How should the propensity score be estimated when some confounders are partially observed?

Clémence Leyrat[1],
James Carpenter[1,2], **Elizabeth Williamson**[1,3], **Helen Blake**[1]

[1]Department of Medical statistics, LSHTM, London
[2]MRC Clinical Trials Unit at UCL, London
[3]Farr Institute of Health Informatics, London

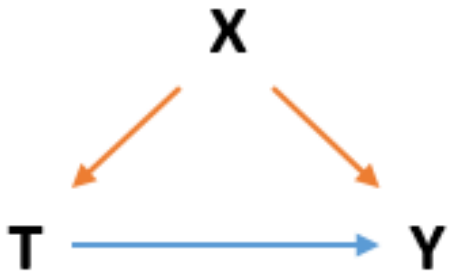
MRC project grant MR/M013278/1

Content

The problem of confounding

Observational studies are a useful source of information to **establish causal effects** of a treatment/exposure on a health-related outcome

Because of the lack of randomisation, study groups may be **unbalanced** $\Longrightarrow$ Risk of confounding bias



T: treatment
Y: outcome
X: confounder

Propensity scores (PS) proposed in 1983 to **balance groups** in observational studies

The propensity score

The PS is the **individual's probability of receiving the treatment** rather than the control conditionally to their baseline characteristics

$$e(x) = P(T = 1|\mathbf{X} = \mathbf{x})$$

The true value of the PS is **unknown** but can be estimated:
$\Longrightarrow$ individual predictions from a logistic model

Covariates to be included:

- true confounders
- risk factors

Assumptions required

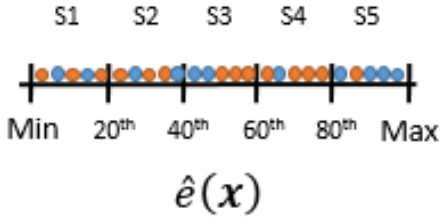
3 **assumptions required** to estimate unbiased causal effects using the PS:

- Positivity: each patient has a non null probability of receiving the treatment or the control
- SITA (conditional exchangeability): no unmeasured confounders
- SUTVA (consistency):
  - the potential outcome for a patient is not affected by the treatment received by the other patients
  - the treatment has **always the same effect** on a given patient

Under these assumptions, the PS is a **balancing score**

PS-based approaches

Stratification

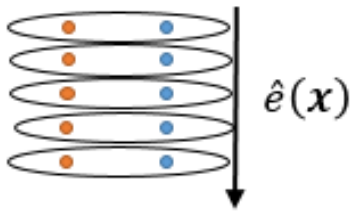


Adjustment

$$Y = \alpha_0 + \alpha_1 T + \alpha_2 \hat{e}(\mathbf{x})$$

- Treated
- Untreated

Matching



Inverse weighting (IPTW)

- $w = \frac{1}{\hat{e}(x)}$
- $w = \frac{1}{1-\hat{e}(x)}$

The issue of missing data

If some confounders are partially observed, the PS **cannot be estimated** for individuals without a complete record

T	X_1	X_2	X_3	$\hat{e}(\mathbf{x})$
0	x_{11}	x_{21}	x_{31}	$\hat{e}_1(\mathbf{x})$
0	x_{12}	?	x_{32}	?
1	x_{13}	?	?	?
0	x_{14}	x_{24}	x_{34}	$\hat{e}_4(\mathbf{x})$
1	x_{15}	x_{25}	?	?
...	...	...	...	...

Missingness mechanisms

The PS estimation and analysis strategy depend on the association between the missing value and observed and unobserved variables, the **missingness mechanism**

Following Rubin's taxonomy, missing confounders can be:

- **MCAR** (missing completely at random)
- **MAR** (missing at random)
- **MNAR** (missing not at random)

What can be done?

Strategies investigated

Focus on 3 approaches (with a binary outcome) for IPTW:

- Complete case analysis
- The missingness pattern approach
- Multiple imputation

For each of them:

- What are the assumptions required?
- What is the best way to implement the method?

In practice

A quick check of the literature showed that, among 132 identified papers:

- 46% used complete case analysis
- 5% used the missingness pattern approach
- 36% used multiple imputation

A systematic review would be needed for a better overview of the different methods implemented in practice.

