Compensating for Non-response in Biosocial Research:

Simulation Study from a Cross-sectional Analysis

Dr Tina Hannemann

Biomarkers in Longitudinal Studies in the UK

- Understanding Society
- Millennium Cohort Study
- British Cohort Study
- Child Development Study
- English Longitudinal Study of Ageing

English Longitudinal Study of Ageing (ELSA)

- Data from 11,391 core sample members of men and women over 50 years living in private households
- Multistage stratified probability sample from the Health Survey of England (HSE) with refreshment samples
- Further health examinations including blood samples (W2, W4 & W6) and hair samples (W6)

Missing Information in ELSA

- Potential non-response at three different stages:
  - Main interview (CAPI)
  - Health examination (Nurse visit)
  - Blood/Hair sample collection

Patterns of Missing information

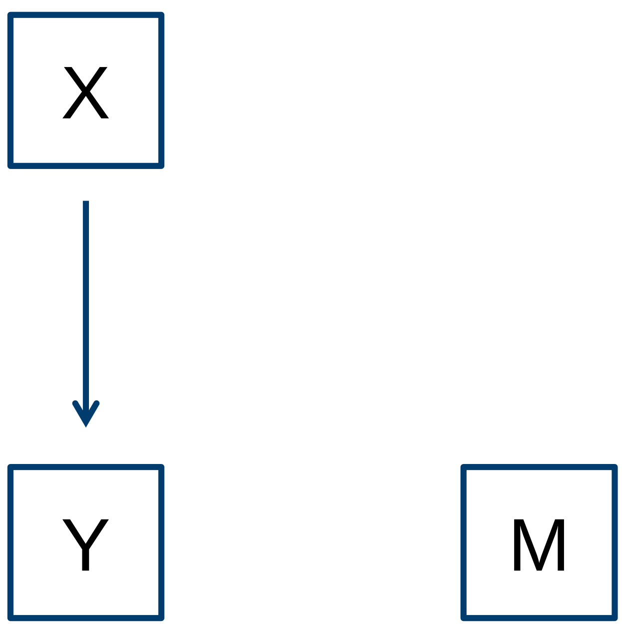
- Missing completely at random (MCAR)

- Missing at random (MAR)

- Missing not at random (MNAR)

Patterns of Missing information

- Missing completely at random

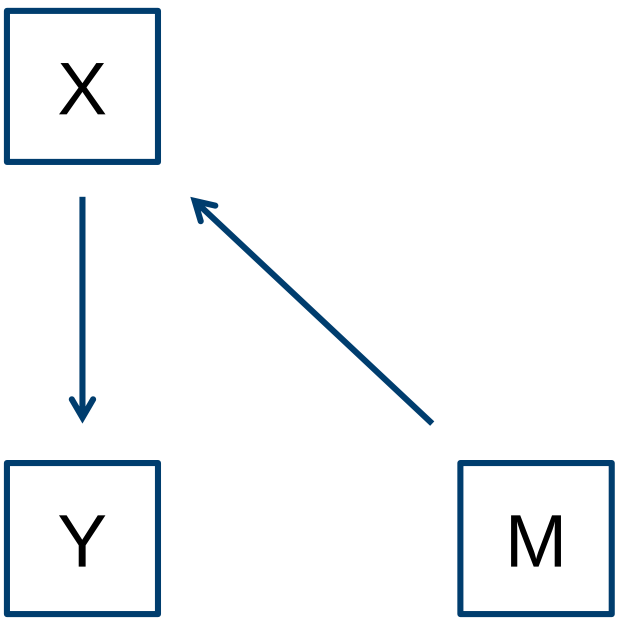


X explanatory variable
Y depend variable

M missingness of Y

Patterns of Missing information

- Missing at random

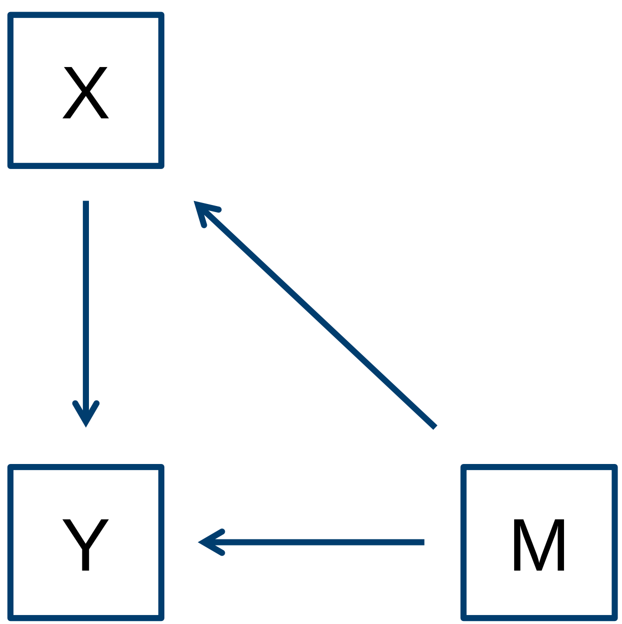


X explanatory variable
Y depend variable

M missingness of Y

Patterns of Missing information

- Missing not at random



X explanatory variable
Y depend variable

M missingness of Y

Accounting for missing data bias

- Several methods for missing data compensation
- Confusion over effectiveness of methods
- Problem of missing data is ignored (assumption of MAR)

Objective

- Test and evaluate several analytical approaches
- Simulation study based on data from ELSA (wave 2)
- Comparison of five different approaches
- Evaluation over all three missingness patterns

