Causal Inference Methods

José R. Zubizarreta
Harvard University

09/04/2023
CUSO Doctoral School in Statistics and Applied Probability
Saignelégier, Switzerland

Overview

- Causal inference
  - Which treatments work?
  - Form whom?
  - When?
  - And why?
- Interventions
  - Point exposures
  - Time-varying
- Designs
  - Experimental
  - Observational
- Strategies
  - Randomization
  - Observation, assumptions
    - E.g., instruments
- Core methods
  - Matching
  - Regression
  - Weighting
- Sensitivity analyses
- Evidence integration
- Issues throughout
  - Missingness
  - Mismeasurement
  - Fairness...
- Perspectives
  - Statistics, biostatistics
  - Economics, political science
  - Computer science

References

- Chattopadhyay, A., and Z. (2023), "On the Implied Weights of Linear Regression for Causal Inference," *Biometrika*, 110, 615–629.
  - Chattopadhyay, A., and Z. (2023), "Notes on Causation, Comparison, and Regression," *arXiv:2305.14118v1*.
- Cohn, E. R., and Z. (2022), "Profile Matching for the Generalization and Personalization of Causal Inferences," *Epidemiology*, 33, 678–688.
- Wang, Y., and Z. (2020), "Minimal Dispersion Approximate Balancing Weights: Asymptotic Properties and Practical Considerations," *Biometrika*, 107, 93–105.

Outline

The experimental ideal

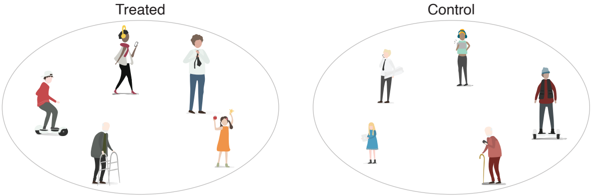
- No amount of being smart is a substitute for a randomized experiment
- But we can still learn from observational data

Cochran's advice

"The planner of an observational study should always ask himself the question, How would the study be conducted if it were possible to do it by controlled experimentation?"

Randomized experiments

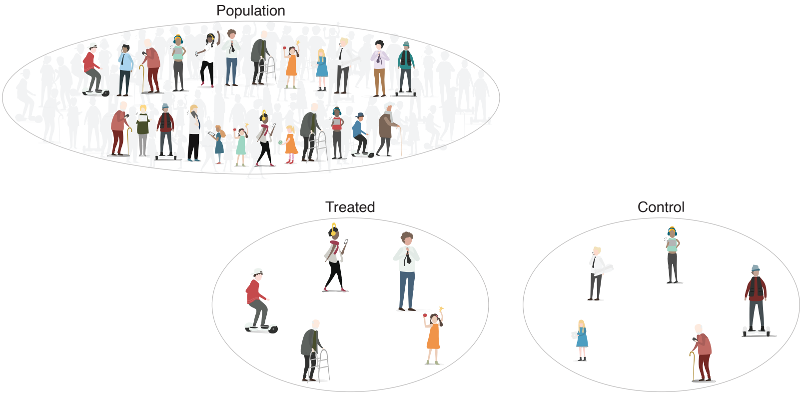
- In a randomized experiment, the treatment and control groups tend to be similar in terms of both observed and unobserved covariates



Designs by rawpixel.com / Freepik

Random sampling, random assignment

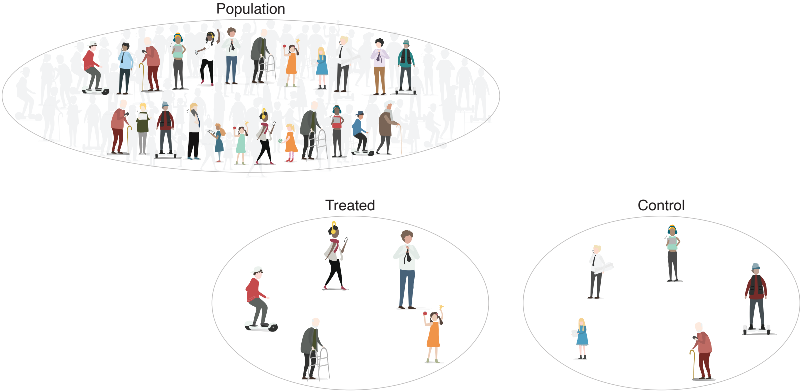
- Under random sampling, the treatment and control groups are representative of a target population



Designs by rawpixel.com / Freepik

Key features

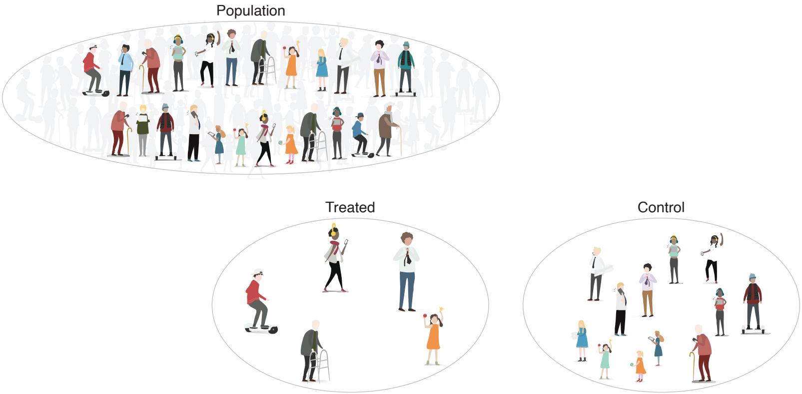
- Some key features of a randomized experiment are: covariate balance, study representativeness, self-weighted sampling, sample boundedness



Designs by rawpixel.com / Freepik

Observational studies

- In an observational study, treatment assignment is not at random, and groups tend to differ systematically in their covariates



Designs by rawpixel.com / Freepik

Motivating questions

- How are common methods for adjustment in observational studies approximating key features of a hypothetical experiment?

Motivating questions

- How are common methods for adjustment in observational studies approximating key features of a hypothetical experiment?
  - Three fundamental methods:

Motivating questions

- How are common methods for adjustment in observational studies approximating key features of a hypothetical experiment?
  - Three fundamental methods: matching, regression, weighting

Motivating questions

- How are common methods for adjustment in observational studies approximating key features of a hypothetical experiment?
  - Three fundamental methods: matching, regression, weighting
- Specifically, how are they acting on the individual (unit) level data at hand?
  - Closed-form expressions, mathematical optimization procedures

Motivating questions

- How are common methods for adjustment in observational studies approximating key features of a hypothetical experiment?
  - Three fundamental methods: matching, regression, weighting
- Specifically, how are they acting on the individual (unit) level data at hand?
  - Closed-form expressions, mathematical optimization procedures
- How are these methods different, and what are their weaknesses and strengths?
  - Study design, computational tractability, and statistical efficiency

Setup

- Estimand (for the most):
  - Average treatment effect (ATE)
    - $\text{ATE} := \mathbb{E}[Y_i(1) - Y_i(0)]$

Setup

- Estimand (for the most):
  - Average treatment effect (ATE)
    - ATE $:= \mathbb{E}[Y_i(1) - Y_i(0)]$
- Assumptions:
  - Strong ignorability
    - Positivity: $0 < P(Z_i = 1|\boldsymbol{X}_i = \boldsymbol{x}) < 1$ for all $\boldsymbol{x} \in \text{Supp}(\boldsymbol{X}_i)$
    - Unconfoundedness: $Y_i(1), Y_i(0) \perp\!\!\!\perp Z_i|\boldsymbol{X}_i$

Setup

- Estimand (for the most):
  - Average treatment effect (ATE)
    - $\text{ATE} := \mathbb{E}[Y_i(1) - Y_i(0)]$
- Assumptions:
  - Strong ignorability
    - Positivity: $0 < P(Z_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}) < 1$ for all $\boldsymbol{x} \in \text{Supp}(\boldsymbol{X}_i)$
    - Unconfoundedness: $Y_i(1), Y_i(0) \perp\!\!\!\perp Z_i | \boldsymbol{X}_i$
- Extensions:
  - Instrumental variables, difference-in-differences, discontinuity designs

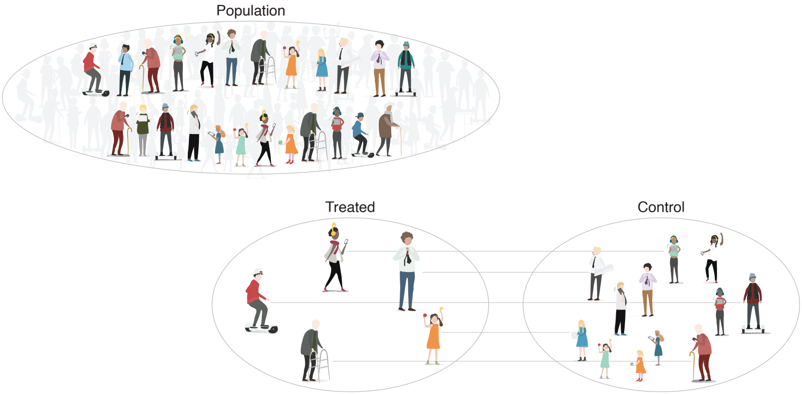
Outline

Matching methods

- [Rubin, 1973, *Biometrics*; Abadie and Imbens, 2006, *Econometrica*]
- [Rosenbaum, 1989; Hansen, 2004; *J. Am. Stat. Assoc.*]
- [Iacus et al., 2012; *Polit. Anal.*]
- [Diamond and Sekhon, 2013; *Rev. Econ. Stat.*]
- [Nikolaev et al., 2013; *Oper. Res.*]
- [Pimentel et al., 2015; *J. Am. Stat. Assoc.*]
- [Imai and Ratkovic, 2015; *J. R. Stat. Soc. B*]
- [King et al., 2016; *Am. J. Political Sci.*]
- [Parikh et al, 2022; *J. Mach. Learn. Res.*]
- Reviews: [Stuart, 2010, *Stat. Sci.*; Imbens, 2015, *J. Hum. Resour.*; Rosenbaum, 2020, *Annu. Rev. Stat. Appl.*]