Complete case analysis

CC for multivariable regression

Complete case (CC) analysis: analysis on the subgroup of patients with **complete records**:

- **Loss of efficiency** because of a loss in sample size
- Risk of **bias** of the treatment effect estimate

CC analysis leads to an **unbiased estimate**:

- when data are **MCAR**
- when missingness **does not depend on Y and T** in the context of multivariable logistic regression

CC for multivariable regression (2)

Table 1. Bias of Estimates Derived From Complete Records Analysis Logistic Regression Under Different Missingness Assumptions

Quantity on Which Missingness Is Dependent	Parameter		
	β_0	β_X	β_C
Neither Y nor X nor C	Asymptotically unbiased	Asymptotically unbiased	Asymptotically unbiased
Outcome (Y)	Biased	Asymptotically unbiased	Asymptotically unbiased
Covariates (X, C, or both)	Asymptotically unbiased	Asymptotically unbiased	Asymptotically unbiased
Outcome (Y) and confounders (C)	Biased	Asymptotically unbiased	Biased
Outcome (Y), exposure (X), and possibly confounders (C)	Biased	Biased[a]	Biased

J.W. Bartlett, O. Harel, and J.R. Carpenter. Asymptotically Unbiased Estimation of Exposure Odds Ratios in Complete Records Logistic Regression. American Journal of Epidemiology. 2015 Oct 15;182(8):730-6.

Are these results generalizable to PS analysis?

Simulation study

Setting:
n=10000, binary outcome Y, binary treatment T, and two binary confounders C_1 and C_2

R is the complete case indicator (R=1 if complete case, 0 otherwise)

Comparison of 3 approaches:

- Multivariable logistic regression to estimate the conditional OR
- Multivariable logistic regression to estimate the marginal OR
- IPTW to estimate the marginal OR

Results

Bias of log(OR). ORcond=2

Variables associated with missingness	Multivariable regression		IPTW	
	ORcond	ORmarg	ORcond	ORmarg
None	0.001	0.000	/	0.001
C1,C2	0.002	0.035	/	0.034
Z	-0.004	0.005	/	0.004
Y	0.000	0.041	/	0.042
C1,C2,Z	0.001	0.040	/	0.036
C1,C2,Y	-0.001	0.130	/	0.134
Z,Y	-0.838	-0.626	/	-0.624
C1,C2,Z,Y	-0.769	-0.579	/	-0.600

Results (2)

Variables associated with missingness	Multivariable regression		IPTW	
	RR	Risk difference	RR	Risk difference
None	0.000	0.000	0.000	0.000
C1,C2	0.070	-0.009	0.069	-0.010
Z	0.016	-0.003	0.015	-0.003
Y	0.125	-0.036	0.127	-0.036
C1,C2,Z	0.076	-0.010	0.073	-0.010
C1,C2,Y	0.227	-0.043	0.230	-0.043
Z,Y	-0.428	-0.130	-0.426	-0.130
C1,C2,Z,Y	-0.390	-0.122	-0.409	-0.124

CC: a bad idea

CC not suitable for the estimation of marginal effects (both with PS and logistic regression) unless:

- MCAR mechanism
- missingness not associated with both Y and Z AND under H0!!

In the literature CC seems to be the **most common approach** for PS analysis...

What else can be done?

The missingness pattern approach

Helen Blake's PhD research

The missingness pattern approach

Proposed by Rosenbaum and Rubin (1984) and D'Agostino and Rubin (2000)

Definition of a **generalized PS** estimated within each pattern of missingness

		X2	
		Observed	Missing
X3	Observed	$\hat{e}(X_1,X_2,X_3)$	$\hat{e}(X_1,X_3)$
	Missing	$\hat{e}(X_1,X_2)$	$\hat{e}(X_1)$

Relies on an **additional assumption**: an extension of SITA

The SITA extension assumption

Let $\mathbf{X}$ the vector of baseline confounders be split in $\mathbf{X} = \{\mathbf{X_{obs}}, \mathbf{X_{mis}}\}$ and $\mathbf{R}$ the vector of the missingness indicators for the confounders.

"Classical" SITA assumption: the potential outcomes and the treatment assignment are independent given the measured characteristics (no unmeasured confounders):