Strategy

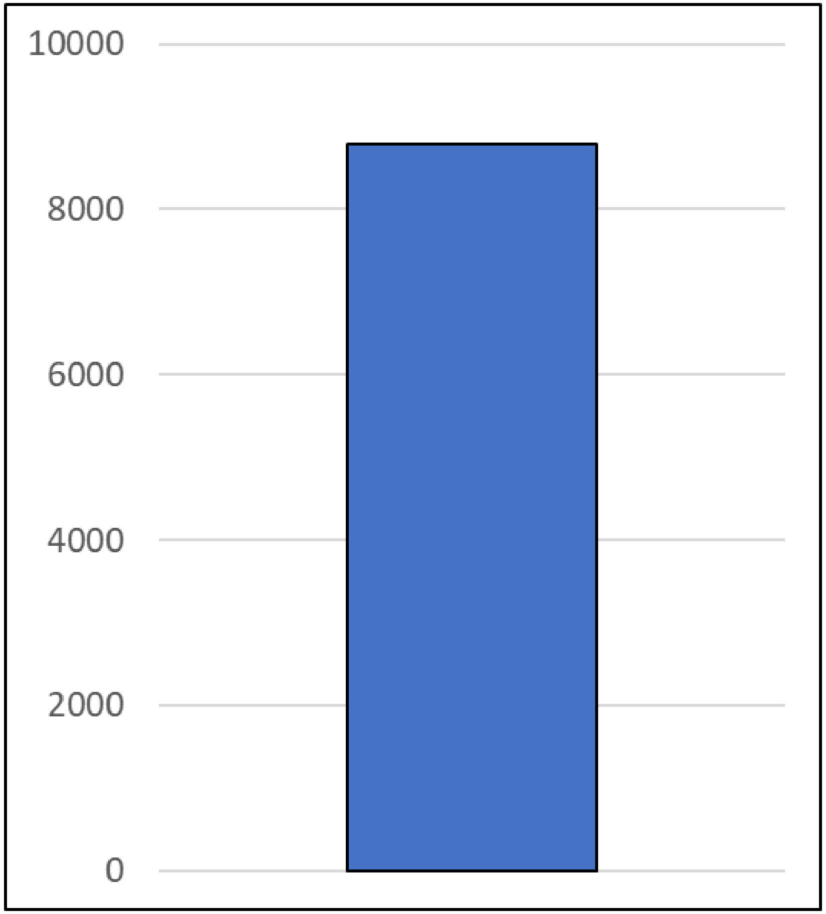
Step 1 Choose substantive model	Step 2 Select data in ELSA, expand sample size	Step 3 Run baseline model on “true” values
Step 4 Create missing data for MCAR, MAR and MNAR	Step 5 Test five different compensation approaches	Step 6 Compare with “true” model and evaluate effectiveness

Step 1 – Substantive model

Impact of socio-economic status on biomarker level

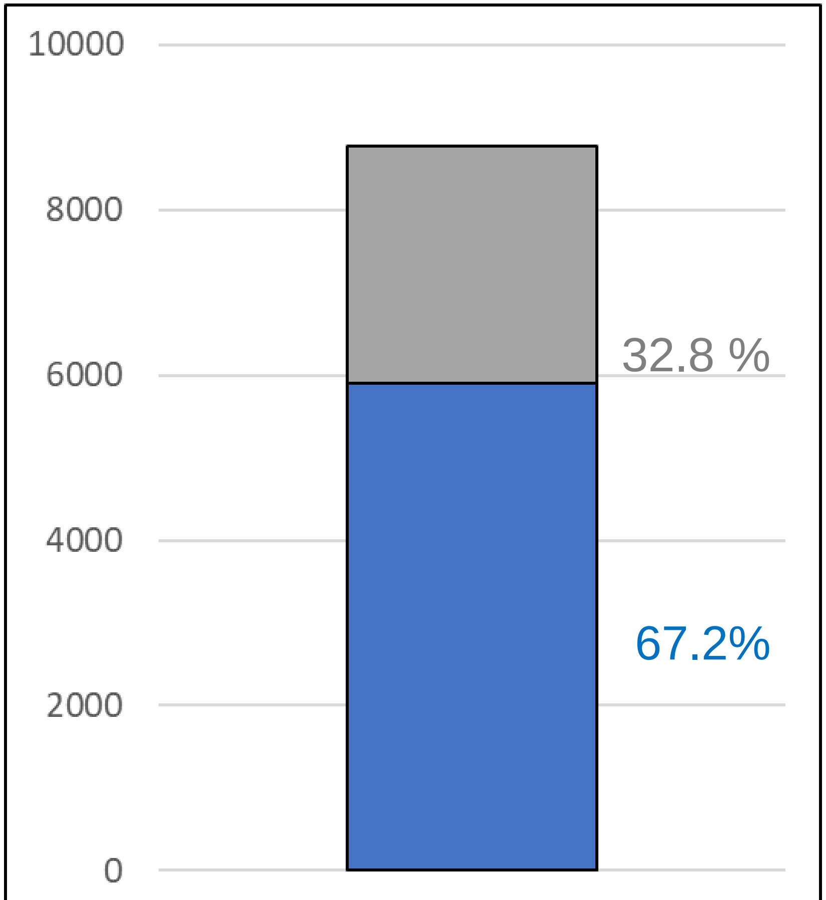
- SES
  - Education, occupational class and wealth quintile
- Biomarker
  - C-reactive protein (CRP)
- Confounding variables
  - Age, sex, physical activity, BMI, Longstanding illness, pain, general health, BP, in paid work

