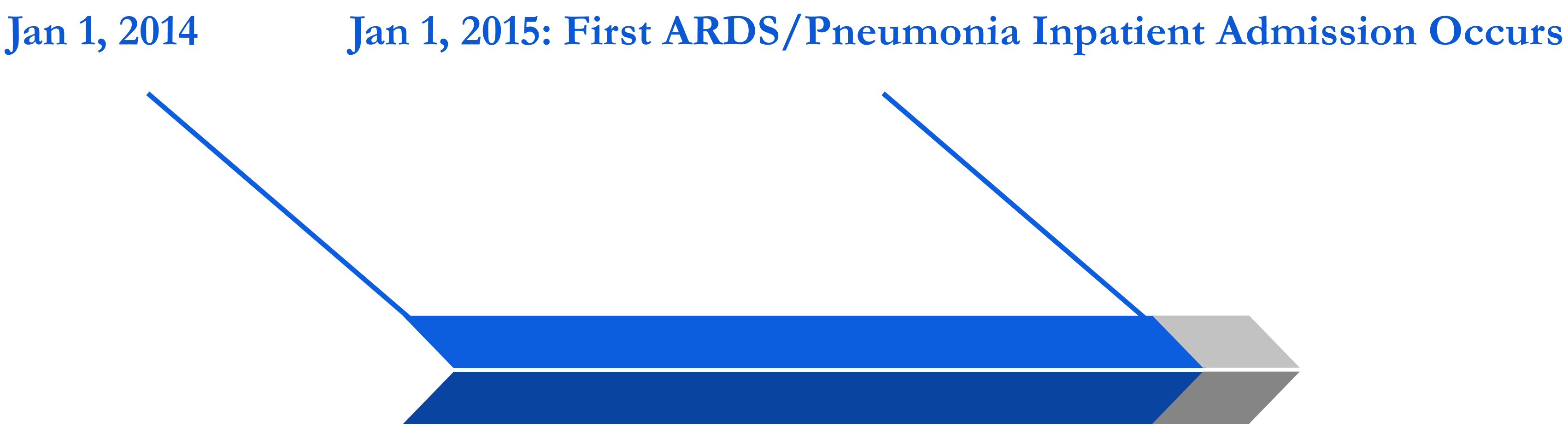
Retrospective analysis



In the past year, has the patient:

1. Taken a 180+ day supply of α-blockers?
2. Presented with comorbidities (e.g. heart failure, PTSD, etc.)?
3. Been admitted to the hospital as an inpatient?

Retrospective analysis



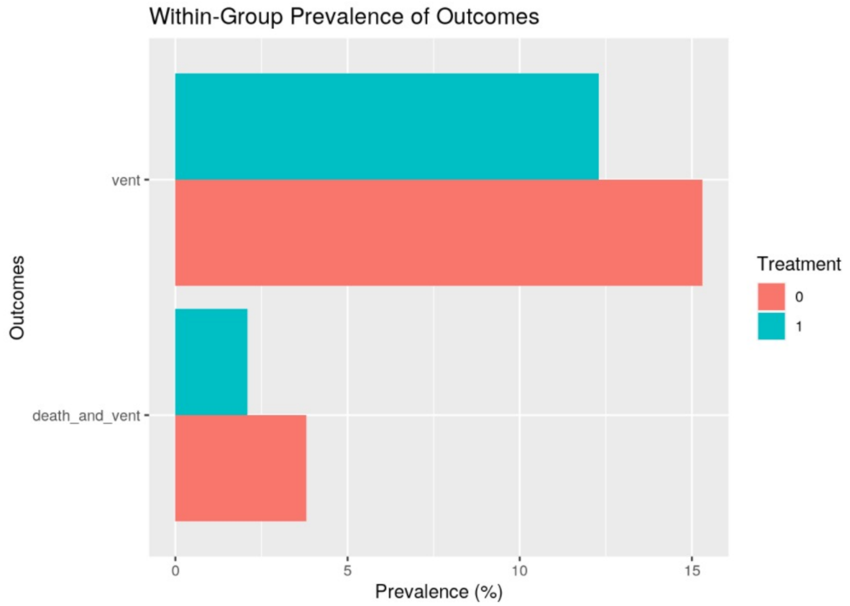
During inpatient admission:

1. Does the patient get ventilated?
2. Does the patient die?

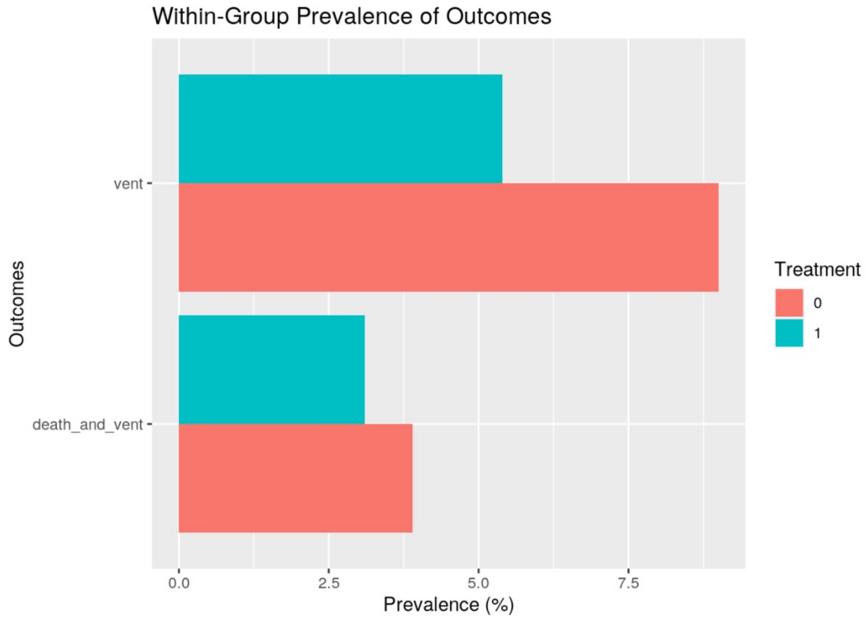
Within-group prevalence of adverse outcomes

- Within-group prevalence of adverse outcomes (ventilation, ventilation followed by death) is lower for the treated group on both datasets

