



Evaluation of multiple imputation for handling missing covariate information in the context of a case-cohort study

Melissa Middleton

Katherine J. Lee, Cattram Nguyen, Margarita Moreno-Betancur, John Carlin

Australian and New Zealand Statistical Conference

5th July 2021



Overview

- Motivating Example
- Objective
- Simulation Study
- Conclusions

Motivating Example: The Barwon Infant Study


- Population-derived birth cohort study with antenatal recruitment¹

Allergy

ORIGINAL ARTICLE

EPIDEMIOLOGY AND GENETICS

Vitamin D insufficiency in the first 6 months of infancy and challenge-proven IgE-mediated food allergy at 1 year of age: a case-cohort study

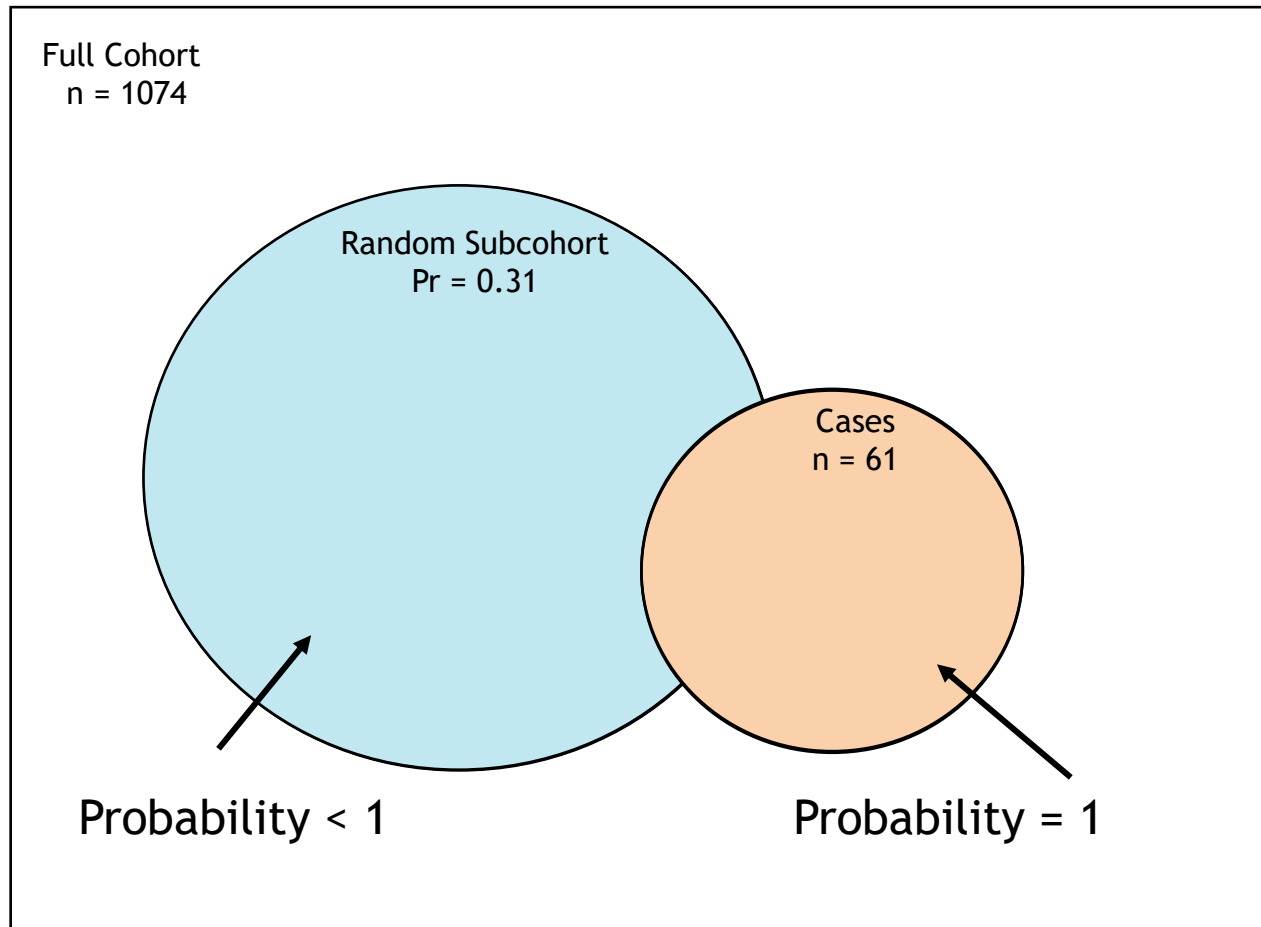
J. Molloy^{1,2,3,4} , J. J. Koplin^{3,4,5}, K. J. Allen^{3,4,6,7}, M. L. K. Tang^{3,4,6,7}, F. Collier^{1,2,3}, J. B. Carlin^{3,5,6}, R. Saffery³, D. Burgner^{3,6,8}, S. Ranganathan^{3,6,9}, T. Dwyer¹⁰, A. C. Ward¹, M. Moreno-Betancur^{3,11}, M. Clarke¹², A. L. Ponsonby^{3,4,†} & P. Vuillermin^{1,2,3,4,†} on behalf of the BIS Investigator Group

