# Robust estimation of the causal risk difference with misclassified outcome data

Di Shu, PhD

Department of Biostatistics, Epidemiology and Informatics
University of Pennsylvania
and
Center for Pediatric Clinical Effectiveness
Children's Hospital of Philadelphia

September 15, 2021

# Disclosures

- Research to be presented was funded in part by
  - The Canadian Institutes of Health Research (CIHR) Drug Safety and Effectiveness Network
  - The Natural Sciences and Engineering Research Council of Canada (NSERC)
  - The Canadian Statistical Sciences Institute (CANSSI)
- All statements in this presentation are mine and do not necessarily represent the views of above funding bodies
- I have no conflicts of interest to disclose

# Aim

To correct outcome misclassification
in robust estimation of the average treatment effect

# Outline

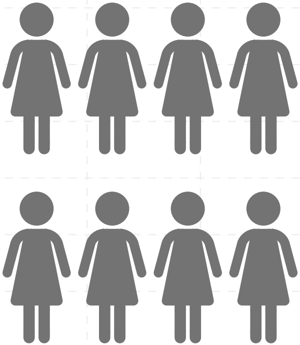
- Impact of outcome misclassification
- Bias correction for
  - inverse probability weighted (IPW) estimation
  - doubly robust (DR) estimation
  - multiply robust (MR) estimation

# Causal inference

- Biomedical research often asks ***causal*** questions
  - Does smoking ***cause*** lung cancer?
  - Does this new drug or vaccine ***cause*** severe adverse reactions?
  - Does treatment A ***cause*** more risk reduction than treatment B?

# Causal inference

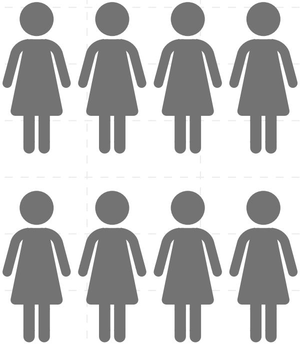
Target population



# Causal inference

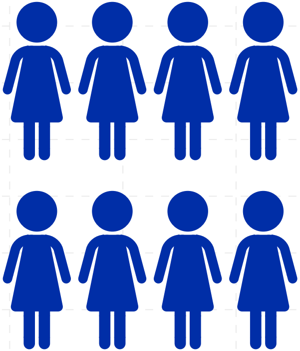
Target population

Drug A vs. Drug B

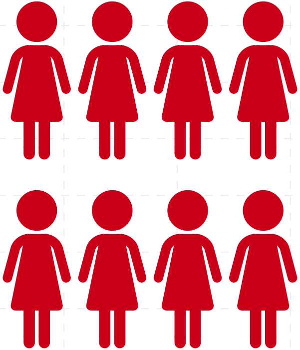


# Causal inference

For the same target population, compare



vs.



# Effect measures for binary outcome

- Causal risk difference or average treatment effect (ATE)
- Causal risk ratio
- Causal odds ratio

# Effect measures for binary outcome

- Causal risk difference or average treatment effect (ATE)
- Causal risk ratio
- Causal odds ratio

# Motivating example: smoking cessation data

- Data structure
  - $X$: a vector of baseline covariates
  - $T$: treatment
  - $Y$: smoking cessation status
- Objective: To estimate $\tau_0 = E\{Y(1)\} - E\{Y(0)\}$, where
  - $Y(1)$: potential outcome that would have been observed had the individual been treated
  - $Y(0)$: potential outcome that would have been observed had the individual been untreated
  - $\tau_0$ is the causal risk difference

# Motivating example: causal interpretation

- $E\{Y(1)\}$: smoking cessation rate that would have been observed had the entire population been treated
- $E\{Y(0)\}$: smoking cessation rate that would have been observed had the entire population been untreated
- Causal effect $\tau_0 = E\{Y(1)\} - E\{Y(0)\}$ compares outcomes in two hypothetical worlds (treatment vs. control) for the same entire population

# Motivating example: challenge

- Outcome of interest: smoking cessation status
- Collected data: self-reported smoking cessation status using
  - phone interview (e.g. Lee et al. 2013)
  - self-administrated online survey
- No biochemical verification of cessation status
- Some smokers might report that they had quit smoking
  - outcomes data are subject to misclassification
  - the true outcome $Y$ is unobserved

Lee SM, Landry J, Jones PM, Buhrmann O, Morley-Forster P. The effectiveness of a perioperative smoking cessation program: a randomized clinical trial. Anesth Analg 2013;117(3):605-13.