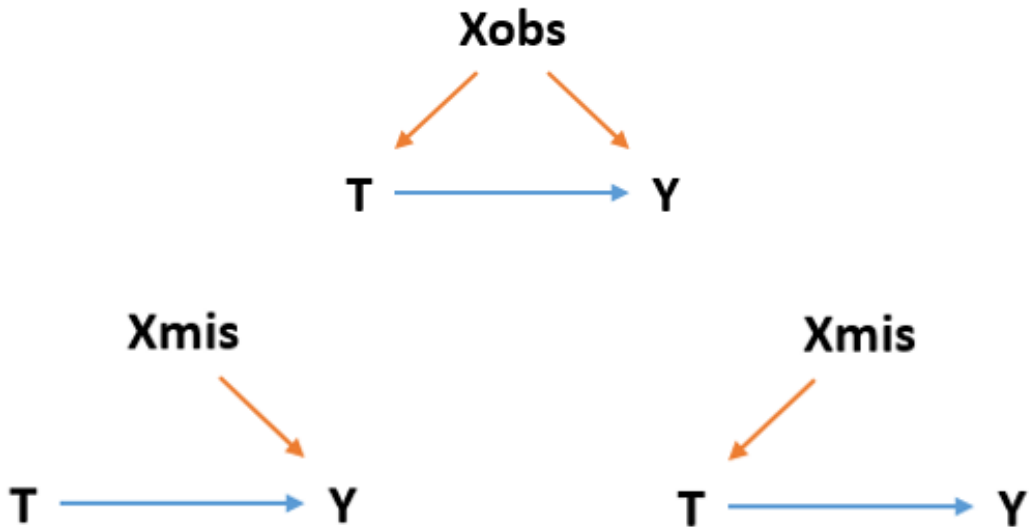
$$(Y^0, Y^1) \perp T|\mathbf{X}$$

SITA extension (Mattei):

$$(Y^0, Y^1) \perp T|\mathbf{X}, \mathbf{R}$$
and either
$$\mathbf{X_{mis}} \perp T|\mathbf{X_{obs}}, \mathbf{R} \quad \text{or} \quad \mathbf{X_{mis}} \perp (Y^0, Y^1)|\mathbf{X_{obs}}, \mathbf{R}$$

The assumption in practice

For the assumption to hold, **X** can be a confounder when observed but not when missing



Assumption required because the generalised PS **balances the observed part** of the covariates only (but not the missing part)

Link with Rubin's taxonomy

It's been shown that:

- if the SITA assumption extension does not hold: invalid inferences even under MCAR
- if the SITA assumption extension holds: valid inferences even under some MNAR mechanisms

Promising approach that requires further investigation to be applicable in a variety of situations combining MAR and MNAR data

Summary

The missingness pattern approach:

- can lead to **valid inferences** if the SITA assumption extension holds
- could be of interest for some **MNAR mechanisms**
- is quite straightforward

However:

- requires a **large sample size**
- difficulties arise with a **lot of patterns**
  $\Longrightarrow$ Pooling?
- has a **specific applicability** in its present form
  $\Longrightarrow$ Combining MPA with other methods?

Multiple imputation

Multiple imputation for PS

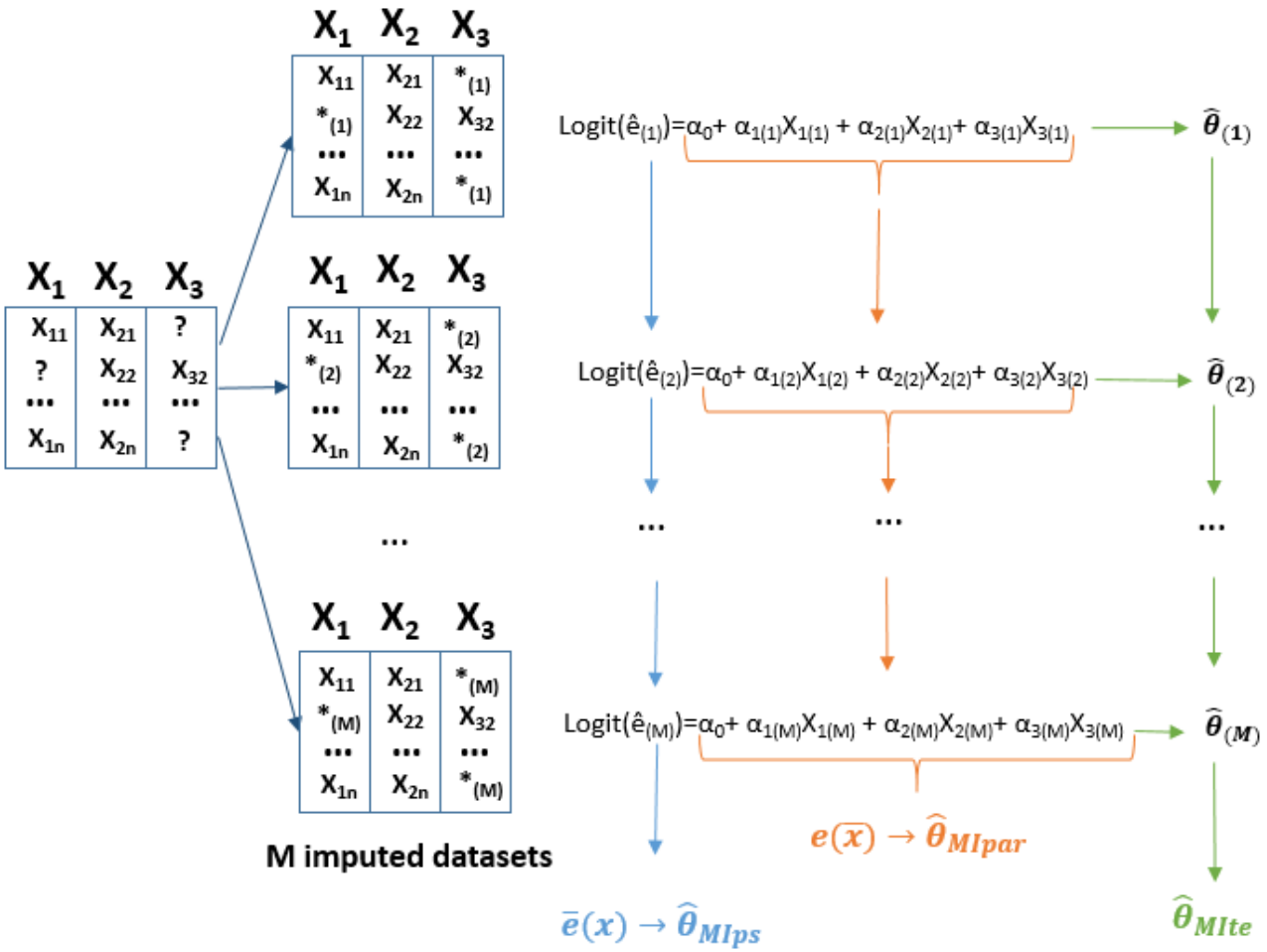
Aim: create M complete datasets to estimate the PS for each participant and apply Rubin's rules to obtain a treatment effect estimate