¹School of Medicine, Deakin University, Waurn Ponds; ²Child Health Research Unit, Barwon Health, Geelong; ³Murdoch Childrens Research Institute, Parkville; ⁴Centre for Food and Allergy Research, Parkville; ⁵Centre for Epidemiology and Biostatistics, The University of Melbourne, Carlton; ⁶Department of Paediatrics, University of Melbourne; ⁷Department of Allergy and Immunology, Royal Children's Hospital, Parkville; ⁸Department of Paediatrics, Monash University, Clayton; ⁹Department of Respiratory Medicine, Royal Children's Hospital, Parkville, VIC, Australia; ¹⁰The George Institute for Global Health, University of Oxford, Oxford, UK; ¹¹Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC; ¹²Biological and Molecular Mass Spectrometry Facility, Centre for Microscopy, Characterisation and Analysis, University of Western Australia, Perth, Western Australia, 6009

1. Vuillermin, .P., et al. (2015) 'Cohort profile: The Barwon Infant Study', *IJE*, 44, pp. 1148-1160
2. Molloy, J., et al. (2017) 'Vitamin D insufficiency in the first 6 months of infancy and challenge-proven IgE-mediated food allergy at 1 year of age: a case-cohort study', *Allergy*, 72, pp. 1222-1231.

Case-Cohort Study Design

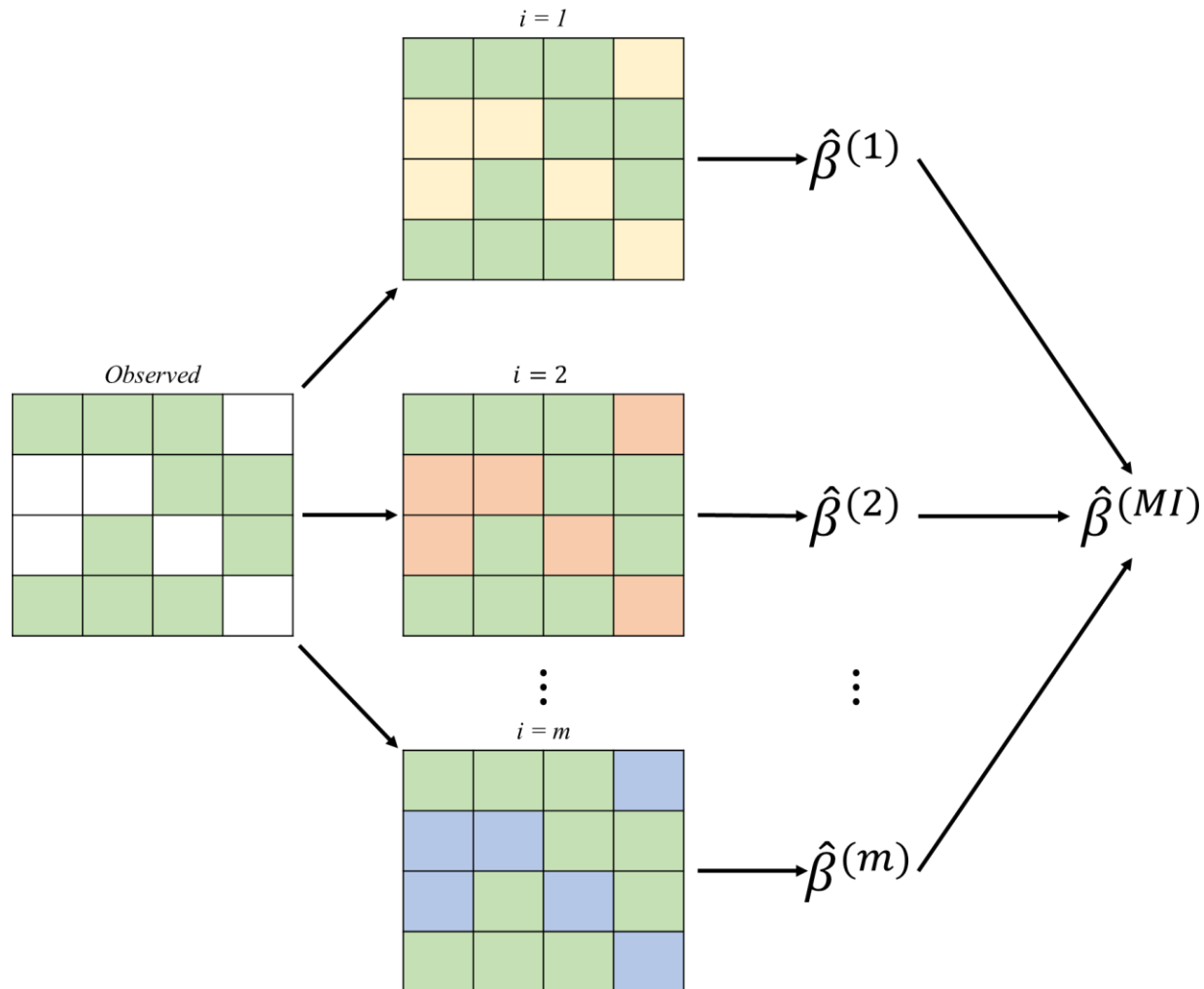
- Expensive exposure collected on a subset



Unequal sampling probabilities

- Standard method of analysis to use inverse probability weighting
- Missing data present in covariates - can be handled using multiple imputation

Multiple Imputation

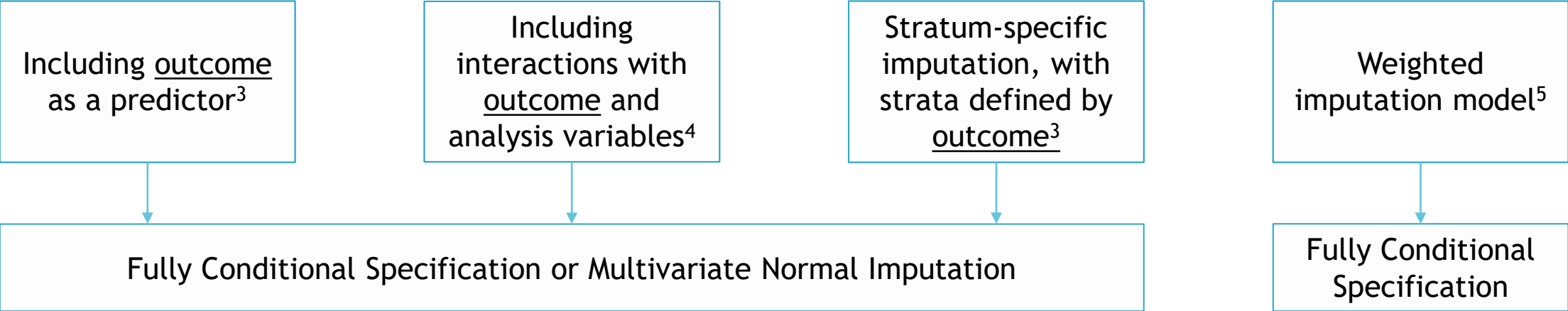


- Joint modelling (multivariate normal)
- Fully conditional specification

Requires compatibility

- Features of the analysis model to be considered in the imputation model - weights!