# Inverse probability weighted (IPW) estimation

- Let $e = P(T = 1|X)$ be the propensity score (Rosenbaum and Rubin 1983)
- IPW estimator:

$$\hat{\tau} = n^{-1} \sum_{i=1}^{n} \frac{T_i Y_i}{\hat{e}_i} - n^{-1} \sum_{i=1}^{n} \frac{(1 - T_i) Y_i}{1 - \hat{e}_i}$$

- $Y_i$ is unobserved in the presence of misclassification
- Let $Y_i^*$ denote the observed outcome and $\hat{\tau}^*$ the naive estimator

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55.

# Inverse probability weighted (IPW) estimation

- Let $e = P(T = 1|X)$ be the propensity score (Rosenbaum and Rubin 1983)
- IPW estimator:

$$\hat{\tau} = n^{-1} \sum_{i=1}^{n} \frac{T_i Y_i}{\hat{e}_i} - n^{-1} \sum_{i=1}^{n} \frac{(1 - T_i) Y_i}{1 - \hat{e}_i}$$

- $Y_i$ is unobserved in the presence of misclassification
- Let $Y_i^*$ denote the observed outcome and $\hat{\tau}^*$ the naive estimator

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55.

# Why misclassification matters

- Let $p_{ab} = P(Y^* = a|Y = b)$ for $a, b = 0, 1$
  - $p_{11}$: sensitivity
  - $p_{00}$: specificity
  - $p_{10} + p_{01}$: total error

**Theorem 1a (Shu and Yi 2019)**:

- *Naive estimator $\hat{\tau}^* \to (p_{11} - p_{10})\tau_0$ in probability as $n \to \infty$*
- *Asymptotic bias of the naive estimator is $-$total error $\times\ \tau_0$*
- *Asymptotic relative bias is $-$total error*

Shu D, Yi GY. Causal inference with measurement error in outcomes: bias analysis and estimation methods. Stat Methods Med Res 2019;28(7):2049-68.

# Closed-form bias correction

**Theorem 1b (Shu and Yi 2019)**:

- *Propose* $\hat{\tau} = \dfrac{\hat{\tau}^*}{p_{11} - p_{10}} = \dfrac{\textit{naive}}{1 - \textit{total error}} = \dfrac{\textit{naive}}{\textit{sens} + \textit{spec} - 1}$
- $\hat{\tau}$ *is consistent* : $\hat{\tau} \to \tau_0$ *in probability as* $n \to \infty$

- Remark 1: attenuation effect of outcome misclassification
- Remark 2: we can specify or estimate $p_{ab}$ using additional info and data (e.g. validation sample, repeated measures)
- Remark 3: extension that allows for more complicated misclassification model (i.e. $Y^*$ depends on $Y$, $T$ and/or $X$) is available

# Closed-form bias correction

**Theorem 1b (Shu and Yi 2019)**:

- *Propose* $\hat{\tau} = \dfrac{\hat{\tau}^*}{p_{11} - p_{10}} = \dfrac{\textit{naive}}{1 - \textit{total error}} = \dfrac{\textit{naive}}{\textit{sens} + \textit{spec} - 1}$
- $\hat{\tau}$ *is consistent* : $\hat{\tau} \to \tau_0$ *in probability as* $n \to \infty$