Pair matching

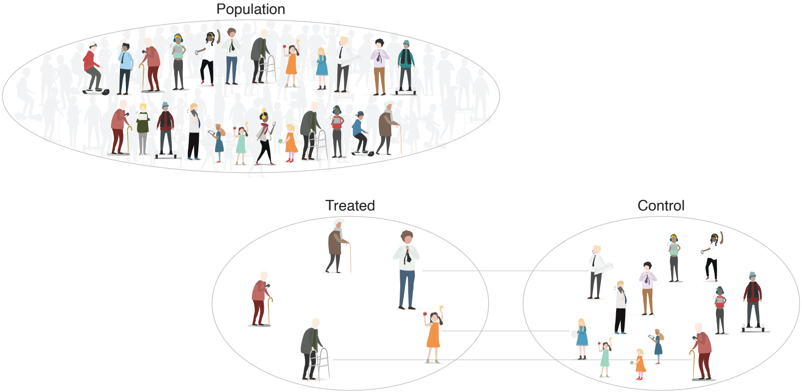
- With matching, we attempt to find the randomized experiment that is "hidden inside" the observational study



Designs by rawpixel.com / Freepik

Subset matching

- When there is limited overlap in covariate distributions we cannot match all the treated units



Designs by rawpixel.com / Freepik

An optimization framework [Z., 2012, *J. Am. Stat. Assoc.*; Z. et al., 2014, *Ann. Appl. Stat.*; Z. and Keele, 2017, *J. Am. Stat. Assoc.*; Wang and Z., 2022, *Stat. Sin.*]

$$\min_{\boldsymbol{m}}\{\mathbb{D}(\boldsymbol{m}) - \lambda\mathbb{I}(\boldsymbol{m}) : \boldsymbol{m} \in \mathcal{M} \cap \mathcal{B} \cap \mathcal{R}\}$$

where:

- $\mathbb{D}(\boldsymbol{m})$ is the total sum of covariate distances between the matched groups
- $\mathbb{I}(\boldsymbol{m})$ is the information content of the matched sample
- λ is a scalar chosen by the investigator
- $\mathcal{M}$, $\mathcal{B}$ and $\mathcal{R}$ are matching, balancing and representativeness constraints, respectively

Cardinality matching [Z. et al., 2014, *Ann. Appl. Stat.*; Kilcioglu and Z., 2016, *Ann. Appl. Stat.*; Visconti and Z., 2018, *Obs. Studies*; Niknam and Z., 2022, *JAMA*]

$$\min_{\boldsymbol{m}} \{ \quad - \quad \mathbb{I}(\boldsymbol{m}) : \boldsymbol{m} \in \mathcal{M} \cap \mathcal{B} \cap \mathcal{R} \}$$

where:

- $\mathbb{I}(\boldsymbol{m})$ is the information content of the matched sample
- $\mathcal{M}$, $\mathcal{B}$ and $\mathcal{R}$ are matching, balancing and representativeness constraints, respectively

Cardinality matching: fine balance [Rosenbaum et al., 2007, *J. Am. Stat. Assoc.*; Z., 2012, *J. Am. Stat. Assoc.*]

$$
\begin{aligned}
\underset{\boldsymbol{m}}{\text{maximize}} \quad & \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c} \\
\text{subject to} \quad & \sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \ \forall c \in \mathcal{C} \\
& \sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \ \forall t \in \mathcal{T} \\
& \sum_{t \in \mathcal{T}_{p,k}} \sum_{c \notin \mathcal{C}_{p,k}} m_{t,c} = \sum_{t \notin \mathcal{T}_{p,k}} \sum_{c \in \mathcal{C}_{p,k}} m_{t,c}, \ \forall p \in \mathcal{P}, k \in \mathcal{K}(p) \\
& m_{t,c} \in \{0, 1\}, \ t \in \mathcal{T}, c \in \mathcal{C}
\end{aligned}
$$

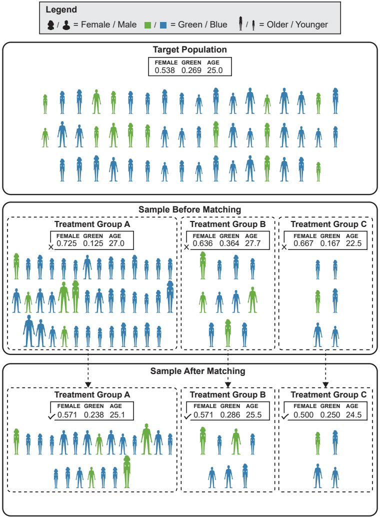
Handling "big data" with cardmatch [Bennett et al., 2020; *J. Comp. Graph. Stat.*]

Target size	Exposure size 70118	140236	210354	280472	350590	420708	490826	560944	631062	701180
1000	0.28	0.50	0.65	0.79	1.11	1.20	1.49	2.13	2.58	2.63
2000	0.20	0.72	0.91	1.14	1.49	1.56	1.87	2.20	2.53	2.67
3000	0.19	0.73	1.08	1.37	1.62	1.51	2.02	2.26	2.53	3.15
4000	0.22	0.44	1.09	1.57	1.74	1.98	2.00	2.29	2.48	2.62
5000	0.18	0.33	0.87	1.26	1.52	1.94	3.05	1.73	2.93	3.51
6000	0.26	0.47	0.64	1.66	2.07	2.40	2.78	2.94	3.18	3.04
7000	0.18	0.36	0.56	0.76	1.62	2.09	2.28	2.36	2.71	8.54
8000	0.25	0.40	0.57	0.82	1.87	2.25	2.42	2.95	3.08	3.66
9000	0.25	0.46	0.74	0.82	0.99	2.18	2.94	3.13	4.13	3.85
10000	0.19	0.39	0.63	0.83	1.08	2.55	2.58	2.93	3.13	3.42

Towards generalization and personalization

- Idea: balancing towards a target covariate profile [Chattopadhyay et al., 2021, *Stat. Med.*; Chattopadhyay and Z., 2022, *Biometrika*; Cohn and Z., 2022, *Epidemiology*]

Profile matching for a target population [Cohn and Z., 2022; *Epidemiology*]

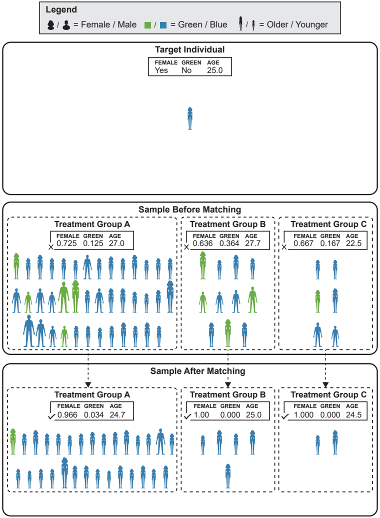


Illustrated by Xavier Alemañy

Profile matching with finite resolution [Cohn and Z., 2022; *Epidemiology*]

Illustrated by Xavier Alemañy

Profile matching for a target individual [Cohn and Z., 2022; *Epidemiology*]



Illustrated by Xavier Alemañy

A multidimensional knapsack problem [Cohn and Z., 2022; *Epidemiology*]

$$\begin{aligned}
\underset{\boldsymbol{m_t}, \boldsymbol{m_c}}{\text{maximize}} \quad & \sum_{t \in \mathcal{T}_T} m_t \\
\text{subject to} \quad & \left| \sum_{t \in \mathcal{T}_T} m_t B_k(\boldsymbol{X}_t) - m_t \boldsymbol{x}^* \right| \leq \sum_{t \in \mathcal{T}_T} m_t \delta_k, k = 1, ..., K \\
& m_t \in \{0, 1\}, \forall t \in \mathcal{T}_T
\end{aligned}$$

Outline

Related works

- [Abadie et al., 2015; *Am. J. Political Sci.*]
- [Angrist 1998; *Econometrica*]
- [Aronow and Samii, 2016; *Am. J. Political Sci.*]
- [Ben-Michael et al., 2021; *J. R. Stat. Soc. B*]
- [Fuller, 2009; *Sampling Statistics*]
- [Gelman and Imbens, 2018; *J. Bus. Econ. Stat.*]
- [Imbens, 2015; *J. Hum. Resour.*]
- [Kline, 2011; *Am. Econ. Rev.*]
- [Rao and Singh, 2009; *Pak. J. Stat.*]
- [Robins et al., 2007; *Stat. Sci.*]
- [Sloczynski., 2020; *Rev. Econ. Stat.*]

Stigler's automobile

"The method of least squares is the automobile of modern statistical analysis."

But when it comes to causal inference...

- Where is the experiment?

But when it comes to causal inference...

- Where is the experiment?

  ... or more specifically...

But when it comes to causal inference...

- Where is the experiment?
  ... or more specifically...
- How do linear regression adjustments in observational studies emulate key features of randomized experiments?

But when it comes to causal inference...