Proposed Imputation Approaches



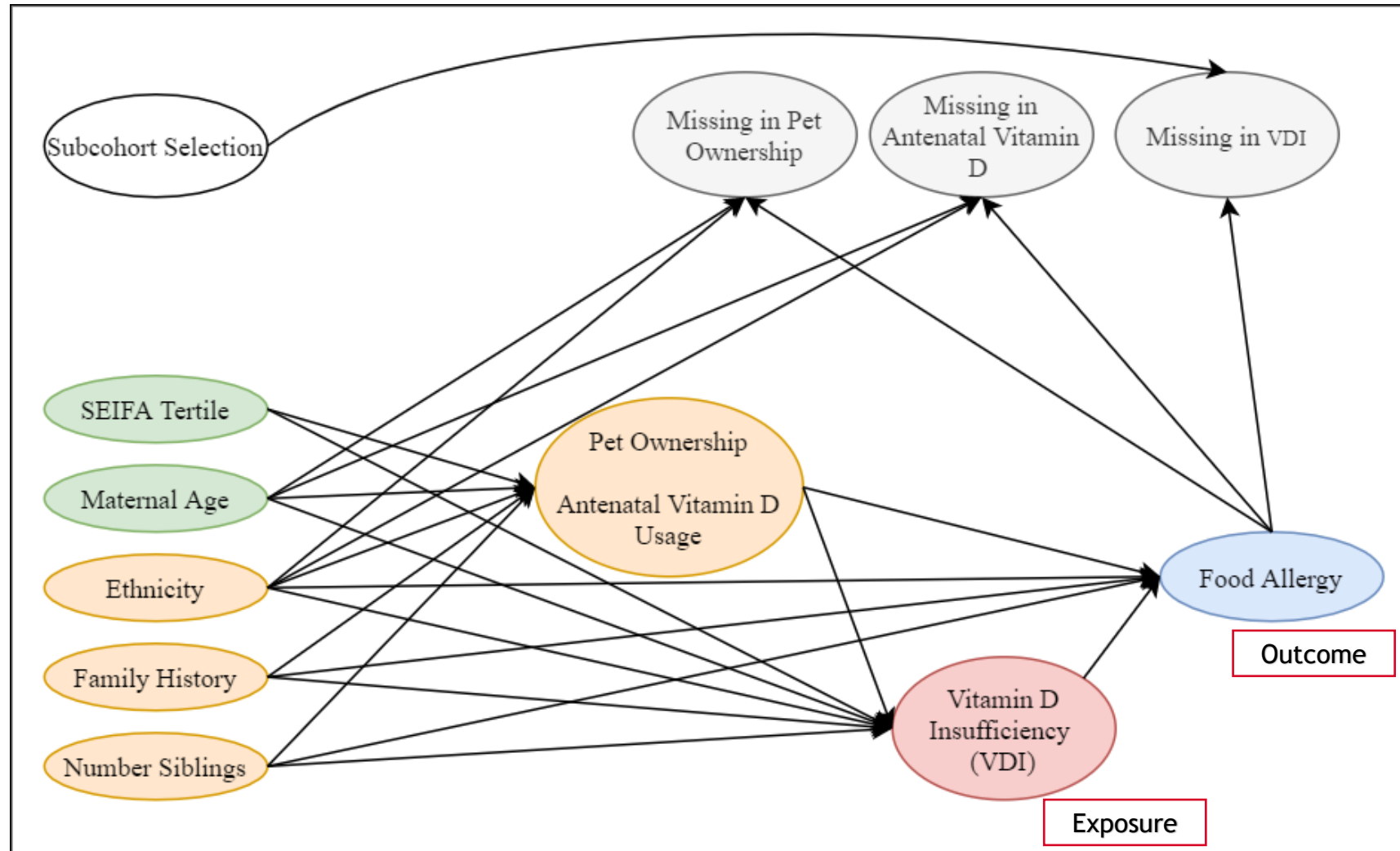
Weighting = Case Status (Outcome)