MarketScan



Optum

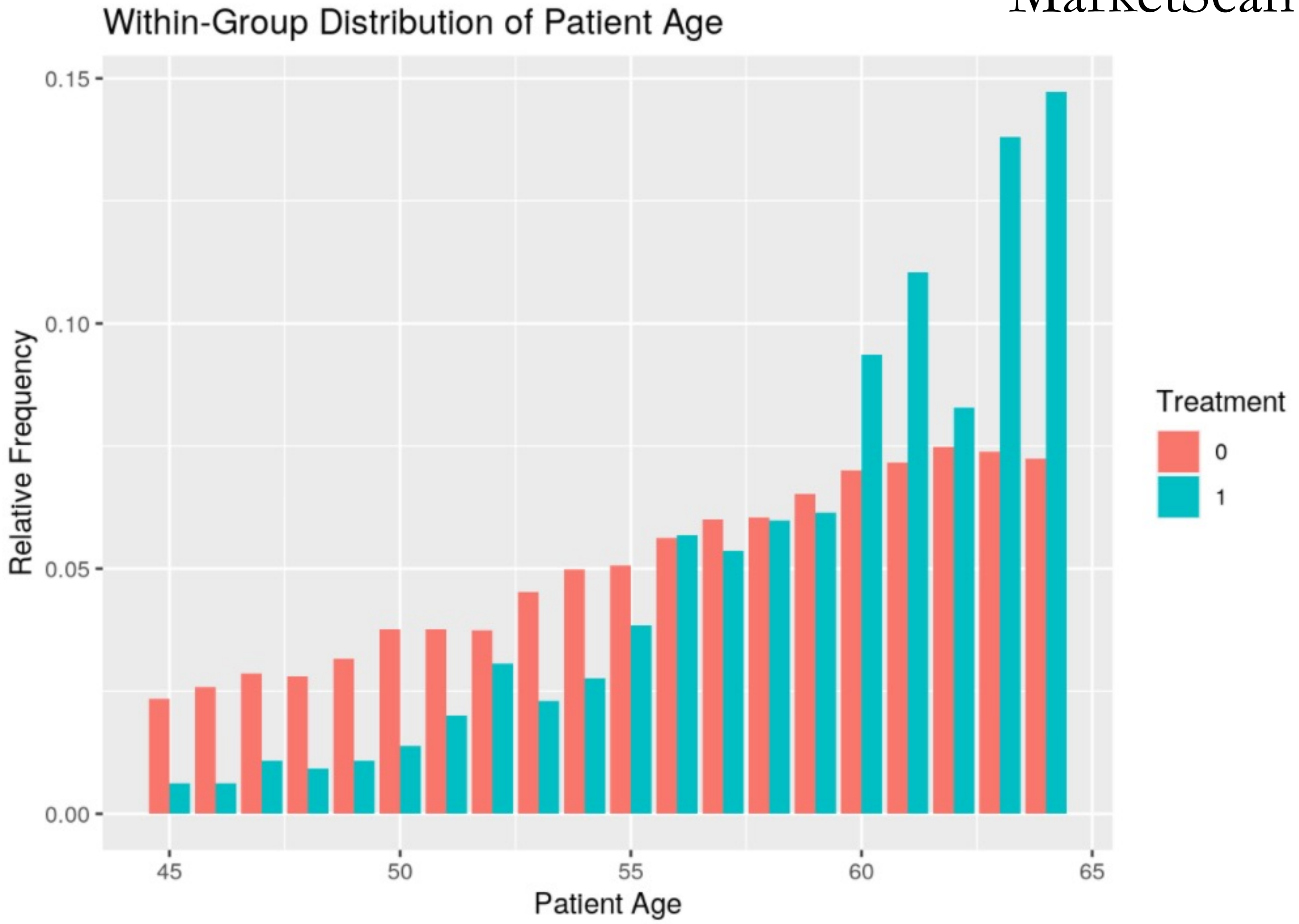


➢ But we need to account for confounding

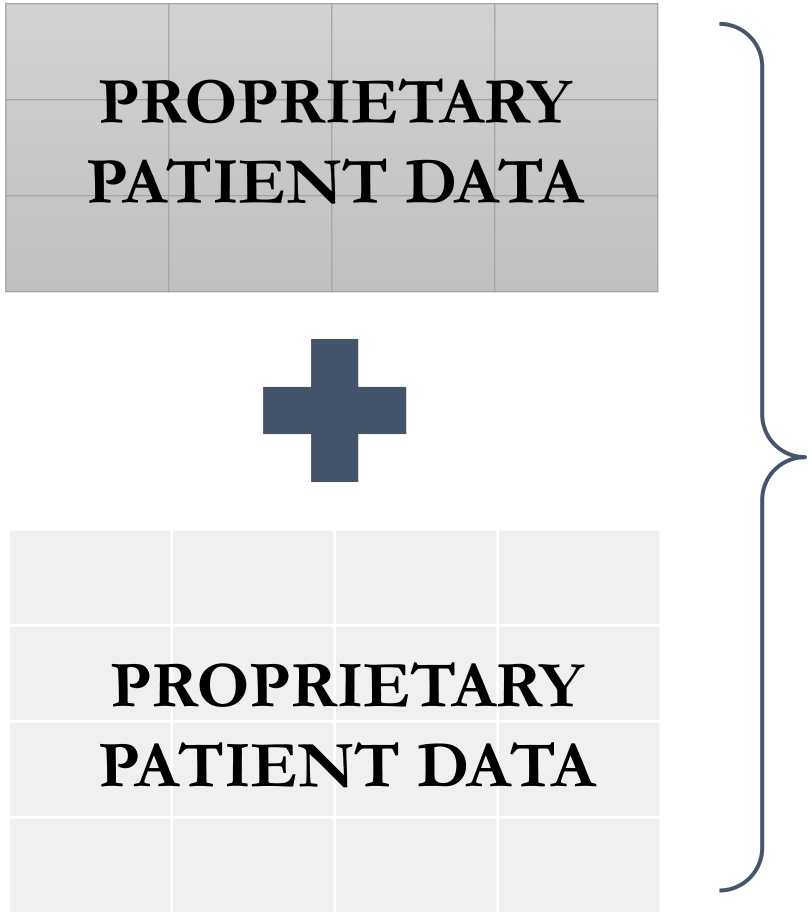
Imbalance between the treated and control groups