- Where is the experiment?

  ... or more specifically...
- How do linear regression adjustments in observational studies emulate key features of randomized experiments?
- In particular, how is linear regression acting on the individual-level data to produce to an average treatment effect estimate?

Contributions [Chattopadhyay and Z., 2022, *Biometrika*; 2021, *arXiv*]

- Closed form, finite sample expressions of the implied weights for a range of regression-based estimators:

Contributions [Chattopadhyay and Z., 2022, *Biometrika*; 2021, *arXiv*]

- Closed form, finite sample expressions of the implied weights for a range of regression-based estimators:
  - Traditional regression adjustments
  - g-computation
  - Augmented inverse probability weighting
  - Regression adjustments with multi-valued treatments
  - Regression adjustments after matching
  - Two-stage least squares with instrumental variables
  - Fixed effects

Contributions [Chattopadhyay and Z., 2022, *Biometrika*; 2021, *arXiv*]

- Closed form, finite sample expressions of the implied weights for a range of regression-based estimators:
  - Traditional regression adjustments
  - g-computation
  - Augmented inverse probability weighting
  - Regression adjustments with multi-valued treatments
  - Regression adjustments after matching
  - Two-stage least squares with instrumental variables
  - Fixed effects
- Analysis of the weights in both finite and large sample regimes

Contributions [Chattopadhyay and Z., 2022, *Biometrika*; 2021, *arXiv*]

- Closed form, finite sample expressions of the implied weights for a range of regression-based estimators:
  - Traditional regression adjustments
  - g-computation
  - Augmented inverse probability weighting
  - Regression adjustments with multi-valued treatments
  - Regression adjustments after matching
  - Two-stage least squares with instrumental variables
  - Fixed effects
- Analysis of the weights in both finite and large sample regimes
- Diagnostics for linear regression in causal inference

Implied weights of linear regression [Chattopadhyay and Z., 2022, *Biometrika*]

- Standard approach to regression adjustment:

$$\underbrace{Y_i^{\text{obs}}}_{\substack{\text{observed}\\ \text{outcome}}} = \beta_0 + \boldsymbol{\beta}_1^\top \underbrace{\boldsymbol{X}_i}_{\substack{\text{observed}\\ \text{covariates}}} + \tau \underbrace{Z_i}_{\substack{\text{treatment}\\ \text{indicator}\\ \in\{0,1\}}} + \epsilon_i$$

Implied weights of linear regression [Chattopadhyay and Z., 2022, *Biometrika*]

- Standard approach to regression adjustment:

$$\underbrace{Y_i^{\text{obs}}}_{\substack{\text{observed}\\\text{outcome}}} = \beta_0 + \boldsymbol{\beta}_1^\top \underbrace{\boldsymbol{X}_i}_{\substack{\text{observed}\\\text{covariates}}} + \tau \underbrace{Z_i}_{\substack{\text{treatment}\\\text{indicator}\\\in\{0,1\}}} + \epsilon_i$$

- $\hat{\tau}^{\text{OLS}}$ is equivalent to uni-regression imputation (URI):

$$\hat{\tau}^{\text{OLS}} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{Y}_i(1) - \hat{Y}_i(0)\}$$

Implied weights of linear regression [Chattopadhyay and Z., 2022, *Biometrika*]

- Standard approach to regression adjustment:

$$\underbrace{Y_i^{\text{obs}}}_{\substack{\text{observed}\\\text{outcome}}} = \beta_0 + \boldsymbol{\beta}_1^\top \underbrace{\boldsymbol{X}_i}_{\substack{\text{observed}\\\text{covariates}}} + \tau \underbrace{Z_i}_{\substack{\text{treatment}\\\text{indicator}\\\in\{0,1\}}} + \epsilon_i$$

- $\hat{\tau}^{\text{OLS}}$ is equivalent to uni-regression imputation (URI):

$$\hat{\tau}^{\text{OLS}} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{Y}_i(1) - \hat{Y}_i(0)\}$$

- In turn, this can be written as a Hájek estimator:

$$\hat{\tau}^{\text{OLS}} = \sum_{i:Z_i=1} w_i^{\text{URI}} Y_i^{\text{obs}} - \sum_{i:Z_i=0} w_i^{\text{URI}} Y_i^{\text{obs}}$$

More formally

Proposition 1. *The URI estimator of the ATE can be expressed as* $\hat{\tau}^{OLS} = \sum_{i:Z_i=1} w_i^{URI} Y_i^{obs} - \sum_{i:Z_i=0} w_i^{URI} Y_i^{obs}$ *where* $\boldsymbol{w_i^{URI} = n_t^{-1} + n n_c^{-1}(X_i - \bar{X}_t)^\top (S_t + S_c)^{-1}(\bar{X} - \bar{X}_t)}$ *for each unit in the treatment group and* $w_i^{URI} = n_c^{-1} + n n_t^{-1}(X_i - \bar{X}_c)^\top (S_t + S_c)^{-1}(\bar{X} - \bar{X}_c)$ *for each unit in the control group. Moreover, within each group the weights add up to one,* $\sum_{i:Z_i=0} w_i^{URI} = 1$ *and* $\sum_{i:Z_i=1} w_i^{URI} = 1$.

Properties of the URI weights

1. Exact balance:

$$\sum_{i:Z_i=1} w_i^{\text{URI}} \boldsymbol{X}_i = \sum_{i:Z_i=0} w_i^{\text{URI}} \boldsymbol{X}_i = \boldsymbol{X}^{*\text{URI}}$$

Properties of the URI weights

1. Exact balance:

$$\sum_{i:Z_i=1} w_i^{\text{URI}} \boldsymbol{X}_i = \sum_{i:Z_i=0} w_i^{\text{URI}} \boldsymbol{X}_i = \boldsymbol{X}^{*\text{URI}}$$

2. Target profile:

$$\boldsymbol{X}^{*\text{URI}} = \boldsymbol{S}_c(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_t + \boldsymbol{S}_t(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_c$$

Properties of the URI weights

1. Exact balance:

$$\sum_{i:Z_i=1} w_i^{\text{URI}} \boldsymbol{X}_i = \sum_{i:Z_i=0} w_i^{\text{URI}} \boldsymbol{X}_i = \boldsymbol{X}^{*\text{URI}}$$

2. Target profile:

$$\boldsymbol{X}^{*\text{URI}} = \boldsymbol{S}_c(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_t + \boldsymbol{S}_t(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_c$$

3. Minimum variance: The variance of weights in the treatment group is

$$\frac{1}{n_t}(\bar{\boldsymbol{X}}_t - \bar{\boldsymbol{X}}_c)^\top(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\boldsymbol{S}_t(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}(\bar{\boldsymbol{X}}_t - \bar{\boldsymbol{X}}_c)$$

Properties of the URI weights

1. Exact balance:

$$\sum_{i:Z_i=1} w_i^{\text{URI}} \boldsymbol{X}_i = \sum_{i:Z_i=0} w_i^{\text{URI}} \boldsymbol{X}_i = \boldsymbol{X}^{*\text{URI}}$$

2. Target profile:

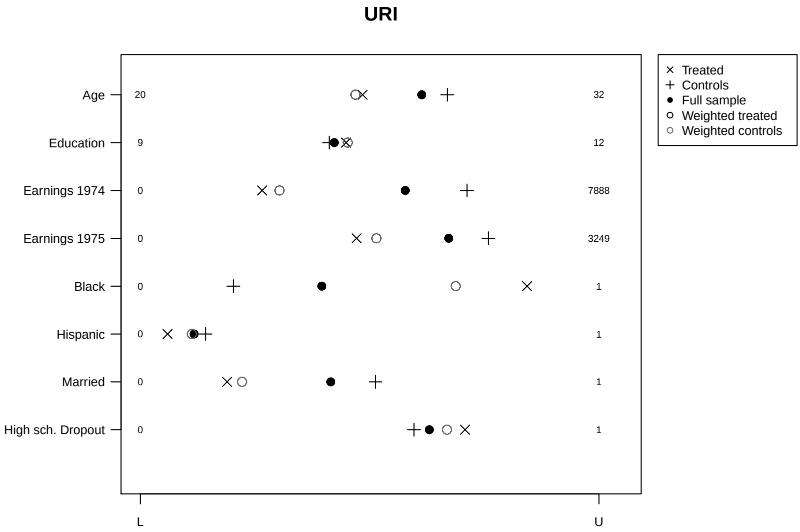
$$\boldsymbol{X}^{*\text{URI}} = \boldsymbol{S}_c(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_t + \boldsymbol{S}_t(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\bar{\boldsymbol{X}}_c$$

3. Minimum variance: The variance of weights in the treatment group is

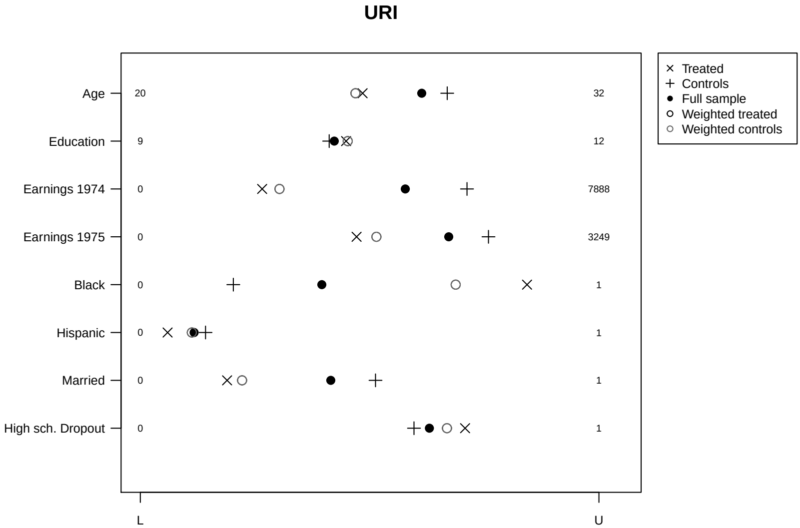
$$\frac{1}{n_t}(\bar{\boldsymbol{X}}_t - \bar{\boldsymbol{X}}_c)^\top(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}\boldsymbol{S}_t(\boldsymbol{S}_t + \boldsymbol{S}_c)^{-1}(\bar{\boldsymbol{X}}_t - \bar{\boldsymbol{X}}_c)$$