- Remark 1: attenuation effect of outcome misclassification
- Remark 2: we can specify or estimate $p_{ab}$ using additional info and data (e.g. validation sample, repeated measures)
- Remark 3: extension that allows for more complicated misclassification model (i.e. $Y^*$ depends on $Y$, $T$ and/or $X$) is available

# Selected simulation results

Performance of the corrected estimator $\hat{\tau}$ in comparison to the naive estimator $\hat{\tau}^*$ with 5000 simulation runs; ReBias$(\%)$: average relative bias; CP$(\%)$: coverage percentage

| setting | | $n = 1000$ | | $n = 5000$ | |
|---|---|---|---|---|---|
| $p_{11} = p_{00}$ | Method | ReBias$(\%)$ | CP$(\%)$ | ReBias$(\%)$ | CP$(\%)$ |
| $0.9$ | Naive | -19.6 | 82.7 | -19.8 | 35.1 |
| | Corrected | 0.6 | 95.1 | 0.2 | 95.4 |
| $0.8$ | Naive | -39.7 | 46.5 | -40.1 | 0.50 |
| | Corrected | 0.5 | 95.3 | -0.2 | 94.7 |
| $0.7$ | Naive | -60.3 | 13.9 | -60.0 | 0.00 |
| | Corrected | -0.7 | 94.6 | 0.1 | 94.6 |

# Real-world application

- Evaluation of a perioperative smoking cessation program (Lee et al. 2013)
- Outcome of interest: smoking cessation status for previous 7 days at the 30-day follow-up postoperatively
- $Y = 1$ if no smoking, $Y = 0$ if still smoking
- Outcome collected: self-reported smoking cessation status answered on the phone, without verification
- Reasonable to assume $p_{11} = 1$
- Misclassification: preoperative self-reported outcomes confirmed by exhaled CO levels $\rightarrow$ specify $p_{10} = 0.075$
- Analysis: $\hat{\tau}^* = 0.170$, $\hat{\tau} = \dfrac{\hat{\tau}^*}{p_{11} - p_{10}} = \dfrac{0.170}{1 - 0.075} = 0.184$

# Real-world application

- Evaluation of a perioperative smoking cessation program (Lee et al. 2013)
- Outcome of interest: smoking cessation status for previous 7 days at the 30-day follow-up postoperatively
- $Y = 1$ if no smoking, $Y = 0$ if still smoking
- Outcome collected: self-reported smoking cessation status answered on the phone, without verification
- Reasonable to assume $p_{11} = 1$
- Misclassification: preoperative self-reported outcomes confirmed by exhaled CO levels $\rightarrow$ specify $p_{10} = 0.075$
- Analysis: $\hat{\tau}^* = 0.170$, $\hat{\tau} = \dfrac{\hat{\tau}^*}{p_{11} - p_{10}} = \dfrac{0.170}{1 - 0.075} = 0.184$

# Double robustness

- The validity of IPW estimators requires propensity score model be correctly specified
- Doubly robust (DR) estimators provide more protection against model misspecification (e.g. Robins et al. 1994, Lunceford and Davidian 2004) by simultaneously using
  - A propensity score model
  - An outcome model

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Am Stat Assoc 1994;89(427):846-66.

Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med 2004;23(19):2937-60.

# Correcting outcome misclassification

- A bias-corrected DR estimator:

$$\hat{\tau}_{DR} = \hat{E}(Y_1) - \hat{E}(Y_0), \text{ where}$$

$$\hat{E}(Y_1) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i^*}{\hat{e}_i(p_{11}-p_{10})} - \frac{T_i - \hat{e}_i}{\hat{e}_i}\hat{q}_{i1} - \frac{T_i}{\hat{e}_i}\left(\frac{p_{10}}{p_{11}-p_{10}}\right)\right\}$$

$$\hat{E}(Y_0) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(1-T_i) Y_i^*}{(1-\hat{e}_i)(p_{11}-p_{10})} + \frac{T_i - \hat{e}_i}{1-\hat{e}_i}\hat{q}_{i0} - \frac{1-T_i}{1-\hat{e}_i}\left(\frac{p_{10}}{p_{11}-p_{10}}\right)\right\}$$