Step 2 – ELSA missing information



8.780
Wave 2 core member

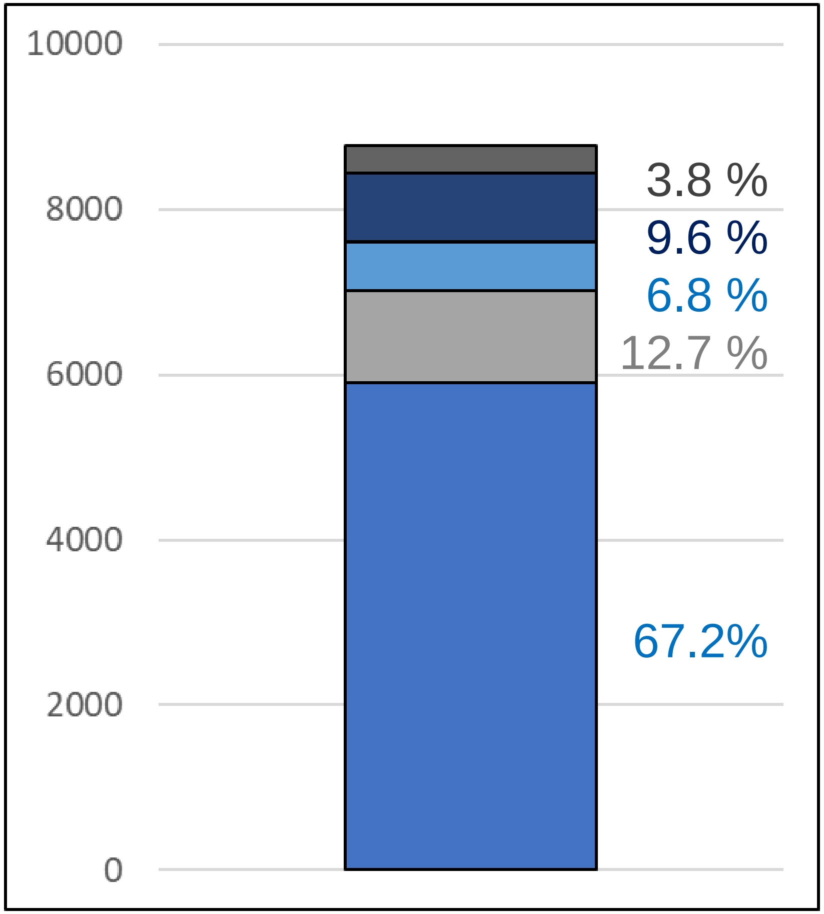
Step 2 – ELSA missing information



2.881
Non-response and missing

5.899 Eligible blood samples

Step 2 – ELSA missing information



332 Unusable blood sample

839 No blood sample

596 Refuse blood sample

1.114 Refuse of nurse visit

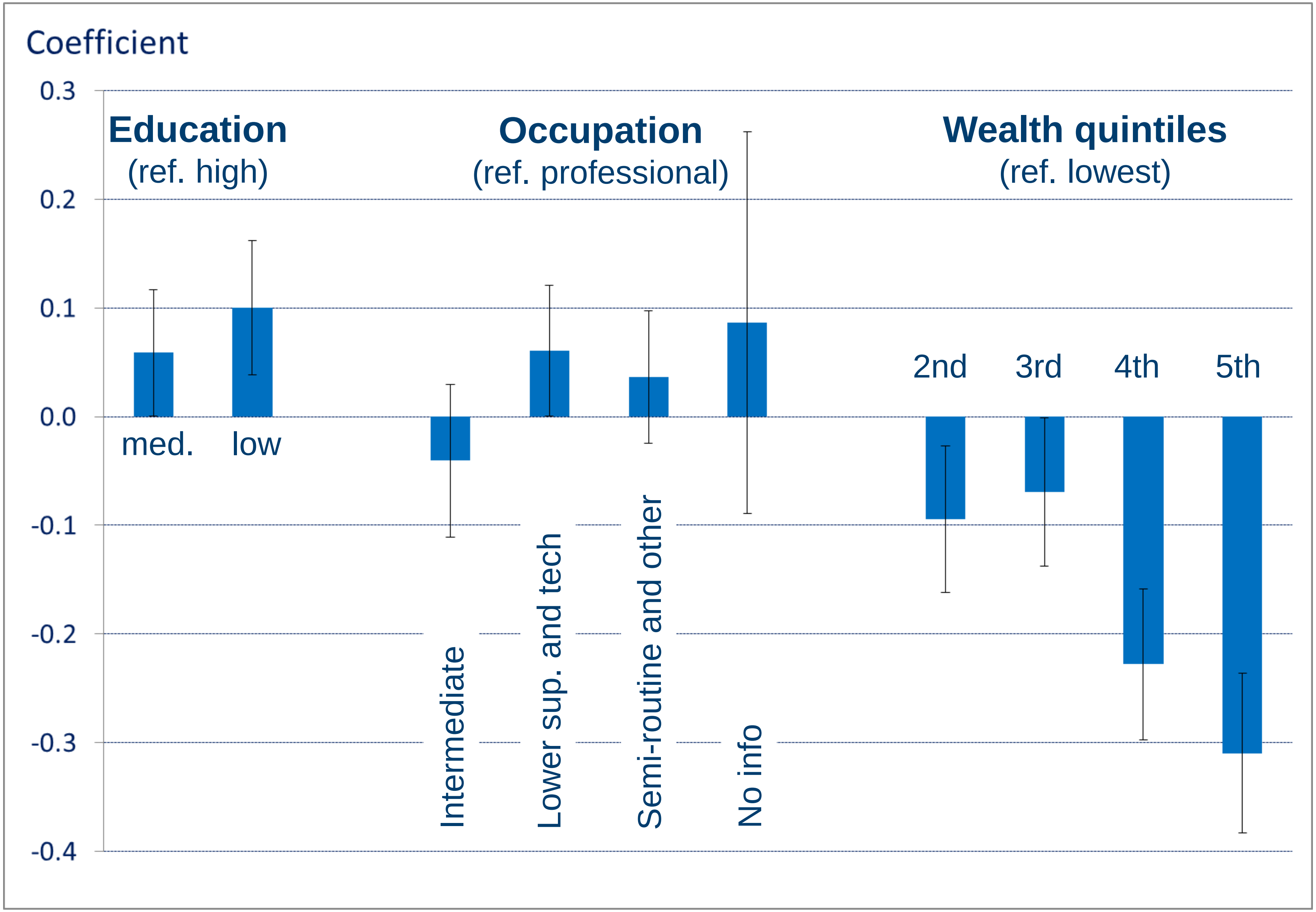
5.899 Eligible blood samples

Step 3 – Model on expanded full dataset

- Logistic regression on logged CRP values (10.000 sample)

- Coefficients for all three SES variables as baseline for simulation study