- Covariates are imbalanced
  - Prostate problems tend to worsen with age
  - Thus, treated patients are generally older

MarketScan



Solution: Combine heterogeneous patient data?



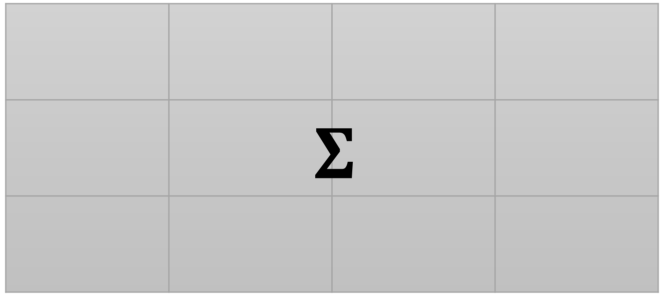
Ideal: Combine patient-level information on **MarketScan** and **Optum**

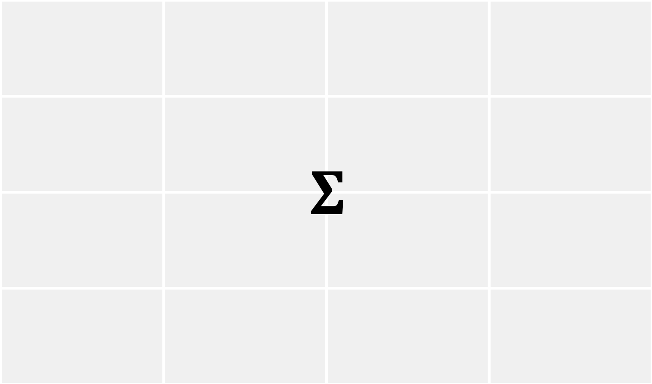
- More data for minority patient groups
- Increases statistical power for treatment effect estimation

Problem: Not allowed!

- **Legal** issues (data use agreements, data owner competition)
- **Ethical** issues (patient privacy)

Solution: Combine summary statistics





This paper proposes categories of federated methods

Method: On each dataset individually, calculate carefully constructed statistics related to both treatment assignment and patient outcomes

Objective: Obtain point and variance estimates that are asymptotically the same as those from the combined individual-level data

Asymptotic distribution of IPW-MLE continued

Lemma 1. Suppose the regularity conditions for the parametric propensity and outcome models hold. As the sample size $n \to \infty$, the IPW-MLE $\hat{\beta}$ is consistent and asymptotically normal

$$n^{1/2} \cdot (\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V_\beta)$$

where $V_\beta = A_{\beta,\varpi}^{-1} \cdot (D_{\beta,\varpi} - M_{\beta,\varpi,\gamma}) \cdot A_{\beta,\varpi}^{-1}$,
$M_{\beta,\varpi,\gamma} = C_{\beta,\varpi,\gamma,1} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T + C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,1}^T - C_{\beta,\varpi,\gamma,2} \cdot V_\gamma \cdot C_{\beta,\varpi,\gamma,2}^T$ and $V_\gamma = A_\gamma^{-1} \cdot B_\gamma \cdot A_\gamma^{-1}$