$\hat{q}_{i1}$ is an estimate of $P(Y_i = 1 | T_i = 1, X_i)$ and $\hat{q}_{i0}$ is an estimate of $P(Y_i = 1 | T_i = 0, X_i)$

# Correcting outcome misclassification

- $\hat{q}_{i1}$ and $\hat{q}_{i0}$ cannot be calculated by fitting the postulated outcome model, because the true value $Y$ is unobserved
- Maximize the observed likelihood instead

- Let $\beta$ be the outcome model parameters
- Observed likelihood function contributed from subject $i$:

$$\begin{aligned} L_i(\beta) &= P(Y_i = 1|T_i, X_i; \beta)\{p_{11}Y_i^* + (1-p_{11})(1-Y_i^*)\} \\ &\quad + P(Y_i = 0|T_i, X_i; \beta)\{p_{10}Y_i^* + (1-p_{10})(1-Y_i^*)\} \end{aligned}$$

# Correcting outcome misclassification

- $\hat{q}_{i1}$ and $\hat{q}_{i0}$ cannot be calculated by fitting the postulated outcome model, because the true value $Y$ is unobserved
- Maximize the observed likelihood instead

- Let $\boldsymbol{\beta}$ be the outcome model parameters
- Observed likelihood function contributed from subject $i$:

$$\begin{aligned} L_i(\boldsymbol{\beta}) \quad = \quad & P(Y_i = 1 | T_i, X_i; \boldsymbol{\beta})\{p_{11}Y_i^* + (1 - p_{11})(1 - Y_i^*)\} \\ & + P(Y_i = 0 | T_i, X_i; \boldsymbol{\beta})\{p_{10}Y_i^* + (1 - p_{10})(1 - Y_i^*)\} \end{aligned}$$

# Correcting outcome misclassification

**Theorem 2 (Shu and Yi 2019)**: *The proposed estimator $\hat{\tau}_{DR}$ is doubly robust, i.e., it is consistent when either the propensity score model $T \sim X$ or the outcome model $Y \sim (X, T)$ is correctly specified.*

- Remark: extension that allows for more complicated misclassification model (i.e. $Y^*$ depends on $Y$, $T$ and/or $X$) is available

# Selected simulation results

Performance of the DR estimator $\hat{\tau}_{DR}$ with 5000 simulation runs;
Scenario: whether the treatment and outcome models are correct or not

| | | $p_{11} = 0.9,\ p_{10} = 0.1$ | | $p_{11} = 0.8,\ p_{10} = 0.2$ | |
|---|---|---|---|---|---|
| $n$ | Scenario | ReBias(%) | CP% | ReBias(%) | CP% |
| 2000 | ✔ ✔ | 0.0 | 95.1 | -0.9 | 94.2 |
| | ✔ ✘ | -0.9 | 95.0 | 0.5 | 93.5 |
| | ✘ ✔ | 0.1 | 96.4 | 0.9 | 94.4 |
| 5000 | ✔ ✔ | 0.1 | 95.5 | -0.4 | 94.8 |
| | ✔ ✘ | 1.0 | 96.1 | 0.3 | 95.1 |
| | ✘ ✔ | 0.3 | 95.4 | 0.3 | 96.0 |

# Multiple robustness

- What if both the treatment and outcome models are wrong?
- Han and Wang (2013) developed a multiply robust (MR) method by simultaneously using
  - A set of propensity score models
  - A set of outcome models
- Only one model needs to be correct

Han P, Wang L. Estimation with missing data: beyond double robustness. Biometrika 2013;100(2):417-30.

# Multiple robustness

- What if both the treatment and outcome models are wrong?
- Han and Wang (2013) developed a multiply robust (MR) method by simultaneously using
  - A set of propensity score models
  - A set of outcome models
- Only one model needs to be correct

Han P, Wang L. Estimation with missing data: beyond double robustness. Biometrika 2013;100(2):417-30.