Baseline Model



Step 4 – Creating missing data patterns

- MCAR: Missingness is completely random
- MAR: Missingness dependent on age, sex, wealth and health
- MNAR: Missingness dependent also on level of CRP
- 100 simulation datasets for each missing patterns

Step 5 – Five compensation methods

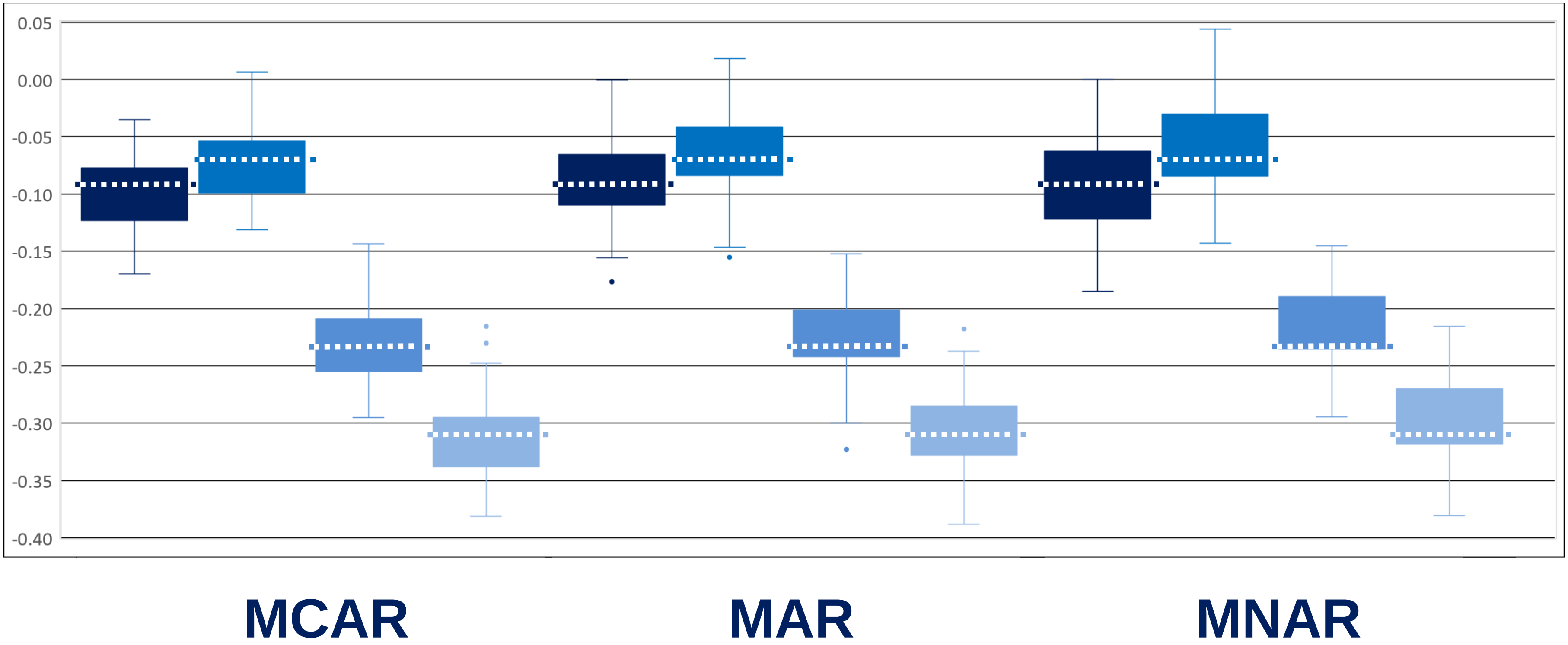
- Complete Case (CC)
- Inverse Propensity Weighting (IPW)
- Selection model using Mills Ratio (MR)
- Multiple Imputation (MI)
- Multiple Imputation with Mills Ratio (MI+MR)

