Propensity score matching after multiple imputation when a confounder has missing data

Corentin Ségalas, Clémence Leyrat, James Carpenter and Elizabeth Williamson

corentin.segalas@u-bordeaux.fr

JdS 2023, Bruxelles

Statistical framework

Observational study with:

- Y_i a binary outcome
- Z_i a binary exposure (1 if patient i treated, 0 if not)
- X_i a vector of baseline covariates (all potential confounders)

Statistical framework

Observational study with:

- Y_i a binary outcome
- Z_i a binary exposure (1 if patient i treated, 0 if not)
- X_i a vector of baseline covariates (all potential confounders)

Average treatment effect on the treated (ATT)

$$ATT = E(Y_i^1|Z_i = 1) - E(Y_i^0|Z_i = 1)$$

Y_i^1 and Y_i^0 are potential outcomes

Propensity score (PS)

Definition (Rosenbaum and Rubin, 1983)

For patient i,

$$\pi_i = P(z_i = 1 | x_i)$$

estimated using a logistic regression or more advanced techniques (Westreich et al., 2010)

Propensity score (PS)

Definition (Rosenbaum and Rubin, 1983)

For patient i,

$$\pi_i = P(z_i = 1|x_i)$$

estimated using a logistic regression or more advanced techniques (Westreich et al., 2010)

Unbiased estimator of the true ATT: PS matching, PS stratification, inverse probability of treatment weighting, etc.

PS matching

Match each treated patient to at least one untreated patient based on the distance between their PS.

PS matching

Match each treated patient to at least one untreated patient based on the distance between their PS.

- matching algorithm
- metric for the distance
- caliper whose size limits the difference between a pair
- number of non-treated patients matched to each treated patient
- sampling with or without replacement

PS matching

Match each treated patient to at least one untreated patient based on the distance between their PS.

- matching algorithm
- metric for the distance
- caliper whose size limits the difference between a pair
- number of non-treated patients matched to each treated patient
- sampling with or without replacement

+ direct estimation of the ATT

PS matching

Match each treated patient to at least one untreated patient based on the distance between their PS.

- matching algorithm
- metric for the distance
- caliper whose size limits the difference between a pair
- number of non-treated patients matched to each treated patient
- sampling with or without replacement

+ direct estimation of the ATT
− loss of power

Multiple imputation and Rubin's rules

How to obtain $\hat{\theta}$, estimate of the ATT, when X has missing data.

Multiple imputation and Rubin's rules

How to obtain $\hat{\theta}$, estimate of the ATT, when X has missing data.
If missing at random (MAR): multiple imputation. $(\hat{\theta}_k)_k$ are aggregated using Rubin's rules (Leyrat et al. 2019, Granger et al. 2019):

$$\hat{\theta} = \frac{1}{m} \sum_{k=1}^{m} \hat{\theta}_k$$

Multiple imputation and Rubin's rules

How to obtain $\hat{\theta}$, estimate of the ATT, when X has missing data. If missing at random (MAR): multiple imputation. $(\hat{\theta}_k)_k$ are aggregated using Rubin's rules (Leyrat et al. 2019, Granger et al. 2019):

$$\hat{\theta} = \frac{1}{m}\sum_{k=1}^{m}\hat{\theta}_k, \quad \widehat{Var}(\hat{\theta}) = W + \left(1 + \frac{1}{m}\right) B$$

where

$$W = \frac{1}{m}\sum_{k=1}^{m}\widehat{Var}(\hat{\theta}_k), \quad B = \frac{1}{m-1}\sum_{k=1}^{m}(\hat{\theta}_k - \hat{\theta})^2.$$

Motivation and objectives

In a simulation study, over-coverage of the ATT was observed with classic PS matching.

Motivation and objectives

In a simulation study, over-coverage of the ATT was observed with classic PS matching.

Literature insight (Reiter, 2008)

Rubin's rules could lead to inflated variance when some patients contributed to the imputation model but not to the analysis model

Motivation and objectives

In a simulation study, over-coverage of the ATT was observed with classic PS matching.