# Specifying two model sets

- Let $e(X) = P(T = 1|X)$ be the true treatment model and $q_t(X) = P(Y = 1|X, T = t)$ the true outcome model
- Individuals $i = 1, \ldots, m$ are treated, $i = m + 1, \ldots, n$ are untreated
- Postulate
  - $\mathcal{E} = \{e^j(\boldsymbol{\gamma}^j; X), j = 1 \ldots, J\}$: a set of $J$ treatment models
  - $\mathcal{Q} = \{q_t^k(\boldsymbol{\beta}^k; X), k = 1 \ldots, K\}$: a set of $K$ outcome models

# MR estimation without misclassification

- Applying the method of Han and Wang (2013):
  - Step 1: obtain $\hat{\gamma}^j$ by fitting the $j$th treatment model, for each $j$
  - Step 2: obtain $\hat{\beta}^k$ by fitting the $k$th outcome model, for each $k$
  - Step 3: calculate $\hat{w}_i$ and $\tilde{w}_i$ [that use $\hat{\gamma}^j$ and $\hat{\beta}^k$]
  - Step 4: calculate

$$\hat{E}(Y_1) = \sum_{i=1}^{m} \hat{w}_i Y_i \quad \text{and} \quad \hat{E}(Y_0) = \sum_{i=m+1}^{n} \tilde{w}_i Y_i$$

  - Step 5: calculate $\hat{\tau} = \hat{E}(Y_1) - \hat{E}(Y_0)$

# MR estimation without misclassification

- Define
  - $\hat{\theta}^j = n^{-1} \sum_{i=1}^n e^j(\hat{\boldsymbol{\gamma}}^j; X_i)$ for $j = 1, \ldots, J$
  - $\hat{\eta}_1^k = n^{-1} \sum_{i=1}^n q_1^k(\hat{\boldsymbol{\beta}}^k; X_i)$ and $\hat{\eta}_0^k = n^{-1} \sum_{i=1}^n q_0^k(\hat{\boldsymbol{\beta}}^k; X_i)$ for $k = 1, \ldots, K$
- $\hat{w}_i = \left\{ \frac{1}{m} \frac{1}{1 + \hat{\rho}^T \hat{g}_i(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}})} \right\} \Big/ \left\{ \frac{1}{m} \sum_{i=1}^m \frac{1}{1 + \hat{\rho}^T \hat{g}_i(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}})} \right\}$
  - $\hat{\rho}$ solves $\sum_{i=1}^m \hat{g}_i(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}) / \{1 + \rho^T \hat{g}_i(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}})\} = \mathbf{0}$ for $\rho$, with $\hat{g}_i(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}) = (e^1(\hat{\boldsymbol{\gamma}}^1; X_i) - \hat{\theta}^1, \ldots, e^J(\hat{\boldsymbol{\gamma}}^J; X_i) - \hat{\theta}^J, q_1^1(\hat{\boldsymbol{\beta}}^1; X_i) - \hat{\eta}_1^1, \ldots, q_1^K(\hat{\boldsymbol{\beta}}^K; X_i) - \hat{\eta}_1^K)^T$
- Calculate $\tilde{w}_i$ similarly

# MR estimation without misclassification

- Applying the method of Han and Wang (2013):
  - Step 1: obtain $\hat{\gamma}^j$ by fitting the $j$th treatment model, for each $j$
  - Step 2 : obtain $\hat{\beta}^k$ by fitting the $k$th outcome model, for each $k$
  - Step 3: calculate $\hat{w}_i$ and $\tilde{w}_i$ [that use $\hat{\gamma}^j$ and $\hat{\beta}^k$]
  - Step 4 : calculate

$$\hat{E}(Y_1) = \sum_{i=1}^{m} \hat{w}_i Y_i \quad \text{and} \quad \hat{E}(Y_0) = \sum_{i=m+1}^{n} \tilde{w}_i Y_i$$