3. Marti, H. and Chavance, M. (2011) 'Multiple imputation analysis of case-cohort studies', *Statistics in Medicine*, 30, pp. 1595-1607.
4. Carpenter, J. and Kenward, M. (2012) *Multiple imputation and its application*. John Wiley & Sons
5. StataCorp (2017a) *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

Objective

- To evaluate methods for implementing multiple imputation to handle missing covariate data in the context of a case-cohort study where the target analysis is estimating a:
 1. Risk Ratio
 - estimated from modified Poisson regression model
 2. Odds Ratio
 - estimated from logistic regression model

Simulation Study - Data Generation



- Exposure-Outcome
- Weak association
 - Strong association**

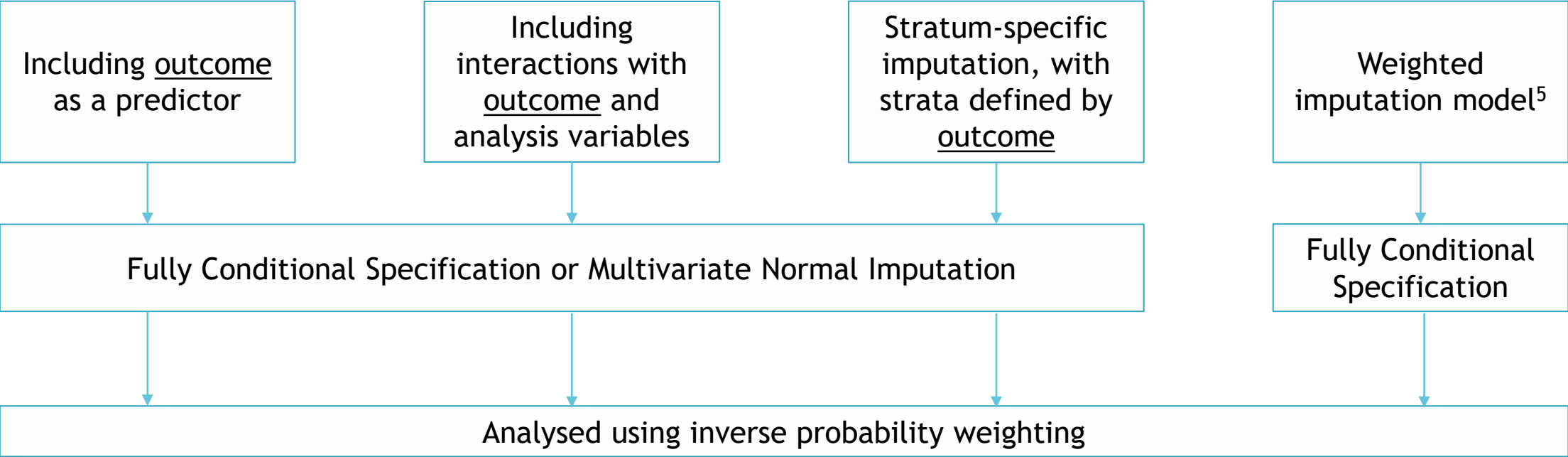
- Missingness Mechanism
- Independent
 - Observed dependent
 - Inflated dependent**