Step 6 – Run and evaluate simulations

	CC	IPW	MR	MI	MI+MR
MCAR					
MAR					
MNAR					

Expectation

Complete Case – Wealth Quintiles



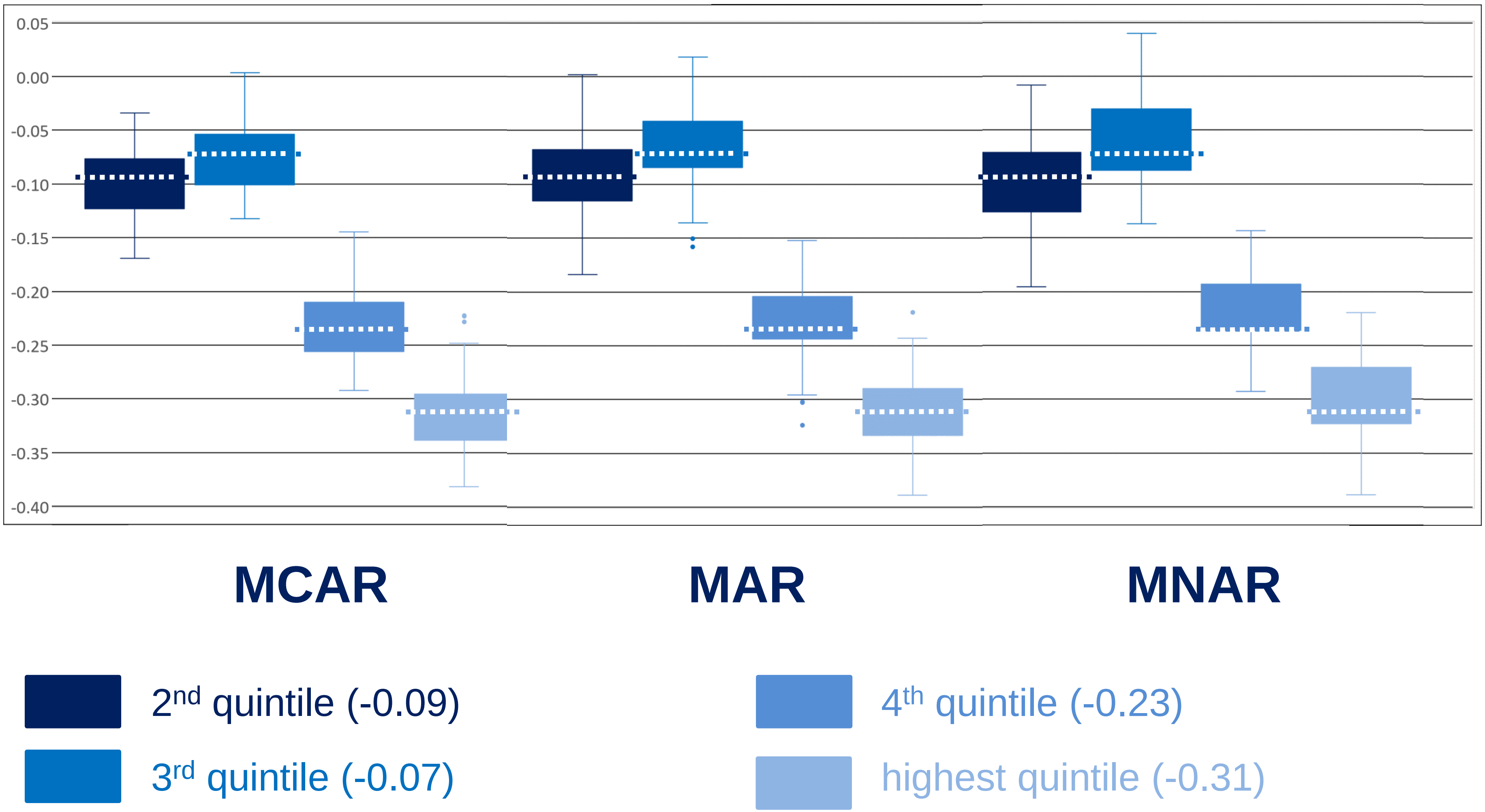
2nd quintile (-0.09)

3rd quintile (-0.07)