  - Step 5: calculate $\hat{\tau} = \hat{E}(Y_1) - \hat{E}(Y_0)$

# MR estimation with outcome misclassification

- Misclassification: $p_{ab} = P(Y^* = a|Y = b)$ for $a, b = 0, 1$.
- Correction method (Shu and Yi 2021 submitted)
  - Step 1: obtain $\hat{\gamma}^j$ by fitting the $j$th treatment model, for each $j$
  - Step 2: obtain $\hat{\beta}^k$ by maximizing the observed likelihood, for each $k$
  - Step 3: calculate $\hat{w}_i$ and $\tilde{w}_i$
  - Step 4: calculate $\hat{E}(Y_1) = \sum_{i=1}^{m} \frac{\hat{w}_i Y_i^*}{p_{11} - p_{10}} - \frac{p_{10}}{p_{11} - p_{10}}$

$$\hat{E}(Y_0) = \sum_{i=m+1}^{n} \frac{\tilde{w}_i Y_i^*}{p_{11} - p_{10}} - \frac{p_{10}}{p_{11} - p_{10}}$$

  - Step 5: calculate $\hat{\tau}_{MR} = \hat{E}(Y_1) - \hat{E}(Y_0)$

# MR estimation with outcome misclassification

**Theorem 3 (Shu and Yi 2021 submitted)**: *The proposed estimator $\hat{\tau}_{MR}$ is multiply robust, i.e., it is consistent when either $\mathcal{E}$ or $\mathcal{Q}$ contains a correctly specified model.*

Shu D, Yi GY. Multiply robust estimation of causal treatment effects with binary outcome data subject to misclassification. Submitted.

# Selected simulation results

Performance of the MR estimator $\hat{\tau}_{MR}$ with 5000 simulation runs when $(p_{11}, p_{10}) = (0.7, 0.3)$; Scenarios I or II: only the treatment model set or only the outcome model set contains a correct model

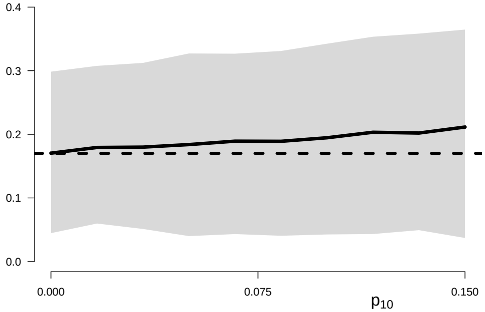
| | | $n = 2000$ | | $n = 5000$ | |
|---|---|---|---|---|---|
| Scenario | Method | ReBias% | CP% | ReBias% | CP% |
| I | Naive | -59.76 | 1.50 | -60.39 | 0.1 |
| | $\hat{\tau}_{MR}$ | 0.64 | 93.6 | -0.90 | 93.7 |
| II | Naive | -60.32 | 1.50 | -60.47 | 0.0 |
| | $\hat{\tau}_{MR}$ | -0.48 | 94.9 | -0.59 | 94.2 |

# Smoking cessation data revisited

- Naive: $\hat{\tau}^* = 0.170$ (CI $0.041, 0.307$)

- Reasonable to assume $p_{11} = 1$
- Misclassification: use preoperative data $\rightarrow$ specify $p_{10} = 0.075$
- Corrected: $\hat{\tau}_{MR} = 0.189$ (CI $0.051, 0.324$)

# Smoking cessation data revisited




Sensitivity analysis with $p_{10}$ ranging from 0 to 0.15. Solid line: point estimates; Grey region: bootstrap percentile 95% CIs. Dashed line: naive estimate (i.e. 0.170)

# Outcome misclassification + X

DOI: 10.1002/sim.8419

RESEARCH ARTICLE

WILEY Statistics in Medicine

**Causal inference with noisy data: Bias analysis and estimation approaches to simultaneously addressing missingness and misclassification in binary outcomes**