- Missing Covariates
- 15% overall
 - 30% overall**

- Subcohort selection probability
- 0.2
 - 0.3
 - 0.4

- Extreme Scenario
- Enhanced auxiliary associations**

Proposed Imputation Approaches



FCS-WO
MVN-WO

FCS-WX
MVN-WX

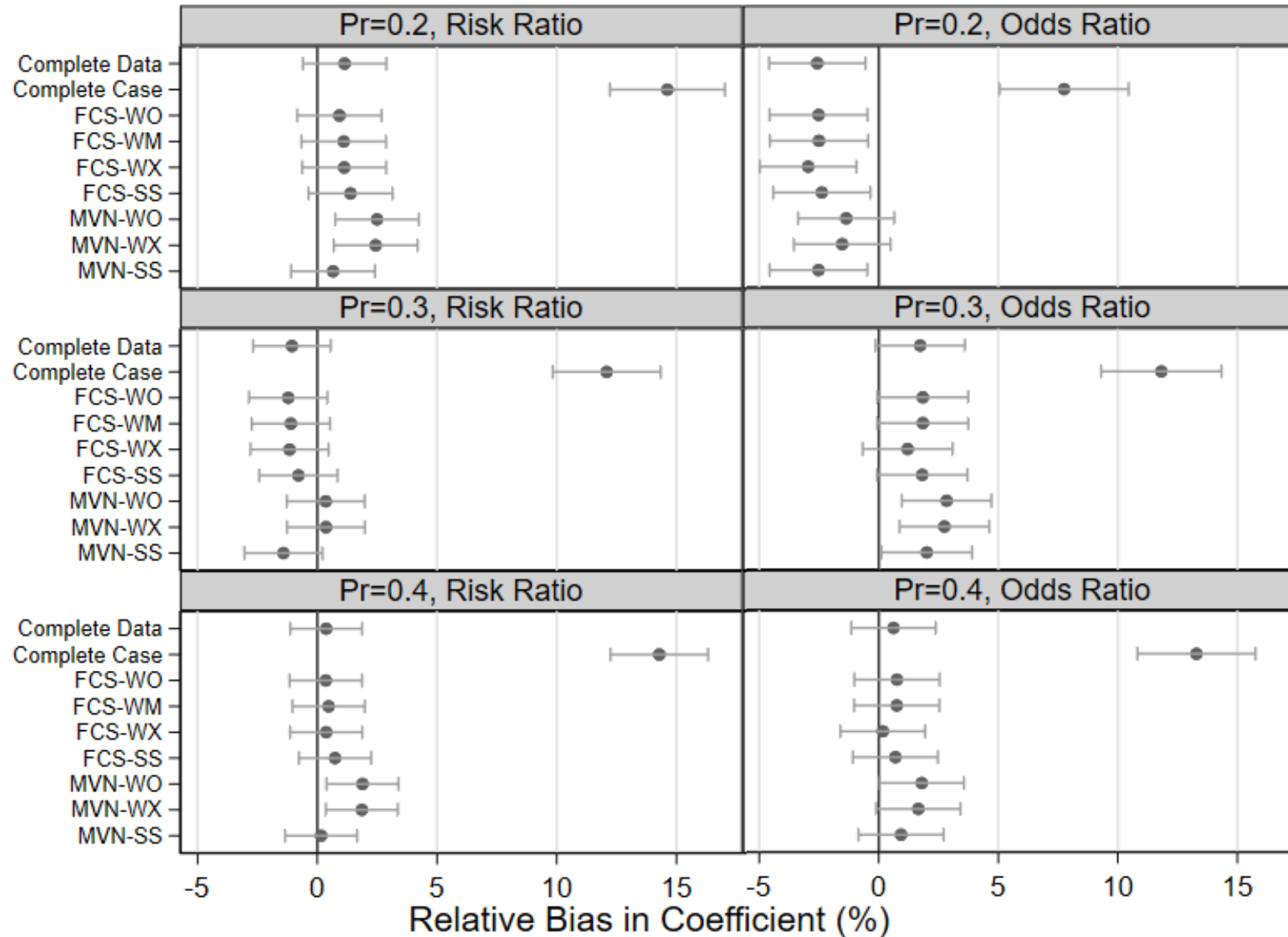
FCS-SS
MVN-SS

FCS-WM

Weighting = Case Status (Outcome)