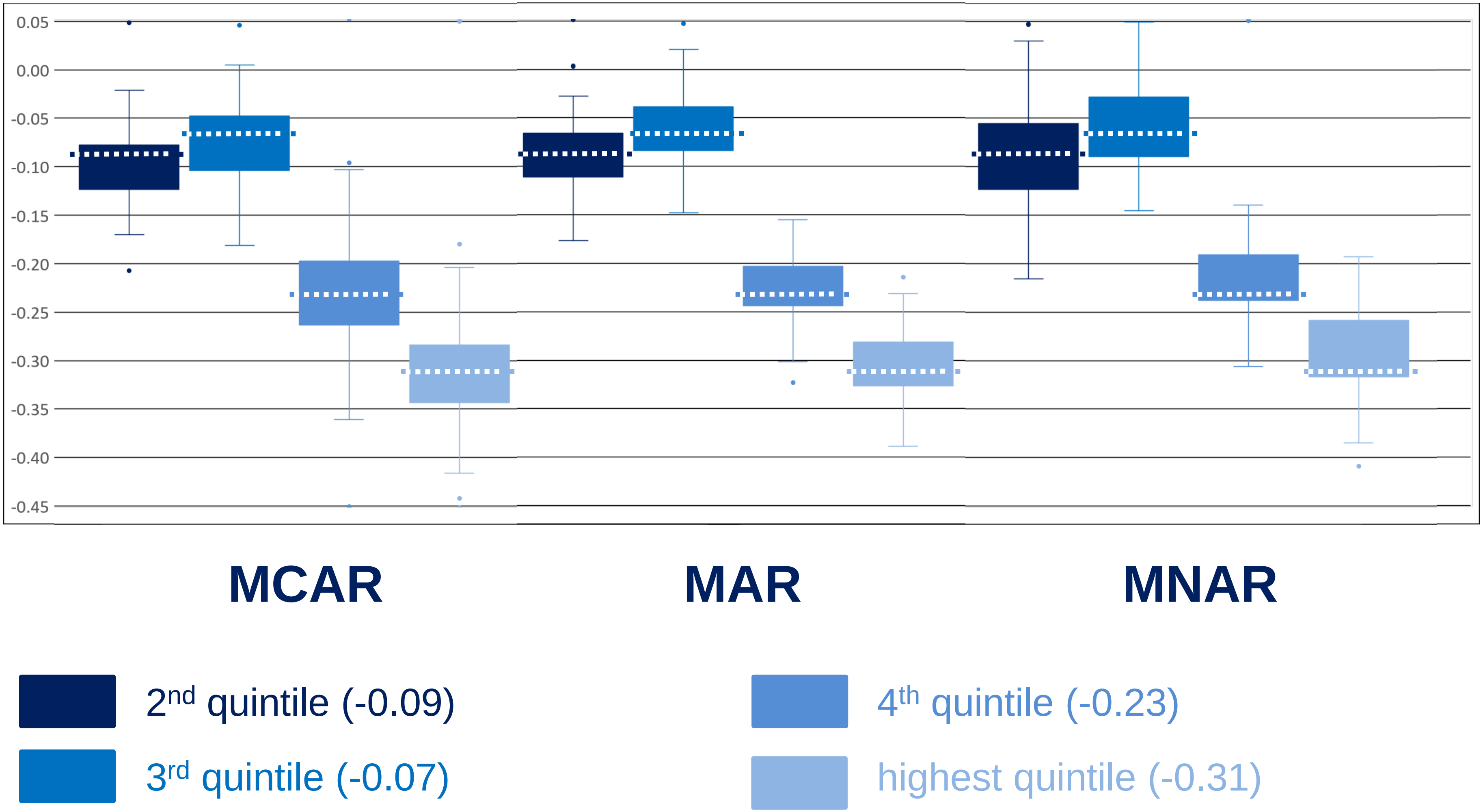
4th quintile (-0.23)

highest quintile (-0.31)