4. Model extrapolation: The weights can take negative values and produce estimators that are not sample bounded

Return to the Lalonde (1986) example



Implied target



Multi-regression imputation (MRI)

- Fit two linear models:
  - Treatment group, $Y_i^{\text{obs}} = \beta_{0t} + \boldsymbol{\beta}_{1t}^\top \boldsymbol{X}_i + \epsilon_{it}$
  - Control group, $Y_i^{\text{obs}} = \beta_{0c} + \boldsymbol{\beta}_{1c}^\top \boldsymbol{X}_i + \epsilon_{ic}$

$$\widehat{\text{ATE}} = \hat{\mathbb{E}}[Y_i(1)] - \hat{\mathbb{E}}[Y_i(0)] = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{0t} + \hat{\boldsymbol{\beta}}_{1t}^\top \boldsymbol{X}_i) - \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{0c} + \hat{\boldsymbol{\beta}}_{1c}^\top \boldsymbol{X}_i)$$

Properties of the MRI weights

1. Exact balance:

$$\sum_{i:Z_i=1} w_i^{\text{MRI}} \boldsymbol{X}_i = \sum_{i:Z_i=0} w_i^{\text{MRI}} \boldsymbol{X}_i = \boldsymbol{X}^{*\text{MRI}}$$

2. Target profile:

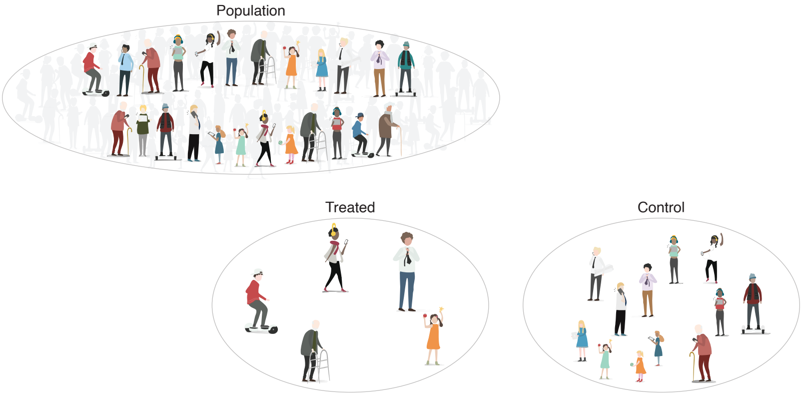
$$\boldsymbol{X}^{*\text{MRI}} = \bar{\boldsymbol{X}}$$

3. Minimum variance: The variance of weights in the treatment group is

$$\frac{1}{n_t}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{X}}_t)^\top \boldsymbol{S}_t^{-1}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{X}}_t)$$

4. Model extrapolation: The weights can take negative values and produce estimators that are not sample bounded

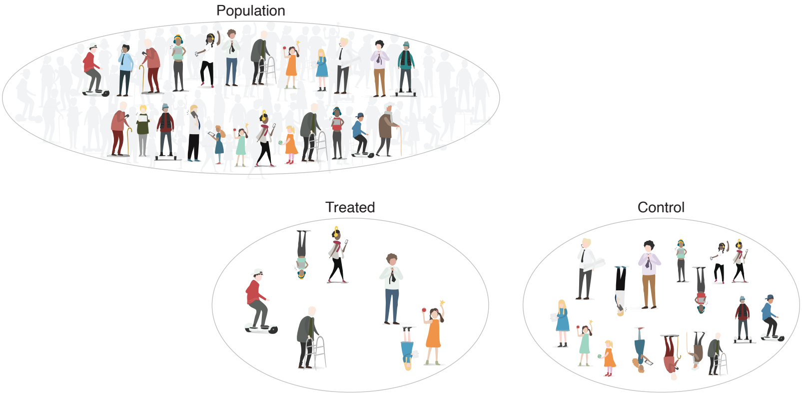
Observational studies



Designs by rawpixel.com / Freepik

Uni-regression imputation (URI)

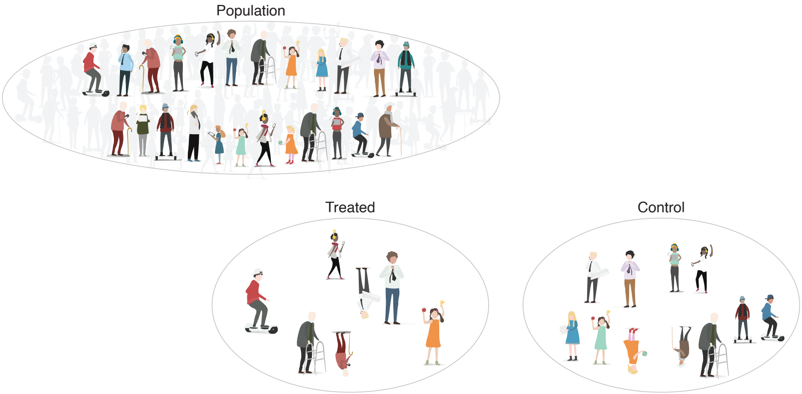
- URI adjustments: exact mean balance; hidden population; weights of minimum variance; negative weights



Designs by rawpixel.com / Freepik

Multi-regression imputation (MRI)

- MRI adjustments: exact mean balance; overall population; weights of minimum variance; negative weights



Designs by rawpixel.com / Freepik

Linear regression as a quadratic programming problem

[Chattopadhyay and Z., 2022, *Biometrika*]

THEOREM 3. *Consider the following quadratic programming problem in the control group*

$$\underset{w}{\textit{minimize}} \sum_{i:Z_i=0} \frac{(w_i - \tilde{w}_i^{base})^2}{w_i^{scale}} \textit{ subject to } \Big| \sum_{i:Z_i=0} w_i X_i - X^* \Big| \leq \delta, \; \sum_{i:Z_i=0} w_i = 1$$

where $\tilde{w}_i^{base}$ *are normalized base weights in the control group,* w_i^{scale} *are scaling weights, and* $X^* \in \mathbb{R}^k$ *is a covariate profile, all of them determined by the investigator. Then, for* $\delta = 0$ *the solution to this problem is*

$$w_i = \tilde{w}_i^{base} + w_i^{scale}(X_i - \bar{X}_c^{scale})^\top (S_c^{scale}/n_c)^{-1}(X^* - \bar{X}_c^{base}),$$

where $\bar{X}_c^{scale} = (\sum_{i:Z_i=0} w_i^{scale} X_i)/(\sum_{i:Z_i=0} w_i^{scale})$, $\bar{X}_c^{base} = \sum_{i:Z_i=0} \tilde{w}_i^{base} X_i$, *and* $S_c^{scale} = n_c \sum_{i:Z_i=0} w_i^{scale}(X_i - \bar{X}_c^{scale})(X_i - \bar{X}_c^{scale})^\top$. *Further, as special cases the implied weights of the weighted-URI, weighted-MRI, and AIPW estimators for the ATE are*

(a) ***weighted-URI***: $\tilde{w}_i^{base} = w_i^{base}/(\sum_{j:Z_j=0} w_j^{base})$, $w_i^{scale} = w_i^{base}$, $X^* = n_c^{-1} S_c^{scale}(n_t^{-1} S_t^{scale} + n_c^{-1} S_c^{scale})^{-1} \bar{X}_t^{scale} + n_t^{-1} S_t^{scale} \left(n_t^{-1} S_t^{scale} + n_c^{-1} S_c^{scale}\right)^{-1} \bar{X}_c^{scale}$.
(b) ***weighted-MRI***: $\tilde{w}_i^{base} = w_i^{scale} = w_i^{base}$, $X^* = \bar{X}$.
(c) ***AIPW***: $\tilde{w}_i^{base} = w_i^{base} = \{1 - \hat{e}(X_i)\}^{-1} / \sum_{j:Z_j=0}\{1 - \hat{e}(X_j)\}^{-1}$, $w_i^{scale} = 1$, $X^* = \bar{X}$.

Here, $\bar{X}_t^{scale} = (\sum_{i:Z_i=1} w_i^{scale} X_i)/(\sum_{i:Z_i=1} w_i^{scale})$ *and* $S_t^{scale} = n_t \sum_{i:Z_i=1} w_i^{scale}(X_i - \bar{X}_t^{scale})(X_i - \bar{X}_t^{scale})^\top$. *The weights for the treated units are obtained analogously.*

Multiple robustness of simple regression estimators

[Chattopadhyay and Z., 2022, *Biometrika*]

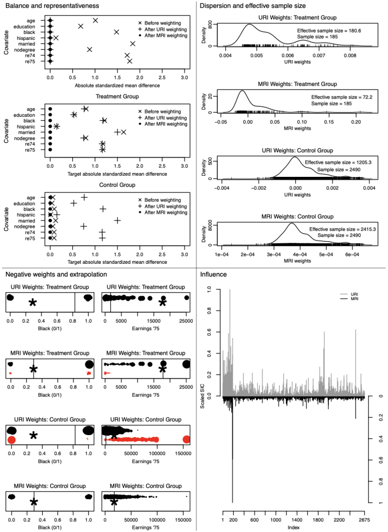
THEOREM 2.