Conducted complete case analysis

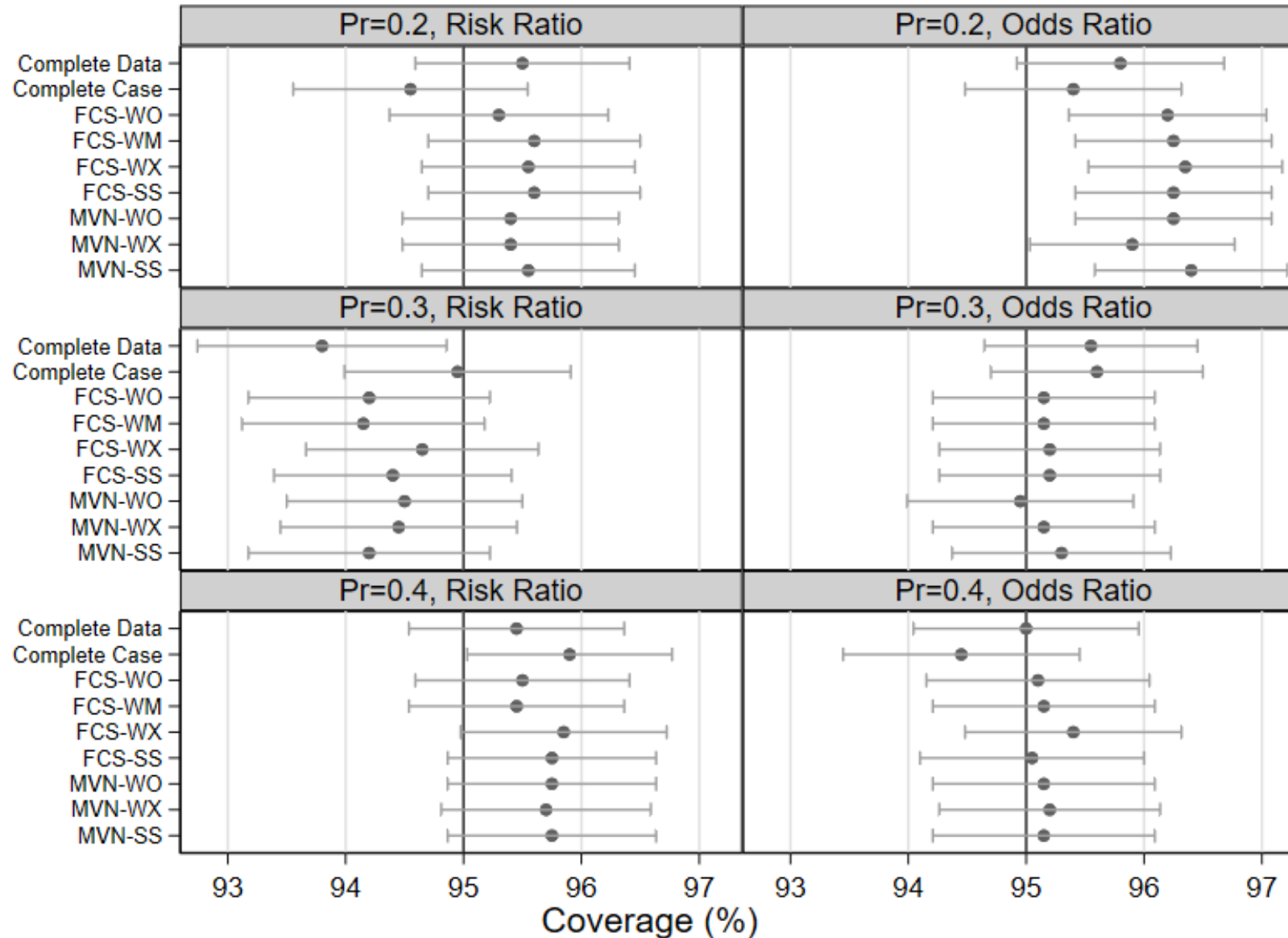
Simulation Study - Relative Bias



Extreme Scenario:

- Enhanced Exposure-Outcome association
- Inflated dependent missingness mechanism
- 30% missing covariates
- Enhanced auxiliary associations