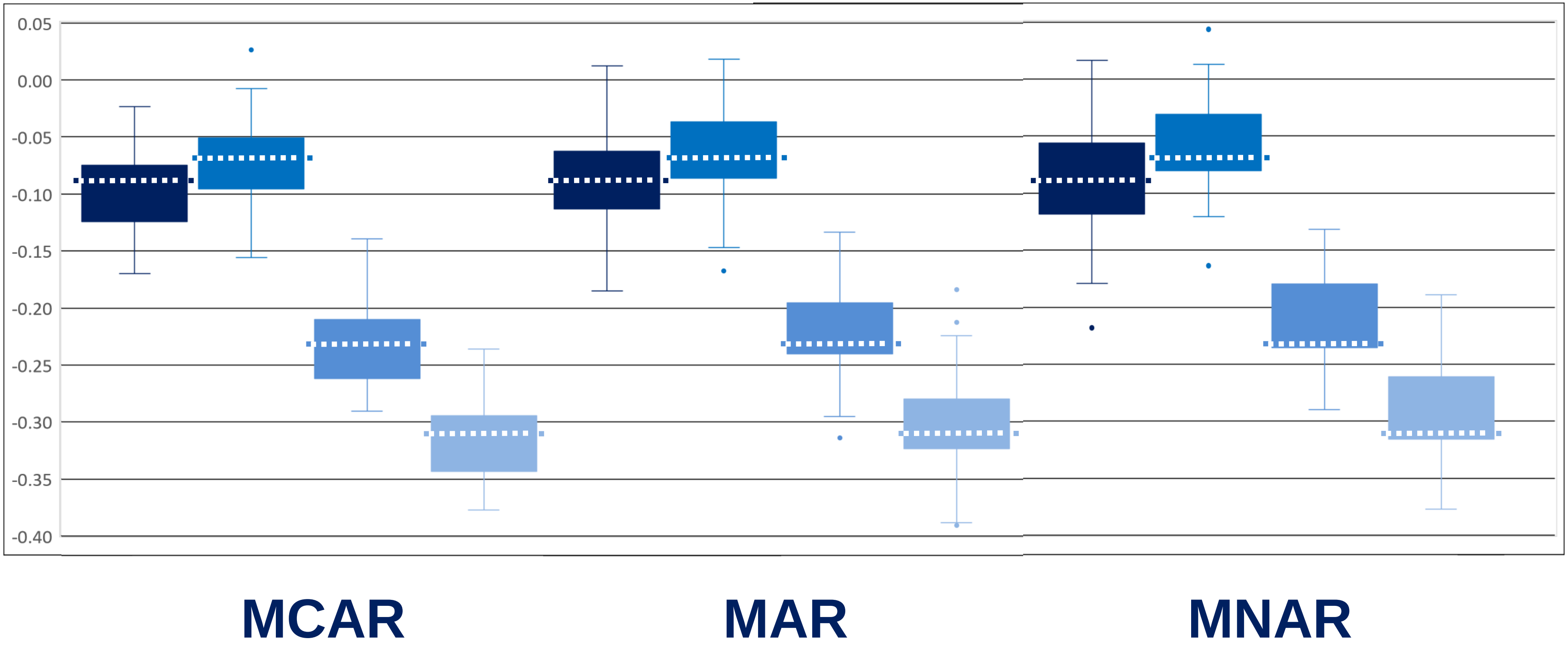
IPW – Wealth Quintiles



MR – Wealth Quintiles



MI – Wealth Quintiles



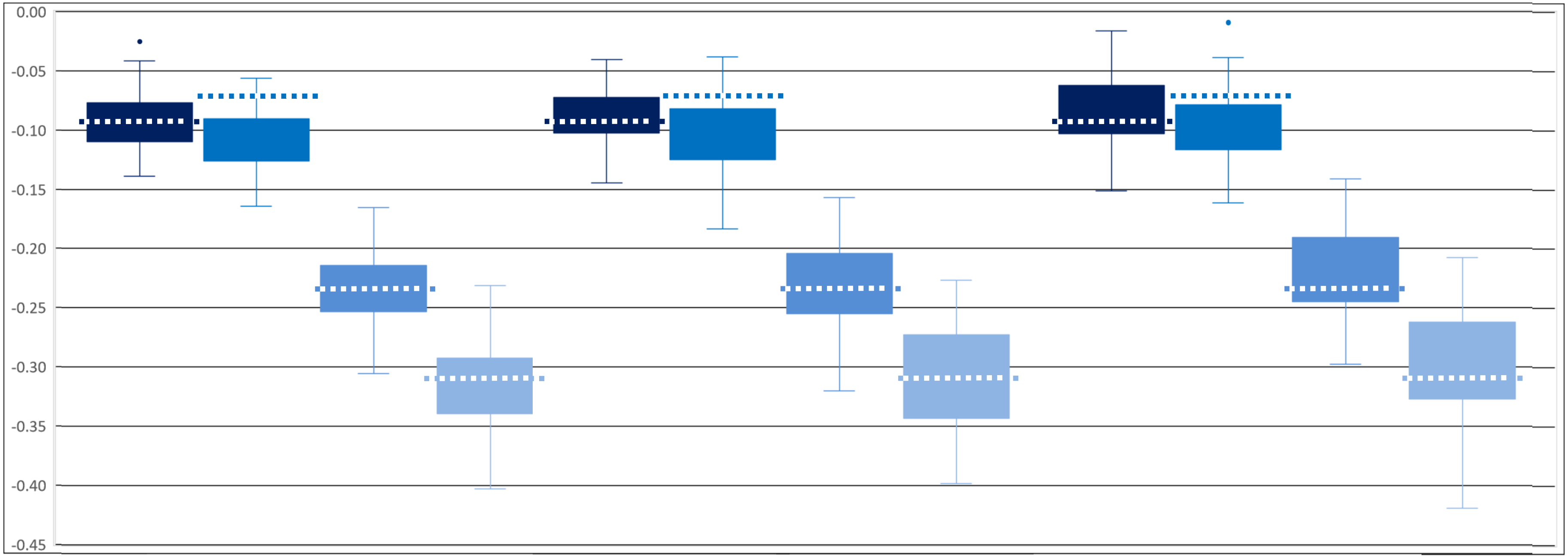
2nd quintile (-0.09)

3rd quintile (-0.07)

4th quintile (-0.23)

highest quintile (-0.31)

MI+MR – Wealth Quintiles



MCAR **MAR** **MNAR**

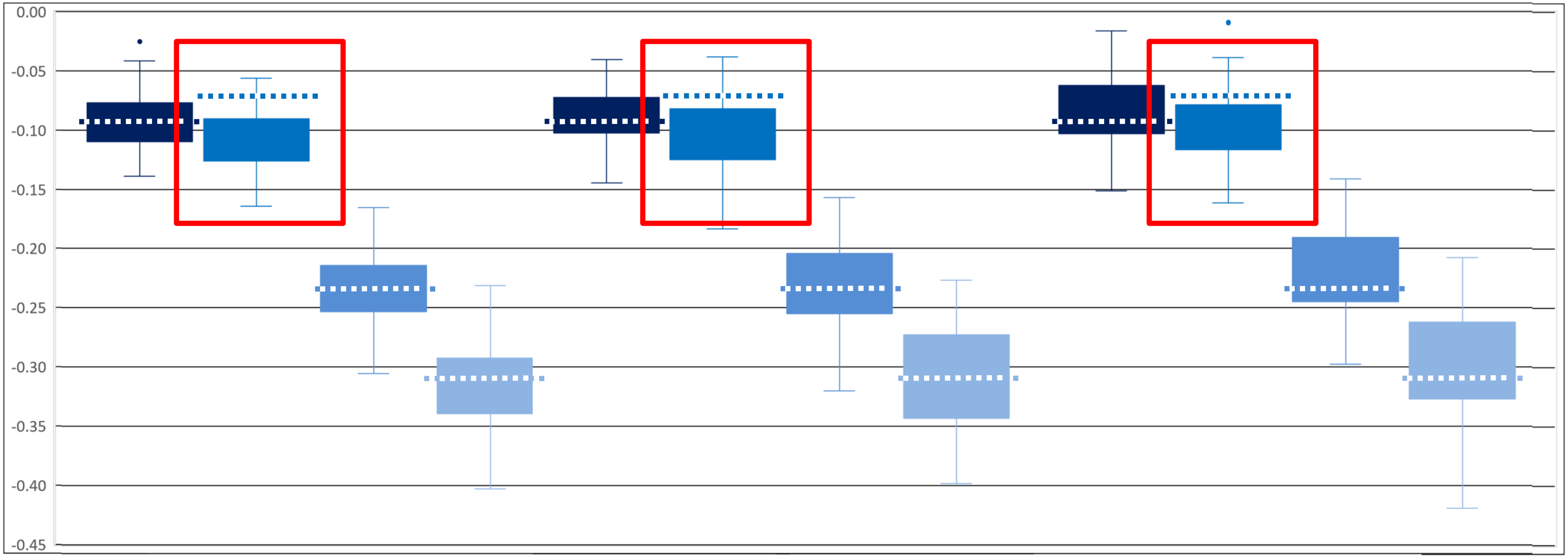
2nd quintile (-0.09)