Two key questions:

- Should the outcome be included in the imputation model ?
  $\Longrightarrow$ PS paradigm $\neq$ Missing data paradigm

- How to apply Rubin's rules?
  $\Longrightarrow$ pooled treatment effect or pooled PS?

What should we combine?



$\hat{\theta}$: treatment effect estimate

In the literature...

Existing studies:

- Mitra & Reiter[1]: for PS matching, MIps>MIte but opposite conclusion for IPTW
  $\Longrightarrow$ Outcome not included in the imputation model

- Hill[2]: MIte>MIps and outcome in the imputation model
  $\Longrightarrow$ PS matching only

Simulation study but **no theoretical arguments** about the validity of these estimators when data are MAR

[1] Mitra R, Reiter JP. A comparison of two methods of estimating propensity scores after multiple imputation. SMMR. 2016 Feb;25(1):188-204.

[2] Hill J. Reducing Bias in Treatment Effect Estimation in Observational Studies Suffering from Missing Data; 2004. ISERP working paper 04-01.

Balancing properties

Are the 3 estimated PS **balancing scores**?
$\Longrightarrow$ requirement for valid inferences

For MIte, we showed that **within each imputed dataset**:

$$\mathbf{X}_{obs} \perp Z \mid e(\mathbf{X}_{obs}, \mathbf{X}_m^{(k)})$$
$$\mathbf{X}_m^{(k)} \perp Z \mid e(\mathbf{X}_{obs}, \mathbf{X}_m^{(k)}).$$

For MIps and MIpar:

- the pooled PS is not a function of the covariates
- the true PS is not a function of the estimated PS
  $\Longrightarrow$ the pooled PS **is not a balancing score**

Consistency

Consistency comes from the ability of the PS to balance groups: **MIps and MIpar are not consistent** estimators

MIte: Seaman and White: the consistent estimator for an infinite number of imputations

In practice: how well these 3 estimators perform?