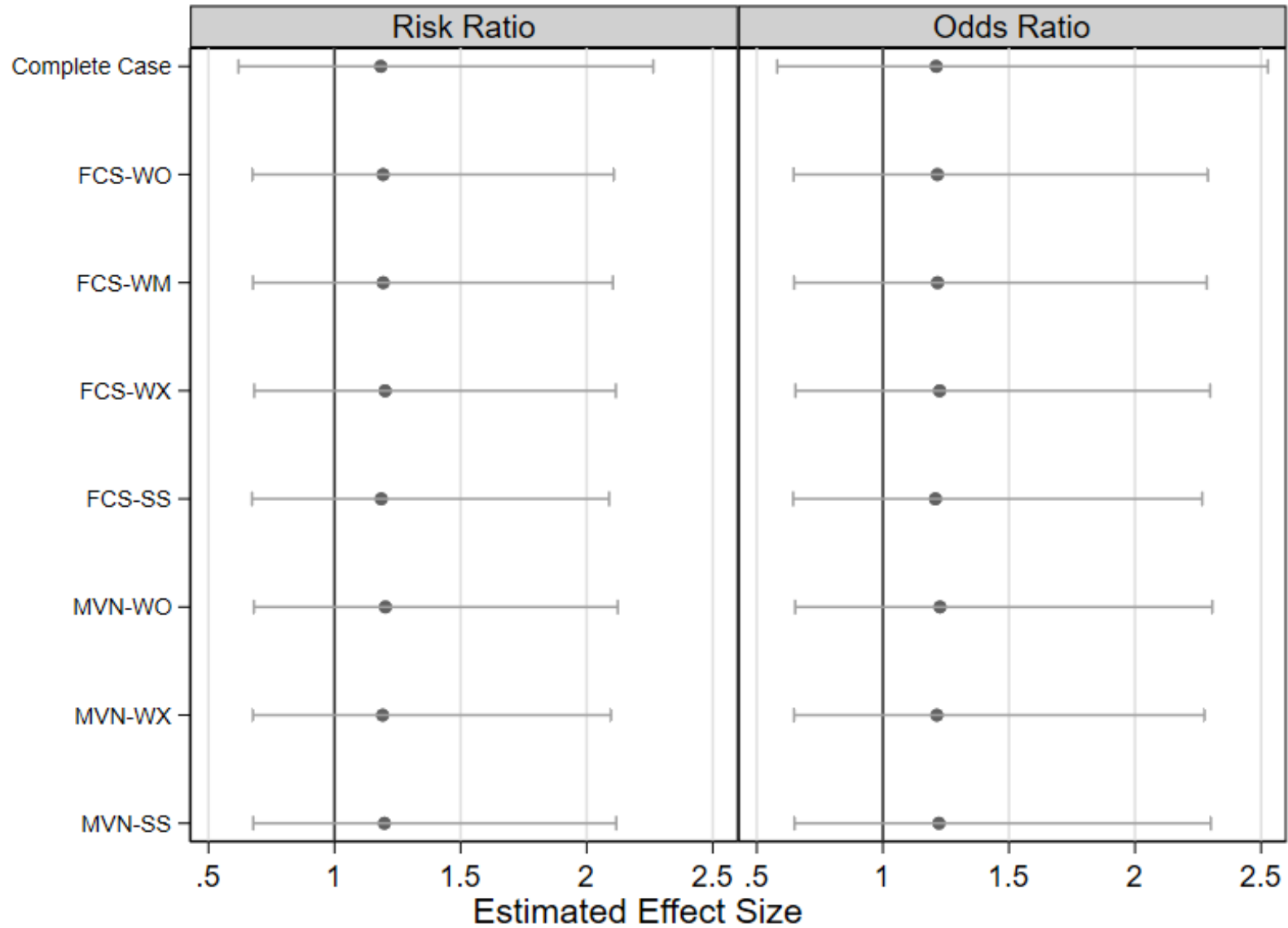
Simulation Study - Coverage



Extreme Scenario:

- Enhanced Exposure-Outcome association
- Inflated dependent missingness mechanism
- 30% missing covariates
- Enhanced auxiliary associations

Case Study



*Error Bars 95% Confidence Interval

Conclusions

- How weighting is included in MI made little difference to analysis, potentially due to only two weight classes in case-cohort studies
- Including outcome in imputation model is a sufficient approach

Next Steps

- Only considered missingness in covariates – often in exposure and outcome as well
- Only considered combining MI and IPW – can also extend weighting for missing covariates or imputation for missing exposure

Acknowledgements