3rd quintile (-0.07)

4th quintile (-0.23)

highest quintile (-0.31)

MI+MR – Wealth Quintiles



MCAR **MAR** **MNAR**

2nd quintile (-0.09)

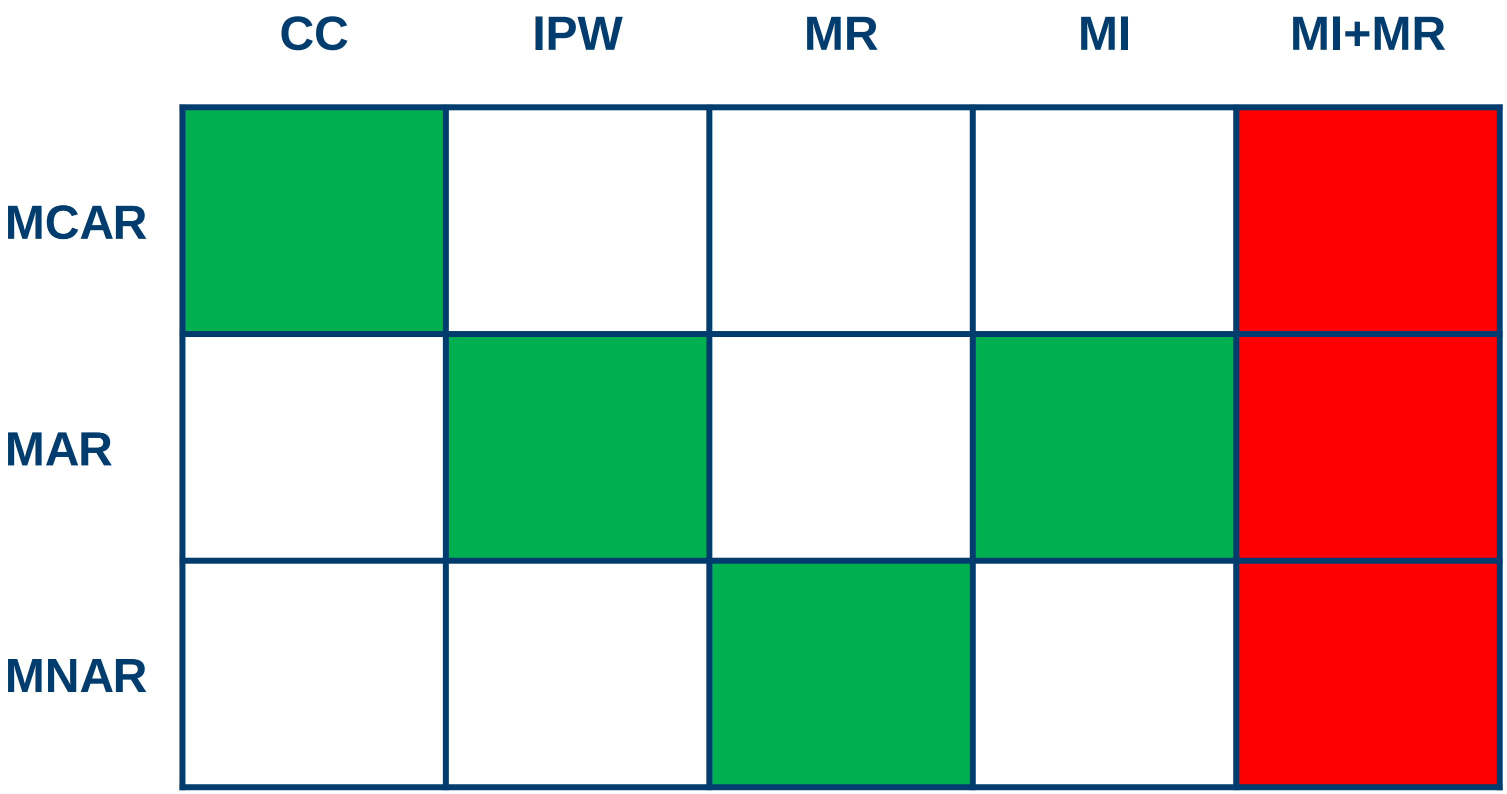
3rd quintile (-0.07)

4th quintile (-0.23)

highest quintile (-0.31)

Expectation

Observation



Conclusions

- No clear patterns
- Multiple Imputation and Mills Ratio method has no advantage
- MCAR and MAR rather robust
- MNAR bias not compensated with these methods

Future Research

- Expand the simulation study to larger iteration number
- Simplify substantial model
- Test and evaluate effectiveness of methods depending on the proportion of missingness in dataset
- Longitudinal analysis including wave 4 and wave 6 using growth curve modelling

Thank you for your attention

Tina.Hannemann@manchester.ac.uk