(a) The URI estimator for the ATE is consistent if any of the following conditions holds: (i) $m_0(x)$ *is linear,* $e(x)$ *is inverse linear, and* $p^2 var(X_i \mid Z_i = 1) = (1-p)^2 var(X_i \mid Z_i = 0)$*; (ii)* $m_1(x)$ *is linear,* $1 - e(x)$ *is inverse linear, and* $p^2 var(X_i \mid Z_i = 1) = (1-p)^2 var(X_i \mid Z_i = 0)$*; (iii)* $m_1(x)$ *and* $m_0(x)$ *are linear and* $p^2 var(X_i \mid Z_i = 1) = (1-p)^2 var(X_i \mid Z_i = 0)$*; (iv)* $e(x)$ *is a constant function of* x*; (v)* $m_0(x)$ *and* $m_1(x)$ *are linear and* $m_1(x) - m_0(x)$ *is a constant function; (vi)* $m_1(x) - m_0(x)$ *is a constant function and* $e(x)$ *is linear in* x*.*

(b) The MRI estimator is consistent for the ATE if any of the following conditions holds: (i) $m_0(x)$ *is linear and* $e(x)$ *is inverse linear; (ii)* $m_1(x)$ *is linear and* $1 - e(x)$ *is inverse linear; (iii)* $m_1(x)$ *and* $m_0(x)$ *are linear; (iv)* $e(x)$ *is constant; (v)* $m_1(x) - m_0(x)$ *is a constant function,* $e(x)$ *is linear, and* $p^2 var(X_i \mid Z_i = 1) = (1-p)^2 var(X_i \mid Z_i = 0)$*.*

New regression diagnostics for causal inference

[Chattopadhyay and Z., 2022, *Biometrika*]



Beyond strong ignorability

- These considerations carry over to other settings and designs, e.g.:
  - Difference-in-differences
  - Instrumental variables

Outline

Weighting methods

- Deville and Särndal [1992, *J. Amer. Stat. Assoc.*]
- Kang and Schafer [2007, *Stat. Sci.*], Hirano et al. [2003, *Econometrica*], Kang and Schafer [2007, *Stat. Sci.*], Robins et al. [1994, *J. Amer. Stat. Assoc.*], Rosenbaum [1987, *J. Amer. Stat. Assoc.*]
- Imai and Ratkovic [2014, *J. R. Stat. Soc. B*]
- Athey et al. [2018, *J. R. Stat. Soc. B*], Ben-Michael et al. [2021a, *J. Amer. Stat. Assoc.*; b, *working paper*], Chan et al. [2016, *J. R. Stat. Soc. B*], Hainmueller [2012, *Political Anal.*], Kallus [2020, *J. Mach. Learn. Res.*], Li et al. [2018, *J. Amer. Stat. Assoc.*], Wang and Z. [2020, *Biometrika*], Wong and Chan [2018, *Biometrika*], Yiu and Su [2018, *Biometrika*], Zhao [2018, *Ann. Stat.*], Zhao and Percival [2017], Z. [2015, *J. Amer. Stat. Assoc.*]
- Reviews: Austin and Stuart [2015, *Stat. Med.*], Chattopadhyay et al. [2020, *Stat. Med.*], Ben Michael et al. [2021, *arXiv*]

Two approaches

- Two seemingly unrelated approaches:

Two approaches

- Two seemingly unrelated approaches:
  - The modeling approach:
    - E.g., logistic regression

Two approaches

- Two seemingly unrelated approaches:
  - The modeling approach:
    - E.g., logistic regression
  - The balancing approach:
    - E.g., entropy balancing

Why weighting [Chattopadhyay et al. 2020, *Stat. Med.*]

- Weighting for:
  - Balance
  - Stability
  - Interpolation
  - Generalizability

Bounding bias under general function classes $\mathcal{M}$

[Ben-Michael et al., 2021, *working paper*]

- Estimand:

$$\mu(1) := \mathrm{E}[Y(1)]$$

- Estimator:

$$\hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} Z_i \hat{w}(X_i) Y_i$$

- Error:

$$\hat{\mu}_1 - \mu(1) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} Z_i \hat{w}_i m(X_i,1) - \frac{1}{n}\sum_{i=1}^{n} m(X_i,1)}_{\text{bias from imbalance}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} Z_i \hat{w}_i \varepsilon_i}_{\text{noise}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} m(X_i,1) - \mu(1)}_{\text{sampling variation}}$$

where $\varepsilon_i := Y_i - m(X_i,1)$, $m(x,z) := \mathrm{E}[Y|X=x, Z=z]$

- Bound:

$$|\text{bias}(\hat{\mu}_1)| \leq \text{imbalance}_{\mathcal{M}}(\hat{w}) := \max_{m \in \mathcal{M}} \left|\frac{1}{n}\sum_{i=1}^{n} m(X_i) - \frac{1}{n}\sum_{i=1}^{n} Z_i \hat{w}_i m(X_i)\right|$$

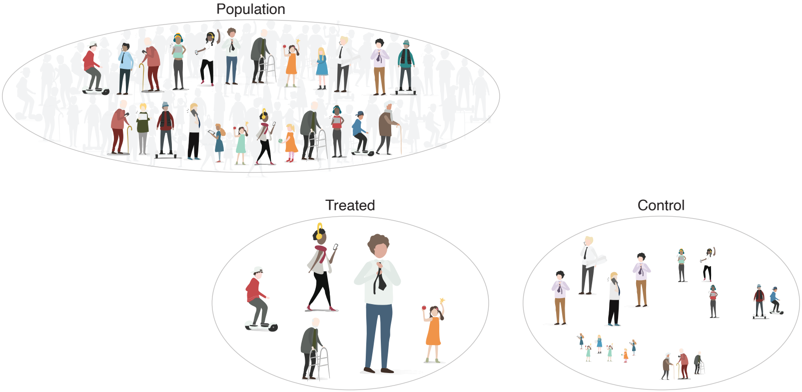
The two approaches to weighting

- Two seemingly unrelated approaches
  - The modeling approach
    - E.g., logistic regression
  - The balancing approach
    - E.g., entropy balancing

Connection

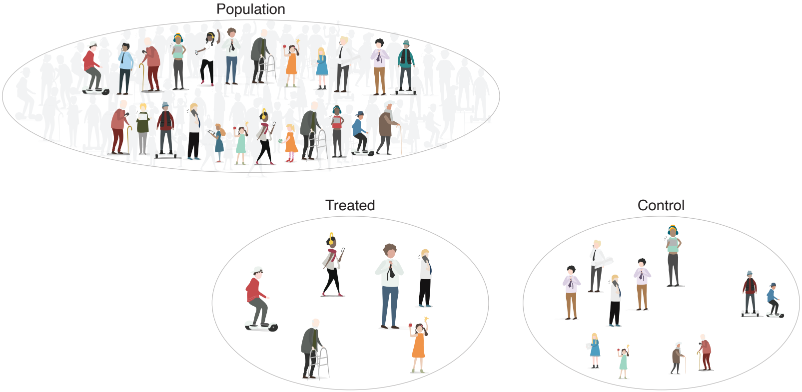
- Both approaches are modeling and balancing
  - But they are solving different optimization problems for the data at hand

Modeling weights



Designs by rawpixel.com / Freepik

Balancing weights



Designs by rawpixel.com / Freepik

Minimal Weights [Wang and Z., 2020, *Biometrika*]

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \sum_{i:Z_i=0} \psi(w_i)$$

$$\text{subject to} \quad \Big| \sum_{i:Z_i=0} w_i B_k(X_i) - \frac{1}{n_t} \sum_{i:Z_i=1} B_k(X_i) \Big| \leq \delta_k, \; k = 1, 2, ..., K$$

Stable Balancing Weights [Z., 2015, *J. Amer. Stat. Assoc.*]

$$
\begin{aligned}
\underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2 \\
\text{subject to} \quad & \Big| \sum_{i:Z_i=0} w_i B_k(X_i) - \frac{1}{n_t} \sum_{i:Z_i=1} B_k(X_i) \Big| \leq \delta_k, \; k = 1, 2, \ldots, K-2 \\
& \sum_{i:Z_i=0} w_i = 1 \\
& w_i \geq 0, \; i : Z_i = 0
\end{aligned}
$$

A quadratic program [Z., 2015, *J. Amer. Stat. Assoc.*]

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2 \\ \text{subject to} \quad & \Big| \sum_{i:Z_i=0} w_i B_k(X_i) - \frac{1}{n_t} \sum_{i:Z_i=1} B_k(X_i) \Big| \leq \delta_k, \ k = 1, 2, ..., K-2 \\ & \sum_{i:Z_i=0} w_i = 1 \\ & w_i \geq 0, \ i : Z_i = 0 \end{aligned}$$

Small weights for big data [Kim et al., 2022, *working paper*]

- Via ADMM and OSQP [Stellato et al., 2020, *Math. Program. Comput.*], we can solve problems with >1M observations in seconds

Small weights for big data [Kim et al., 2022, *working paper*]