Summary

Different ways to apply Rubin's rules after MI of the partially observed covariates for IPTW

$\Longrightarrow$ MIte only is a **consistent estimator** of the treatment effect (MAR mechanism)

Simulation results found in the literature are not clear so **need to empirically assess** these methods:

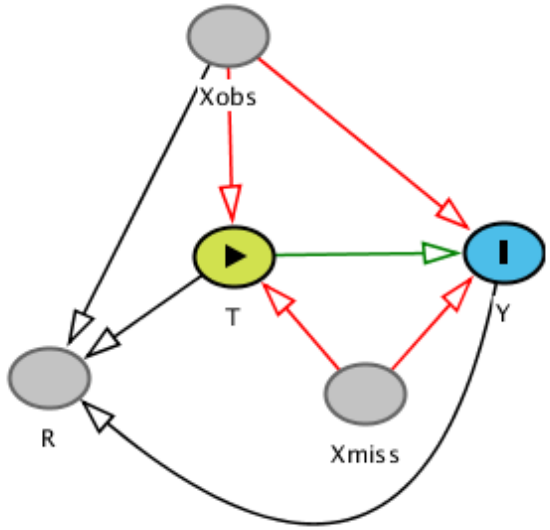
- variance estimation?
- outcome in the imputation model?
- strength of the bias for MIps and MIpar

Simulation study

Simulation plan

Observational study:

- estimation of the effect of a binary treatment T on a binary outcome Y (RR), n=5000
- 3 confounders (2 with 30% of data missing)



Y: binary outcome
T: treatment
R: missingness indicator
Xobs: observed confounders
Xmiss: missing confounders

Multiple imputation:

- Chained equations (FCS)
- M=10
- Imputation model: X_1, X_2, X_3, T, Y

Analysis strategies

IPTW estimator:

- Estimation of the weighted marginal proportions $\hat{P}_0$ and $\hat{P}_1$ and $RR = \frac{\hat{P}_1}{\hat{P}_0}$
- Use of Williamson *et al.*[1] variance estimator for IPTW (two-step estimator)

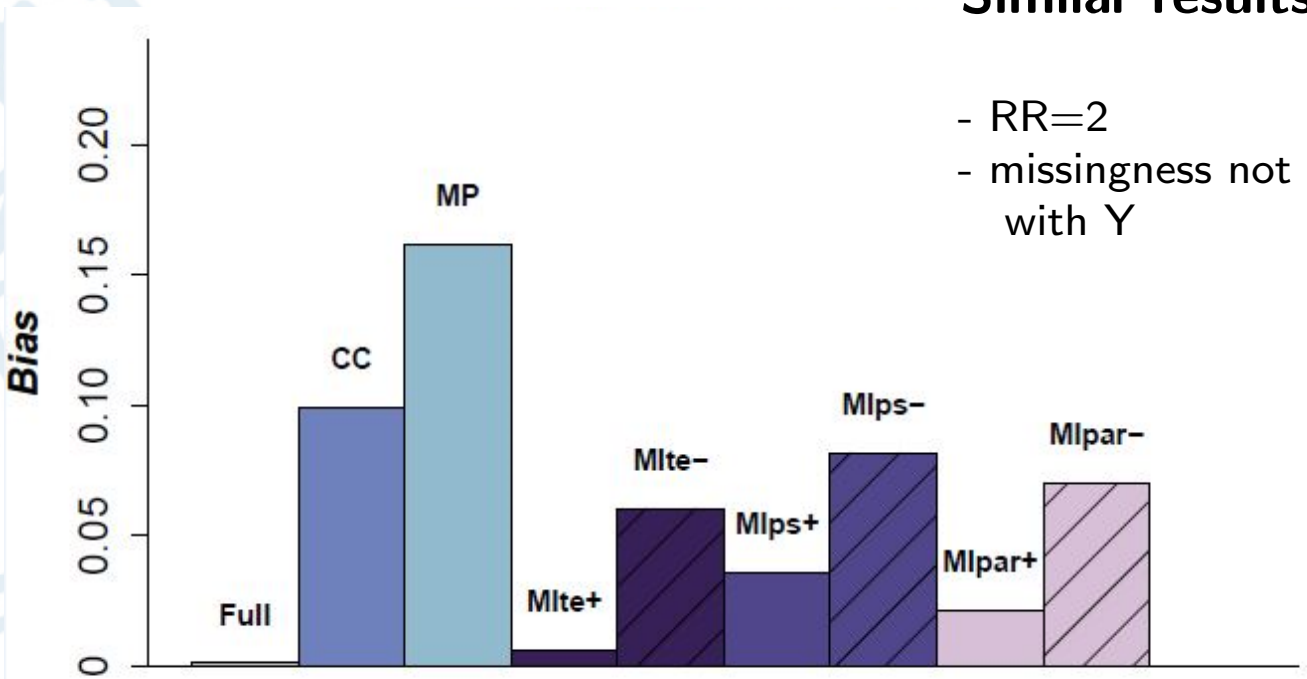
Compared approaches:

- Complete case: exclusion of participants with partial data
- Missingness pattern: 4 different PS models
- MIte: the M IPTW estimates of the treatment effect are pooled according Rubin's rules
- MIps: 1 IPTW estimate obtained from the average PS
- MIpar: 1 IPTW estimate obtained from the PS of the average covariates

[1] Williamson EJ, Forbes A, White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. Stat Med. 2014 Feb;33(5):721-737.