Literature insight (Reiter, 2008)

Rubin's rules could lead to inflated variance when some patients contributed to the imputation model but not to the analysis model

1. Does the discarding of unmatched individuals lead to over-coverage when combining multiple imputation and propensity score matching using Rubin's rules?

Motivation and objectives

In a simulation study, over-coverage of the ATT was observed with classic PS matching.

Literature insight (Reiter, 2008)

Rubin's rules could lead to inflated variance when some patients contributed to the imputation model but not to the analysis model

1. Does the discarding of unmatched individuals lead to over-coverage when combining multiple imputation and propensity score matching using Rubin's rules?

2. Implement the Reiter's correction in the context of propensity score matching and assess its performance

Reiter's rules

Reiter proposed to create r (instead of 1) complete datasets for each parameter draw leading to a total of $m \times r$ complete datasets.

Reiter's rules

Reiter proposed to create r (instead of 1) complete datasets for each parameter draw leading to a total of $m \times r$ complete datasets.

$$\hat{\theta} = \frac{1}{mr} \sum_{k=1}^{m} \sum_{j=1}^{r} \hat{\theta}_{k,j} = \frac{1}{m} \sum_{k=1}^{m} \hat{\theta}_k$$

Reiter's rules

Reiter proposed to create r (instead of 1) complete datasets for each parameter draw leading to a total of $m \times r$ complete datasets.

$$\hat{\theta} = \frac{1}{mr} \sum_{k=1}^{m} \sum_{j=1}^{r} \hat{\theta}_{k,j} = \frac{1}{m} \sum_{k=1}^{m} \hat{\theta}_k,$$

$$\widehat{Var}(\hat{\theta}) = \tilde{W} + \left(1 + \frac{1}{m}\right) \tilde{B} - \left(1 + \frac{1}{r}\right) U,$$

where

$$\tilde{W} = \frac{1}{mr} \sum_{k=1}^{m} \sum_{j=1}^{r} \widehat{Var}(\hat{\theta}_{k,j}), \quad \tilde{B} = \frac{1}{m-1} \sum_{k=1}^{m} \left(\hat{\theta}_k - \hat{\theta}\right)^2,$$

$$U = \frac{1}{m(r-1)} \sum_{k=1}^{m} \sum_{j=1}^{r} \left(\hat{\theta}_{k,j} - \hat{\theta}_k\right)^2.$$

Simulation study

- Aims: assess the impact of discarding patients between imputation and estimation and evaluate Reiter's rules in this context
- Data generation mechanisms:
  - $N = 1,000$ datasets with $10,000$ patients
  - three confounders $x = (x_1, x_2, x_3) \sim \mathcal{N}(0, I_3)$
  - three levels of confounding: strong, moderate and weak
  - 30%, 20% or 10% of treated patients
  - around 15% of missing at random x_2
- Estimands: ATT as an odds-ratio

Simulation study

- Method implemented using `R`:
  - multiple imputation using `mice` (`trace` argument)
  - PS estimation using `glm`
  - PS matching using `MatchIt`
  - ATT estimation using `glm.cluster` from `miceadds`
  - aggregation of the results using Rubin's and Reiter's rules

- Performance measures: relative bias, 95% confidence intervals coverage rate (CR)

Results: Rubin's rules

Table: Results for the 1,000 replicates of the ATT estimation using Rubin's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	-0.010	0.985
Strong	20	-0.001	0.996
Strong	10	0.005	0.999
Moderate	30	-0.001	0.994
Moderate	20	0.002	0.996
Moderate	10	0.004	1.000
Weak	30	-0.001	0.989
Weak	20	0.000	0.990
Weak	10	0.004	0.994

Results: Rubin's rules

Table: Results for the 1,000 replicates of the ATT estimation using Rubin's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	-0.010	0.985
Strong	20	-0.001	0.996
Strong	10	0.005	0.999
Moderate	30	-0.001	0.994
Moderate	20	0.002	0.996
Moderate	10	0.004	1.000
Weak	30	-0.001	0.989
Weak	20	0.000	0.990
Weak	10	0.004	0.994