- Via ADMM and OSQP [Stellato et al., 2020, *Math. Program. Comput.*], we can solve problems with $>$1M observations in seconds
- For us, the bottleneck became memory allocation rather than computation

"Sample bounded ridge regression"

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2 \\ \text{subject to} \quad & \Big| \sum_{i:Z_i=0} w_i B_k(X_i) - \frac{1}{n_t} \sum_{i:Z_i=1} B_k(X_i) \Big| \leq \delta_k, \; k = 1, 2, ..., K-2 \\ & \sum_{i:Z_i=0} w_i = 1 \\ & w_i \geq 0, \; i : Z_i = 0 \end{aligned}$$

Outline

Optimization design of observational studies

Matching (PM1):

$$\underset{\boldsymbol{m}}{\text{maximize}} \sum_{i:Z_i=0} m_i$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$m_i \in \{0, 1\}, i : Z_i = 0$$

Regression (W-MRI...):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \tilde{w}_i^{\text{base}})^2 / w_i^{\text{scale}}$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

Weighting (SBW):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

$$w_i \geq 0, i : Z_i = 0$$

Optimization design of observational studies

Matching (PM1):

$$\underset{\boldsymbol{m}}{\text{maximize}} \sum_{i:Z_i=0} m_i$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \dots, K$$

$$m_i \in \{0, 1\}, i : Z_i = 0$$

Regression (W-MRI...):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \tilde{w}_i^{\text{base}})^2 / w_i^{\text{scale}}$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \dots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

Weighting (SBW):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \dots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

$$w_i \geq 0, i : Z_i = 0$$

Optimization design of observational studies

Matching (PM1):

$$\underset{\boldsymbol{m}}{\text{maximize}} \sum_{i:Z_i=0} m_i$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$m_i \in \{0, 1\}, i : Z_i = 0$$

Regression (W-MRI...):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \tilde{w}_i^{\text{base}})^2 / w_i^{\text{scale}}$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

Weighting (SBW):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

$$w_i \geq 0, i : Z_i = 0$$

Mathematical programs

Matching (PM1):

$$\underset{\boldsymbol{m}}{\text{maximize}} \sum_{i:Z_i=0} m_i$$

subject to

$$\left| \sum_{i:Z_i=0} m_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$m_i \in \{0, 1\}, i : Z_i = 0$$

Regression (W-MRI...):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \tilde{w}_i^{\text{base}})^2 / w_i^{\text{scale}}$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

Weighting (SBW):

$$\underset{\boldsymbol{w}}{\text{minimize}} \sum_{i:Z_i=0} (w_i - \bar{w}_c)^2$$

subject to

$$\left| \sum_{i:Z_i=0} w_i B_k(X_i) - B_k(\boldsymbol{X}^*) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K$$

$$\sum_{i:Z_i=0} w_i = 1$$

$$w_i \geq 0, i : Z_i = 0$$

From weighting to regression to matching

- Weighting as...
  - ... a convex optimization problem
  - ... a quadratic programming problem
  - `sbw` package for `R`

From weighting to regression to matching

- Weighting as...
  - ... a convex optimization problem
  - ... a quadratic programming problem
  - ◦ `sbw` package for R
- Regression as...
  - ... a least squares optimization problem
  - ... a quadratic programming problem
  - ◦ `lmw` package for R

From weighting to regression to matching

- Weighting as...
  - ... a convex optimization problem
  - ... a quadratic programming problem
  - sbw package for R
- Regression as...
  - ... a least squares optimization problem
  - ... a quadratic programming problem
  - lmw package for R
- Matching as...
  - ... an assignment or network flow optimization problem
  - ... a mixed integer programming problem
  - designmatch package for R

Remarks (1)

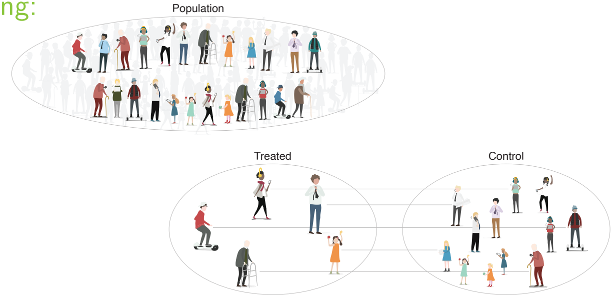
- Where's the experiment?

Remarks (1)

- Where's the experiment?
  - Covariate balance
  - Study representativeness
  - Self-weighted sampling
  - Sample boundedness

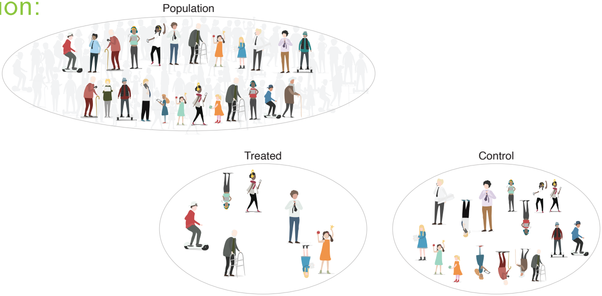
Remarks (1)

- Where's the experiment?
  - Covariate balance
  - Study representativeness
  - Self-weighted sampling
  - Sample boundedness
- Matching:



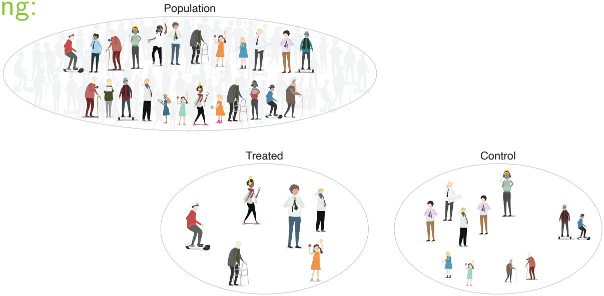
Remarks (1)

- Where's the experiment?
  - Covariate balance
  - Study representativeness
  - Self-weighted sampling
  - Sample boundedness
- Regression:



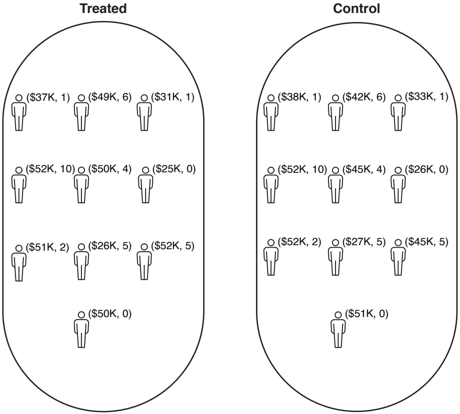
Remarks (1)

- Where's the experiment?
  - Covariate balance
  - Study representativeness
  - Self-weighted sampling
  - Sample boundedness
- Weighting:



Remarks (1): another quick view

Figure 1. Randomized Control Trial

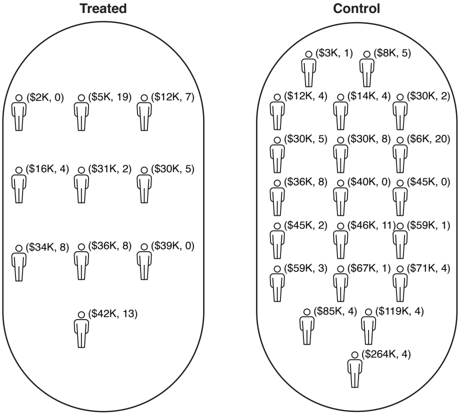


Diagnostic Dashboard: Target Covariate Balance & Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$38,934	$39,325	$39,129
Visits	4.1	4.3	4.2
Sample Size			
Effective	100	100	
Nominal	100	100	

Remarks (1): another quick view

Figure 2. Observational Study, Before Adjustments

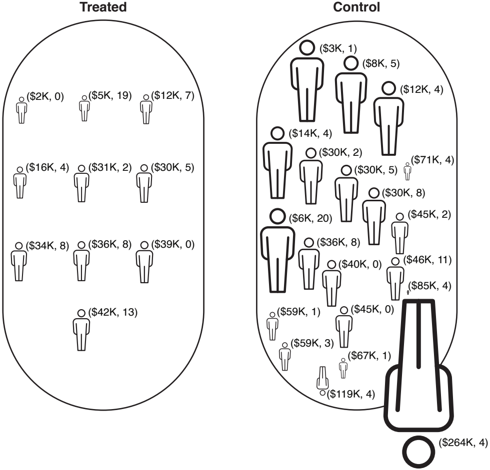


Diagnostic Dashboard: Target Covariate Balance & Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$27,216	$45,086	$27,216
Visits	4.6	3.8	4.6
Sample Size			
Effective	100	200	
Nominal	100	200	

Remarks (1): another quick view

Figure 3. Observational Study, After Regression (URI)

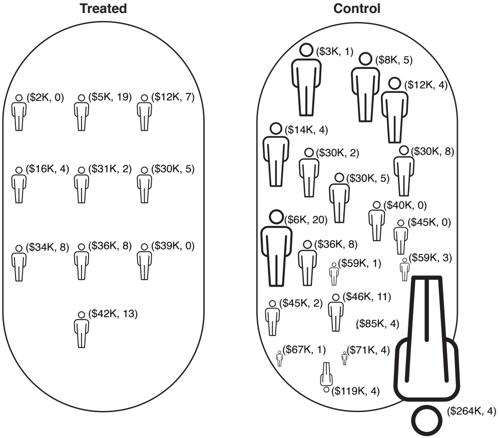


Diagnostic Dashboard:
Target Covariate Balance &
Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$29,600	$29,600	$27,216
Visits	4.3	4.3	4.6
Sample Size			
Effective	98	162	
Nominal	100	200	