Di Shu[1,3] | Grace Y. Yi[2,3]

DOI: 10.1002/sim.8073

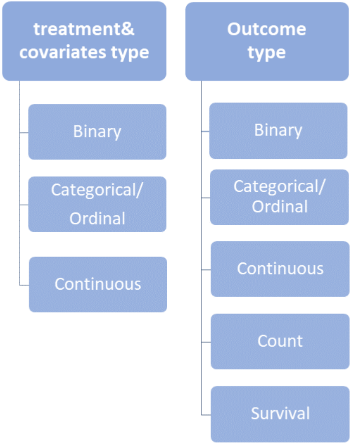
RESEARCH ARTICLE

WILEY Statistics in Medicine

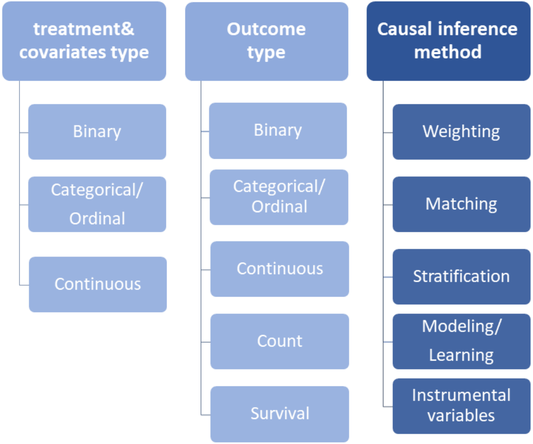
**Weighted causal inference methods with mismeasured covariates and misclassified outcomes**

Di Shu | Grace Y. Yi

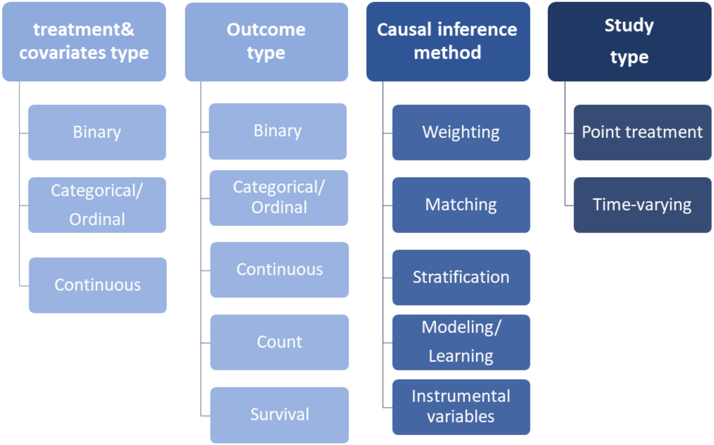
# Future topics



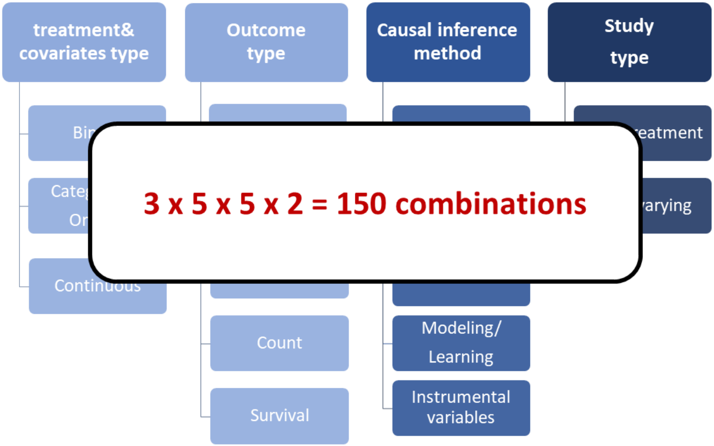
# Future topics



# Future topics



# Future topics



# Acknowledgement

Joint work with Dr. Grace Yi, University of Western Ontario

# Thank you!

**Di.Shu@pennmedicine.upenn.edu**