Results: Rubin's rules

Table: Results for the 1,000 replicates of the ATT estimation using Rubin's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	-0.010	0.985
Strong	20	-0.001	0.996
Strong	10	0.005	0.999
Moderate	30	-0.001	0.994
Moderate	20	0.002	0.996
Moderate	10	0.004	1.000
Weak	30	-0.001	0.989
Weak	20	0.000	0.990
Weak	10	0.004	0.994

Results: Reiter's rules

Table: Results for the 1,000 replicates of the ATT estimation using Reiter's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	0.014	0.937
Strong	20	0.000	0.950
Strong	10	0.000	0.950
Moderate	30	0.001	0.933
Moderate	20	0.001	0.956
Moderate	10	0.002	0.958
Weak	30	0.001	0.946
Weak	20	0.000	0.940
Weak	10	0.002	0.958

Results: Reiter's rules

Table: Results for the 1,000 replicates of the ATT estimation using Reiter's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	0.014	0.937
Strong	20	0.000	0.950
Strong	10	0.000	0.950
Moderate	30	0.001	0.933
Moderate	20	0.001	0.956
Moderate	10	0.002	0.958
Weak	30	0.001	0.946
Weak	20	0.000	0.940
Weak	10	0.002	0.958

Results: Reiter's rules

Table: Results for the 1,000 replicates of the ATT estimation using Reiter's rules

Confounding	% of treated	Rel. bias	CR
Strong	30	0.014	0.937
Strong	20	0.000	0.950
Strong	10	0.000	0.950
Moderate	30	0.001	0.933
Moderate	20	0.001	0.956
Moderate	10	0.002	0.958
Weak	30	0.001	0.946
Weak	20	0.000	0.940
Weak	10	0.002	0.958

Application: context

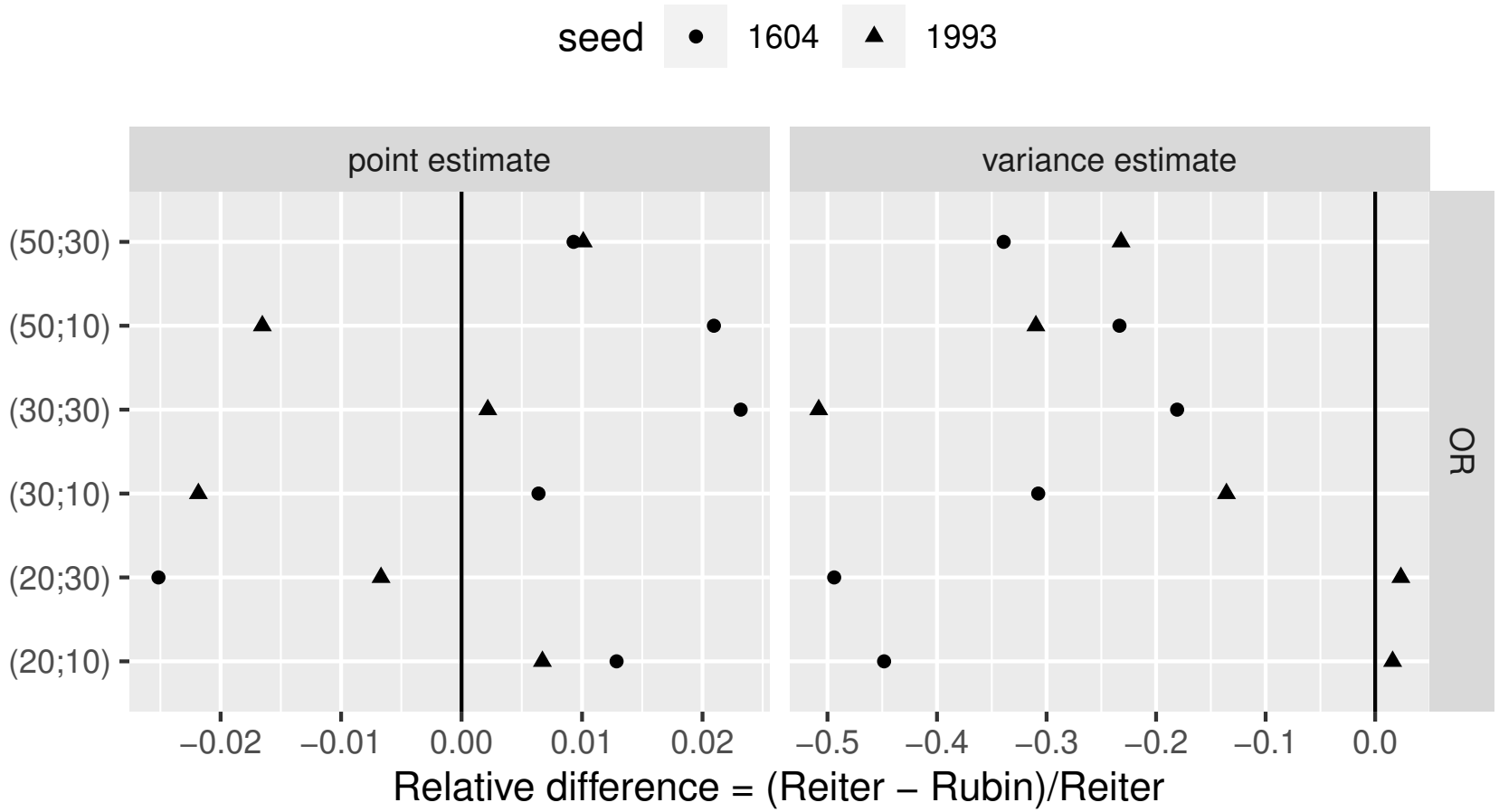
National Cancer Registry of the Office for National Statistics:

- $31,351$ patients diagnosed with cancer
- covariates: stage of the cancer, sex of the patient, patient's level of deprivation, comorbidity (Charlson score) and the patient's performance status
- 25% of performance and 10% of stage data were missing

We have studied the effect of age at diagnosis as a binary variable (median as the cutoff) on the risk of surgery

- impact of $(m; r) = (20; 10)$, $(20; 30)$, $(30; 10)$, $(30; 30)$, $(50; 10)$ and $(50; 30)$
- impact of random fluctuation: 1604 and 1993 as seeds

Application: results



Discussion

- Combination of MI and PS matching using Rubin's rules can lead to inflated variance
- Reiter's rules were able to correct the inflation
- Focus on PS matching only
- Easy to implement in `R`
- Computationally intense with bigger sample sizes ($m \times r$ imputation)
- What about full matching?

Discussion

- Combination of MI and PS matching using Rubin's rules can lead to inflated variance
- Reiter's rules were able to correct the inflation
- Focus on PS matching only
- Easy to implement in `R`
- Computationally intense with bigger sample sizes ($m \times r$ imputation)
- What about full matching?

Take home message

Be careful when combining multiple imputation and propensity score matching

References

- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika
- Westreich, D., Lessler, J., and Funk, M. J. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. Journal of Clinical Epidemiology
- Granger, E., Sergeant, J. C., and Lunt, M. (2019). Avoiding pitfalls when combining multiple imputation and propensity scores. Statistics in Medicine
- Leyrat, C., Seaman, S. R., White, I. R., Douglas, I., Smeeth, L., Kim, J., Resche-Rigon, M., Carpenter, J. R., and Williamson, E. J. (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used? Statistical Methods in Medical Research
- Reiter, J. P. (2008). Multiple Imputation When Records Used for Imputation Are Not Used or Disseminated for Analysis. Biometrika

Paper

Ségalas C, Leyrat C, Carpenter JR, Williamson E. Propensity score matching after multiple imputation when a confounder has missing data. Statistics in Medicine, 2023.

Code

`https://github.com/crsgls/psmatching`

Thanks for your attention!

corentin.segalas@u-bordeaux.fr