Remarks (1): another quick view

Figure 4. Observational Study, After Regression (MRI)

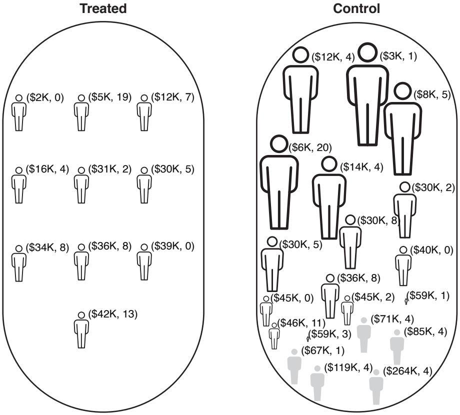


Diagnostic Dashboard: Target Covariate Balance & Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	\$27,216	\$27,216	\$27,216
Visits	4.6	4.6	4.6
Sample Size			
Effective	100	158	
Nominal	100	200	

Remarks (1): another quick view

Figure 5. Observational Study, After Weighting (SBW)

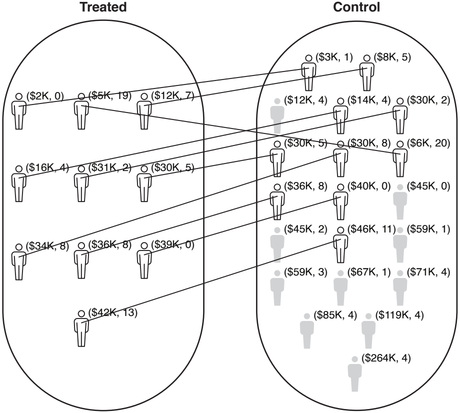


Diagnostic Dashboard:
Target Covariate Balance &
Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$27,216	$27,350	$27,216
Visits	4.6	4.6	4.6
Sample Size			
Effective	100	130	
Nominal	100	200	

Remarks (1): another quick view

Figure 6. Observational Study, After Pair Matching

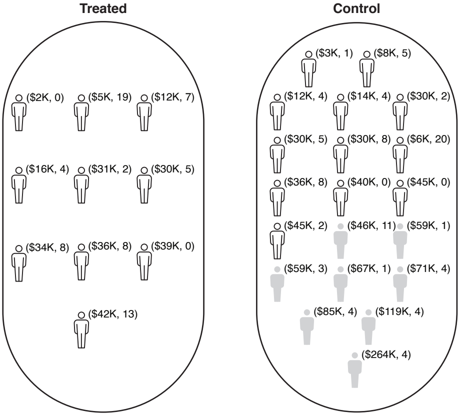


Diagnostic Dashboard: Target Covariate Balance & Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$27,216	$27,829	$27,216
Visits	4.6	4.4	4.6
Sample Size			
Effective	100	100	
Nominal	100	200	

Remarks (1): another quick view

Figure 7. Observational Study, After Profile Matching



Diagnostic Dashboard: Target Covariate Balance & Effective Sample Size

	Treated	Control	Target
Covariate Balance			
Income	$27,216	$27,321	$27,216
Visits	4.6	4.6	4.6
Sample Size			
Effective	100	120	
Nominal	100	200	

Remarks (2)

- Matching, regression, and weighting...
  - ... procedurally, as methods for:
    - Simultaneous covariate adjustment and effect estimation
    - Separate study design and outcome analyses

Remarks (3)

- Matching, regression, and weighting...

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design

$$\text{Matching} \succ \text{weighting} \succ \text{regression}$$

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design

      Matching $\succ$ weighting $\succ$ regression

      A story from the pandemic

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design

      Matching $\succ$ weighting $\succ$ regression

      A story from the pandemic
    - Statistical efficiency

      Regression $\sim$ weighting $\succ$ matching

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design
      Matching $\succ$ weighting $\succ$ regression
      A story from the pandemic
    - Statistical efficiency
      Regression $\sim$ weighting $\succ$ matching
      The implied weights are unconstrained

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design

      Matching $\succ$ weighting $\succ$ regression

      A story from the pandemic
    - Statistical efficiency

      Regression $\sim$ weighting $\succ$ matching

      The implied weights are unconstrained
    - Computational tractability

      Regression $\sim$ weighting $\succ$ matching

Remarks (3)

- Matching, regression, and weighting...
  - ... related but different:
    - Study design
      Matching $\succ$ weighting $\succ$ regression
      A story from the pandemic
    - Statistical efficiency
      Regression $\sim$ weighting $\succ$ matching
      The implied weights are unconstrained
    - Computational tractability
      Regression $\sim$ weighting $\succ$ matching
      In theory, but in practice it depends on the implementation

Outline

Dual representation of the estimand

- Fixed time-point case with binary treatment A
- **Estimand**: $\text{ATE} = E(Y(1)) - E(Y(0))$
- **Assumptions**: Positivity, Exchangeability (Unconfoundedness)

Dual representation of the estimand

- Fixed time-point case with binary treatment A
- **Estimand**: $\text{ATE} = E(Y(1)) - E(Y(0))$
- **Assumptions**: Positivity, Exchangeability (Unconfoundedness)

g-formula

$$E(Y(z)) = \sum_x E(Y|Z=z, X=x)P(X=x) = E\big[E(Y|Z=z,X)\big] = E\big[m_z(X)\big]$$

Dual representation of the estimand

- Fixed time-point case with binary treatment A
- **Estimand**: $\text{ATE} = E(Y(1)) - E(Y(0))$
- **Assumptions**: Positivity, Exchangeability (Unconfoundedness)

g-formula

$$E(Y(z)) = \sum_x E(Y|Z=z, X=x)P(X=x) = E\big[E(Y|Z=z,X)\big] = E\big[m_z(X)\big]$$

Inverse Probability Weighting (IPW)

$$E(Y(z)) = E\left[\frac{\mathbb{1}(Z=z)Y(z)}{P(Z=z|X)}\right] = E\left[\frac{\mathbb{1}(Z=z)Y}{f(Z|X)}\right]$$

Dual representation of the estimand

- Fixed time-point case with binary treatment A
- **Estimand**: $\text{ATE} = E(Y(1)) - E(Y(0))$
- **Assumptions**: Positivity, Exchangeability (Unconfoundedness)

g-formula

$$E(Y(z)) = \sum_x E(Y|Z=z, X=x)P(X=x) = E\left[E(Y|Z=z,X)\right] = E\left[m_z(X)\right]$$

Inverse Probability Weighting (IPW)

$$E(Y(z)) = E\left[\frac{\mathbb{1}(Z=z)Y(z)}{P(Z=z|X)}\right] = E\left[\frac{\mathbb{1}(Z=z)Y}{f(Z|X)}\right]$$

- g-formula-based estimator: $\frac{1}{n}\sum_{i=1}^{n} \hat{m}_1(X_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{m}_0(X_i)$
- IPW-based estimator: $\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{1}(Z_i=1)Y_i}{f(Z_i|X_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{1}(Z_i=0)Y_i}{f(Z_i|X_i)}$

Equivalence between g-formula and IPW

- Assume positivity, i.e. $f(Z|X) > 0$ for all x in the support of X.
- For discrete X, support of X is $\{x : P(X = x) > 0\}$.

$$
\begin{aligned}
E\Big[\frac{\mathbb{1}(Z=z)Y}{f(Z|X)}\Big] &= E\Big[E\Big\{\frac{\mathbb{1}(Z=z)Y}{f(Z|X)}|X\Big\}\Big] \\
&= \sum_x E\Big\{\frac{\mathbb{1}(Z=z)Y}{f(Z|X)}|X=x\Big\}P(X=x) \\
&= \sum_x \frac{1}{f(Z|X)} E\Big\{\mathbb{1}(Z=z)Y|X=x\Big\}P(X=x) \\
&= \sum_x \frac{1}{f(Z|X)} E\Big\{Y|X=x, Z=z\Big\}P(Z=z|X=x)P(X=x) \\
&= \sum_x E(Y|X=x, Z=z)P(X=x) = E(m_z(X))
\end{aligned}
$$

- g-formula-based estimator: $\frac{1}{n}\sum_{i=1}^n \hat{m}_1(X_i) - \frac{1}{n}\sum_{i=1}^n \hat{m}_0(X_i)$
- IPW-based estimator: $\frac{1}{n}\sum_{i=1}^n \frac{\mathbb{1}(Z_i=1)Y_i}{\hat{f}(Z_i|X_i)} - \frac{1}{n}\sum_{i=1}^n \frac{\mathbb{1}(Z_i=0)Y_i}{\hat{f}(Z_i|X_i)}$

Doubly Robust (DR) estimation

- **Estimand**: $E(Y(1))$
- **g-formula-based estimator**: $\frac{1}{n}\sum_{i=1}^{n} m_1(X_i;\hat{\theta})$
- **IPW estimator**: $\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{1}(Z_i=1)Y_i}{f(Z_i|X_i;\hat{\alpha})}$

A DR Estimator of $E(Y(1))$

$$\hat{E}(Y(1))_{DR} = \frac{1}{n}\sum_{i=1}^{n} m_1(X_i,\hat{\theta}) + \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{1}(Z_i=1)Y_i}{f(Z_i|X_i;\hat{\alpha})}(Y_i - m_1(X_i;\hat{\theta}))$$

The estimator is consistent if at least one of the following holds

- $m_1(x;\theta)$ is correctly specified
- $f(z|x;\alpha)$ is correctly specified

Acknowledgments

- Work supported by awards from PCORI and the Alfred P. Sloan Foundation

Causal Inference Methods

José R. Zubizarreta
Harvard University

09/04/2023
CUSO Doctoral School in Statistics and Applied Probability
Saignelégier, Switzerland

Pseudo algorithm of propensity score matching

Algorithm 1 Handling limited overlap with propensity score matching

0. Specify the covariate balance requirements (e.g., mean balance).

Repeat:

1. Estimate the propensity score or another summary of the covariates.
2. Trim the extreme observations according to the summary measure.
3. Match on the summary measure (e.g., using nearest neighbor matching).
4. Assess covariate balance.