Matrix	Expression	Matrix	Expression
$\mathbf{A}_{\boldsymbol{\beta}}$	$\mathbb{E}\Big[-\frac{\partial^2 \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\Big]$	$\mathbf{A}_{\boldsymbol{\gamma}}$	$\mathbb{E}\Big[-\frac{\partial^2 \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\Big]$
$\mathbf{B}_{\boldsymbol{\beta}}$	$\mathbb{E}\Big[\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big(\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big)^\top\Big]$	$\mathbf{B}_{\boldsymbol{\gamma}}$	$\mathbb{E}\Big[\frac{\partial \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big(\frac{\partial \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big)^\top\Big]$
	ATE weighting $\varpi_{i,e_{\boldsymbol{\gamma}}} = \frac{w_i}{e_{\boldsymbol{\gamma}}(\mathbf{x}_i)} + \frac{1-w_i}{1-e_{\boldsymbol{\gamma}}(\mathbf{x}_i)}$		ATT weighting $\varpi_{i,e_{\boldsymbol{\gamma}}} = w_i + \frac{e_{\boldsymbol{\gamma}}(\mathbf{x}_i)}{1-e_{\boldsymbol{\gamma}}(\mathbf{x}_i)}(1-w_i)$
$\mathbf{A}_{\boldsymbol{\beta},\varpi}$	$\mathbb{E}\Big[\Big(\frac{w}{e_{\boldsymbol{\gamma}}} + \frac{1-w}{1-e_{\boldsymbol{\gamma}}}\Big)\frac{\partial^2 \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\Big]$	$\mathbf{A}_{\boldsymbol{\beta},\varpi}$	$\mathbb{E}\Big[\Big(w + \frac{e_{\boldsymbol{\gamma}}(1-w)}{1-e_{\boldsymbol{\gamma}}}\Big)\frac{\partial^2 \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\Big]$
$\mathbf{D}_{\boldsymbol{\beta},\varpi}$	$\mathbb{E}\Big[\Big(\frac{w}{e_{\boldsymbol{\gamma}}} + \frac{1-w}{1-e_{\boldsymbol{\gamma}}}\Big)^2\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \cdot \Big(\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big)^\top\Big]$	$\mathbf{D}_{\boldsymbol{\beta},\varpi}$	$\mathbb{E}\Big[\Big(w + \frac{e_{\boldsymbol{\gamma}}(1-w)}{1-e_{\boldsymbol{\gamma}}}\Big)^2\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \cdot \Big(\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big)^\top\Big]$
$\mathbf{C}_{\boldsymbol{\beta},\varpi}$	$\mathbb{E}\Big[\Big(\frac{w}{e_{\boldsymbol{\gamma}}^2} - \frac{1-w}{(1-e_{\boldsymbol{\gamma}})^2}\Big)\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \cdot \Big(\frac{\partial \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big)^\top\Big]$	$\mathbf{C}_{\boldsymbol{\beta},\varpi,1}$ $\mathbf{C}_{\boldsymbol{\beta},\varpi,2}$	$\mathbb{E}\Big[-\frac{(1-w)}{(1-e_{\boldsymbol{\gamma}})^2}\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \cdot \Big(\frac{\partial \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big)^\top\Big]$ $\mathbb{E}\Big[\Big(\frac{w}{e_{\boldsymbol{\gamma}}} - \frac{e_{\boldsymbol{\gamma}}(1-w)}{(1-e_{\boldsymbol{\gamma}})^2}\Big)\frac{\partial \log f(y\mid\mathbf{x},w,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \cdot \Big(\frac{\partial \log e(\mathbf{x},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big)^\top\Big]$

In the definitions of these matrices, $e_{\boldsymbol{\gamma}}$ denotes $e_{\boldsymbol{\gamma}}(\mathbf{x}_i) = e(\mathbf{x}_i, \boldsymbol{\gamma})$ by a slight abuse of notation.

Research question

- Question: Does taking alpha-blockers reduce the probability of the adverse outcome (ventilation (followed by death))?
- We seek to use both datasets (MarketScan and Optum) to answer this question