Results: bias

RR=1, outcome predictor of missingness



Similar results with:

- RR=2
- missingness not associated with Y

- The **outcome must be included** in the imputation model
- **Pooling the treatment effects** (MIte) performs best

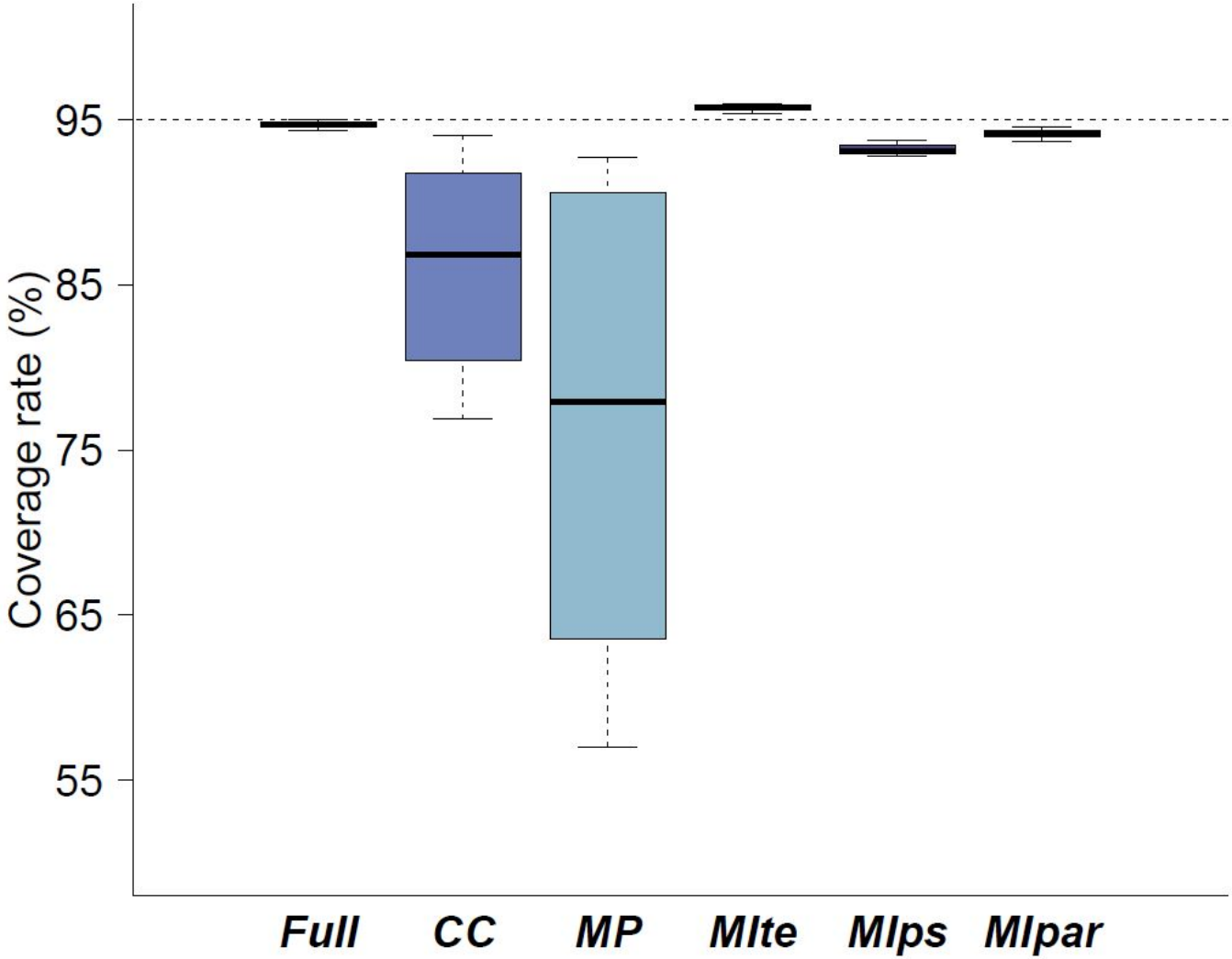
Balancing properties

Standardized differences (in%) between groups: $SD = \frac{100 \times \left|\bar{X}_1 - \bar{X}_0\right|}{\sqrt{\frac{s_0^2 + s_1^2}{2}}}$