Until:

The matched sample satisfies the covariate balance requirements.

Pseudo algorithm of cardinality matching

Algorithm 2 Matching with cardinality matching

0. Specify the covariate balance requirements (e.g., mean balance).
1. Find the largest matched sample that satisfies the covariate balance requirements.
2. Rematch the balanced matched sample to minimize the covariate distances.

Pseudo algorithm of cardinality matching

Algorithm 3 Matching with cardinality matching

0. Specify the covariate balance requirements (e.g., mean balance).
1. Find the largest matched sample that satisfies the covariate balance requirements.
2. Rematch the balanced matched sample to minimize the covariate distances.

Cardinality matching: original formulation

$$
\begin{aligned}
\underset{\boldsymbol{m}}{\text{maximize}} \quad & \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c} \\
\text{subject to} \quad & \sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \ \forall c \in \mathcal{C} \\
& \sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \ \forall t \in \mathcal{T} \\
& \sum_{t \in \mathcal{T}_{p,k}} \sum_{c \notin \mathcal{C}_{p,k}} m_{t,c} = \sum_{t \notin \mathcal{T}_{p,k}} \sum_{c \in \mathcal{C}_{p,k}} m_{t,c}, \ \forall p \in \mathcal{P}, k \in \mathcal{K}(p) \\
& m_{t,c} \in \{0, 1\}, \ t \in \mathcal{T}, c \in \mathcal{C}
\end{aligned}
$$

Cardinality matching: projected [Bennett et al., 2020; *J. Comp. Graph. Stat.*]

$$
\begin{aligned}
\underset{\boldsymbol{x},\boldsymbol{y}}{\text{maximize}} \quad & \sum_{t \in \mathcal{T}} x_t \\
\text{subject to} \quad & \sum_{t \in \mathcal{T}_{p,k}} x_t = \sum_{c \in \mathcal{C}_{p,k}} y_c, \ \forall p \in \mathcal{P}, \ k \in \mathcal{K}(p) \\
& x_t \in \{0,1\}, \ \forall t \in \mathcal{T} \\
& y_c \in \{0,1\}, \ \forall c \in \mathcal{C}
\end{aligned}
$$

A smaller yet equally strong formulation

Proposition

The LP relaxations of the big and small formulations are equivalent.

A smaller yet equally strong formulation

Proposition

The LP relaxations of the big and small formulations are equivalent.

Proposition

The LP relaxations of the big and small formulations are integral if

1. *there are at most two covariates, or*
2. *the covariates are nested.*

In particular, under these conditions the big and small formulations can be solved in polynomial time by solving their LP relaxations.

A smaller yet equally strong formulation

Proposition

The LP relaxations of the big and small formulations are equivalent.

Proposition

The LP relaxations of the big and small formulations are integral if

1. *there are at most two covariates, or*
2. *the covariates are nested.*

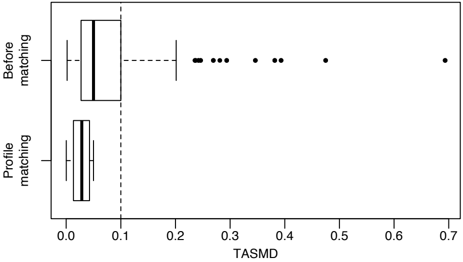
In particular, under these conditions the big and small formulations can be solved in polynomial time by solving their LP relaxations.

Lemma

For three or more covariates, the LP relaxations of the big and small formulations can fail to be integral.

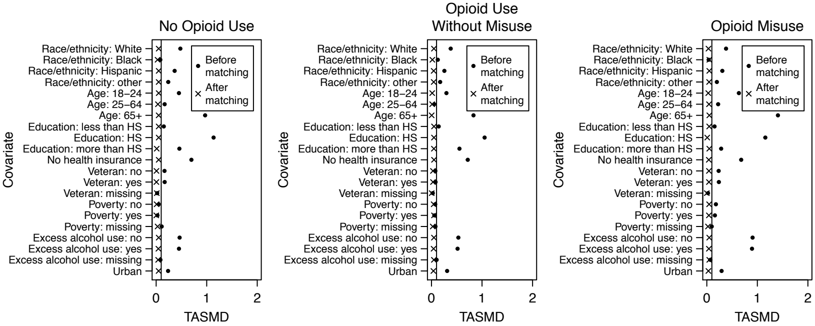
Profile matching toward the sexual minority population

(Cohn and Z., 2022, *Epidemiology*)



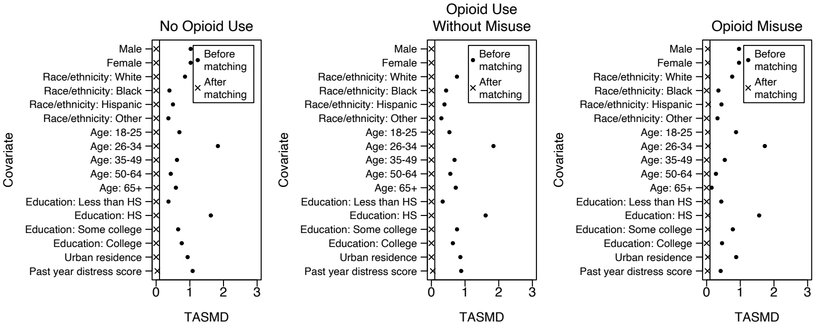
Profile matching toward the Appalachian population

(Cohn and Z., 2022, *Epidemiology*)



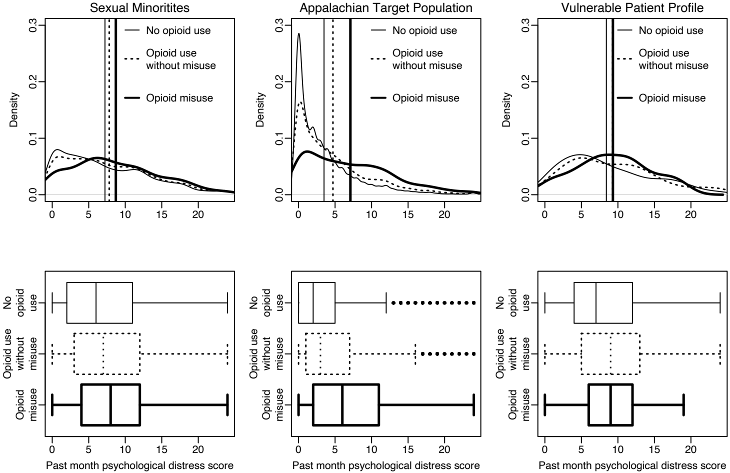
Profile matching toward a vulnerable patient

(Cohn and Z., 2022, *Epidemiology*)



Profile matching toward a vulnerable patient

(Cohn and Z., 2022, *Epidemiology*)



A tuning algorithm

(Wang and Z., 2020, *Biometrika*; Chattopadhyay et al., 2020, *Stat. Med.*)

Algorithm 1 Selection of uniform tuning parameter δ for minimal weights

Fix $\mathcal{D}$, the grid of covariate imbalances (in units of standard deviation).

for $\delta \in \mathcal{D}$, **do**

  Compute the weights w_i, $i = 1, \ldots, n$, for the original sample by solving (4.2.1) with tolerance δ (in standard deviations) for all the balancing constraints.

  for $b \in \{1, \ldots, N_{boot}\}$, *where* N_{boot} *is the number of bootstrap samples,* **do**

    Draw a bootstrap sample S_b from the original sample.

    for $k \in \{1, \ldots, K\}$, *where* K *is the number of balancing constraints,* **do**

      Calculate the covariate imbalance measure $C_{k,b}(\delta)$ corresponding to the kth balancing constraint on S_b.

    end

    Compute the mean imbalance for bth bootstrap sample, i.e., $\xi_b(\delta) = \frac{1}{K} \sum_{k=1}^{K} C_{k,b}(\delta)$.

  end

  Compute the average imbalance over all bootstrap samples, i.e., $\Xi(\delta) = \frac{1}{N_{boot}} \sum_{b=1}^{N_{boot}} \xi_b(\delta)$.

end

Choose $\delta^* = \arg\min_{\delta \in D} \Xi(\delta)$.

Solution methods

- Active-set methods
  - Traditional algorithms for solving QPs
  - Explore the feasible region by adding and dropping constraints

Solution methods

- Active-set methods
  - Traditional algorithms for solving QPs
  - Explore the feasible region by adding and dropping constraints
- Interior-point methods
  - Default algorithm of many commercial solvers
  - Solve unconstrained problem for different barrier functions

Solution methods

- Active-set methods
  - Traditional algorithms for solving QPs
  - Explore the feasible region by adding and dropping constraints
- Interior-point methods
  - Default algorithm of many commercial solvers
  - Solve unconstrained problem for different barrier functions
- Operator splitting methods
  - More recent approach
  - Uses only first order information of the cost function