Method	X_1 (partially observed)	X_2 (fully observed)	X_3 (partially observed)
Crude (without IPTW)	81.3	74.7	51.7
Full data (IPTW)	4.6	4.6	2.4
MIte	4.5	4.5	2.4
MIps (full dataset)	15.9	5.5	10.7
MIps (observed part)	7.6	5.5	4.9
MIpar (full dataset)	14.7	4.8	9.7
MIpar (observed part)	7.7	4.8	5.4

PS obtained from MP, MIps and MIpar **do not balance the missing part** of the covariates

Coverage rate



Real life example

Example

Data: THIN database (records from GP in the UK)

Population: focus on patients with a pneumonia episode, n=9073 (Douglas *et al.*)

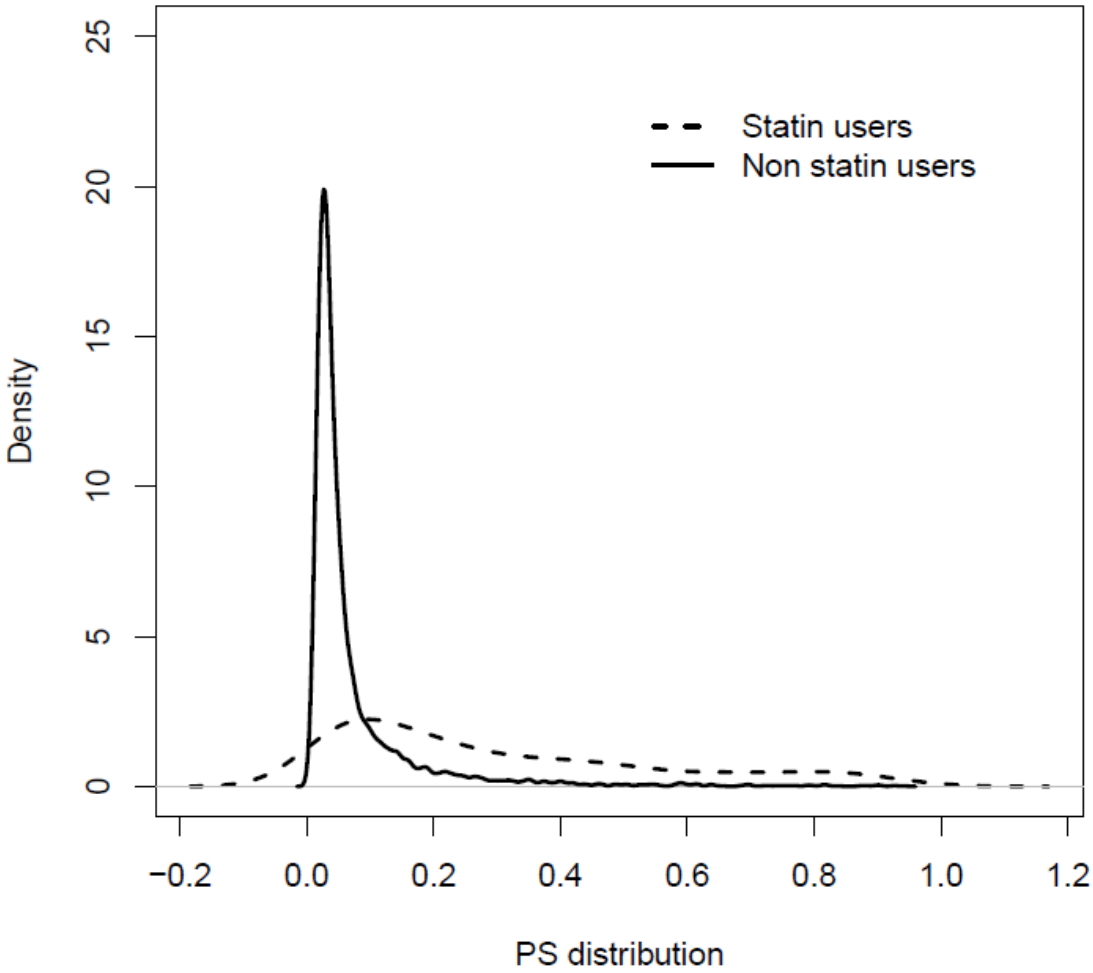
Intervention: statins vs no statins

Outcome: death within 6 months

Confounders: 21 variables (demographic, medical history, treatments)

Missing data: body mass index (19.2%), smoking status (6.2%) and alcohol consumption (18.5%)

Example: PS distribution (CC)



Example: balance

Variable	Missing (%)	Statin users n=599	Missing (%)	Non statin users n=6559	Standardized difference (%) Crude	CC*	MP	MIte	MIps	MIpar
Characteristics										
Age [mean (sd)]		66.9 (10.7)		69.8 (10.9)	27.0	3.8	2.0	1.4	1.4	1.4
Male		322 (53.8)		3173 (48.4)	10.8	2.0	6.2	2.2	2.1	2.2
BMI [mean (sd)]	43 (7.2)	27.6 (5.9)	1444 (22.0)	25.8 (5.9)	31.9	7.8	9.0	9.0	11.4	11.4
Drinkers	67 (11.2)	98 (18.4)	1334 (20.3)	814 (15.6)	7.6	2.1	0.3	2.3	2.9	3.0
Smokers	7 (1.2)	256 (43.2)	505 (7.7)	2728 (45.1)	3.7	1.7	1.5	2.8	3.0	3.0
Medical history										
Diabetes		243 (40.6)		715 (10.9)	72.1	5.0	7.7	7.1	7.2	7.1
Cardiovascular disease		141 (23.5)		651 (9.9)	37.1	11.4	11.4	13.6	13.6	13.6
Circulatory disease		426 (71.1)		3471 (52.9)	38.2	13.6	9.8	16.6	16.7	16.6
Heart failure		51 (8.5)		426 (6.5)	7.7	11.6	6.2	12.8	12.8	12.8
Cancer		37 (6.2)		607 (9.2)	11.5	2.1	0.4	0.4	0.0	0.1
Dementia		6 (1.0)		190 (2.9)	13.7	7.3	13.0	11.6	11.6	11.6
Hypertension		336 (56.1)		1165 (17.8)	52.1	13.3	21.5	18.7	18.7	18.7
Hyperlipidemia		205 (34.2)		182 (2.8)	88.5	1.1	4.1	1.9	2.0	2.0
Treatments										
Antidepressant		108 (18.0)		995 (15.2)	7.7	1.7	5.9	0.3	0.1	0.1
Antipsychotic		11 (1.8)		340 (5.2)	18.3	0.5	11.3	5.0	5.0	5.0
Hormone replacement therapy		37 (6.2)		277 (4.2)	8.8	0.9	0.0	1.0	1.0	1.0
Steroid		93 (15.5)		1090 (16.6)	3.0	1.0	2.2	0.4	0.3	0.3
Antihypertensive		272 (45.4)		1165 (17.8)	62.3	12.6	27.5	18.0	17.8	17.9
Diuretics		319 (53.3)		2416 (36.8)	33.4	14.3	19.8	15.8	15.9	15.9
Betablocker		193 (32.2)		1061 (16.2)	38.1	11.4	7.2	13.8	13.8	13.8
Nitrate		74 (12.4)		334 (5.1)	25.9	17.3	14.8	17.5	17.6	17.6

For CC analysis, n=5168 (503 statin users and 4665 non users).

CC: complete case; MP: missingness pattern; MIte: treatment effects combined after multiple imputation; MIps: propensity scores combined after multiple imputation; MIpar: propensity score parameters combined after multiple imputation

Example: results

Method	$\widehat{RR}$	95% CI($\widehat{RR}$)
Crude	0.587	[0.497;0.684]
CC	0.702	[0.534;0.924]
MP	0.708	[0.555;0.904]
MIte	0.654	[0.513;0.835]
MIps	0.653	[0.512;0.834]
MIpar	0.654	[0.513;0.834]

CC: complete case; MP: missingness pattern; MIte: treatment effects combined after multiple imputation; MIps: propensity scores combined after multiple imputation; MIpar: propensity score parameters combined after multiple imputation; RR: relative risk

The 3 partially observed covariates are **not strong confounders**

MP: Need to **pool some patterns** because of small sample and SITA assumption extension unlikely to be valid

Similar results for MI when increasing artificially the missingness rate

Recommendations

Complete case analysis: **bad idea**, unless MCAR mechanism

Multiple imputation:

- good statistical properties under a MAR mechanism
- the treatment effects should be pooled rather than the PSs
- the outcome must be included in the imputation model

The missingness pattern approach:

- good statistical properties if missing values are not confounders
- promising technique for MNAR mechanisms

Future work

Multiple imputation:

- to study the issue of **compatibility** between the substantive, PS and imputation models
- to study how to assess **covariate balance** after MI

The missingness pattern approach:

- to **combine MPA with MI** when both MAR and MNAR mechanisms
- to study how to **pool patterns** when small sample size
- to develop a **variance estimator**

Thank you!