Sensitivity analysis within the multiple imputation framework: The pattern-mixture method

Julie A Simpson

Head of Biostatistics Unit
Centre for Epidemiology & Biostatistics
Melbourne School of Population and Global Health
The University of Melbourne

Why are the data missing?

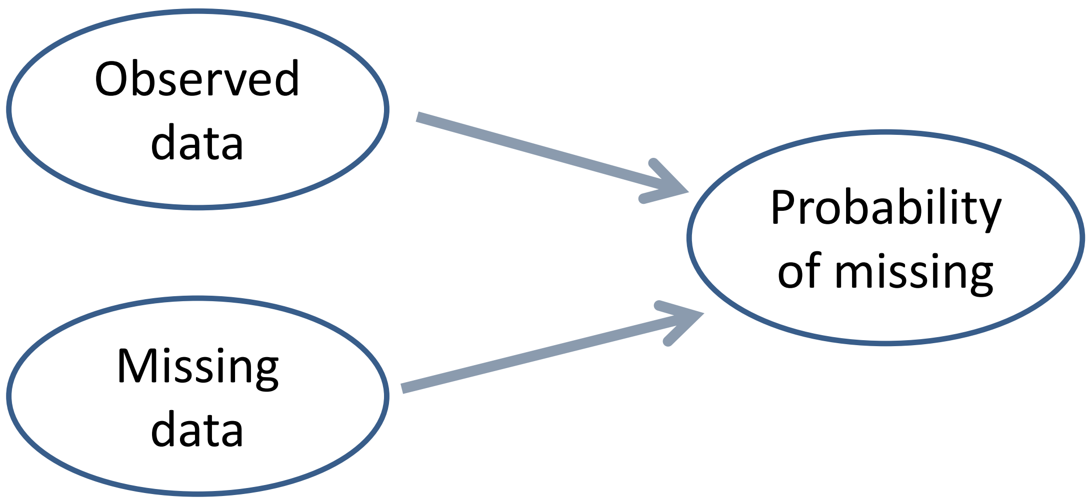
An analysis with missing data must make an assumption some of which are untestable.

There are three assumptions (within Rubin's framework) for the 'distribution of missingness'.

- MCAR – Missing completely at random
- MAR – Missing at random
- MNAR – Missing not at random

Missing not at random (MNAR)

‘Probability of data being missing depends on the values of the missing data, even conditional on the observed data’



Not possible to assess from data whether MAR or MNAR

MNAR models and need for sensitivity analysis

- MNAR: distribution of missing data ≠
  distribution of observed data

- To fit a model under MNAR, **need strong, unverifiable assumptions** about how these two distributions differ (a bit more so than MAR)

- Need to approach this as a **sensitivity analysis** (consider several plausible departures from MAR)

Sensitivity analyses within the MI framework

Assume Y has missing data; R_y indicator for missing Y

MNAR model = model for joint distribution of Y and R_y

Two approaches available:

- Selection-based method (*Carpenter J et al. Stat Methods Med Res 2007*)

$$f\left(Y, R_y \middle| X\right) = f(Y|X).\underbrace{f(R_y|Y, X; \delta_w)}$$

$$logit[P(R_y = 1 \mid X, Y)] = \varphi_0 + \varphi_1 X + \delta_w Y$$

Sensitivity analyses within the MI framework

Assume Y has missing data; R_y indicator for missing Y

MNAR model = model for joint distribution of Y and R_y

Two approaches available:

- Selection-based method (*Carpenter J et al. Stat Methods Med Res 2007*)

$$f\left(Y, R_y \middle| X\right) = f(Y|X).\underbrace{f(R_y|Y, X; \delta_w)}$$

$$logit[P(R_y = 1 \mid X, Y)] = \varphi_0 + \varphi_1 X + \delta_w Y$$

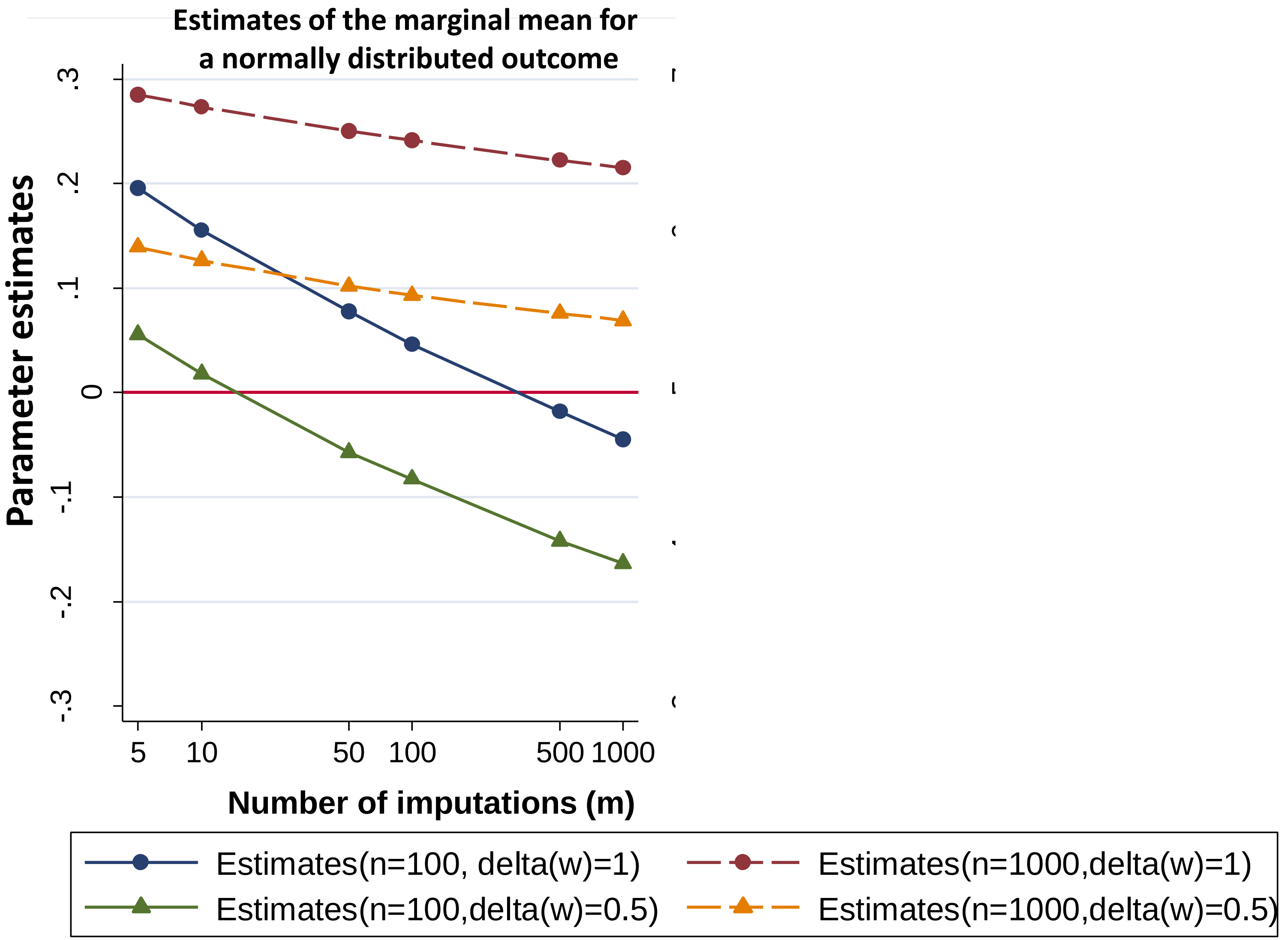
- Pattern-mixture method

$$f\left(Y, R_y \middle| X\right) = f\left(R_y \middle| X\right).\underbrace{f(Y| R_y, X; \delta_{pm})}$$

$$E[\left(Y \middle| R_y, X\right)] = \beta_0 + \beta_1 X + \delta_{pm} R_y$$

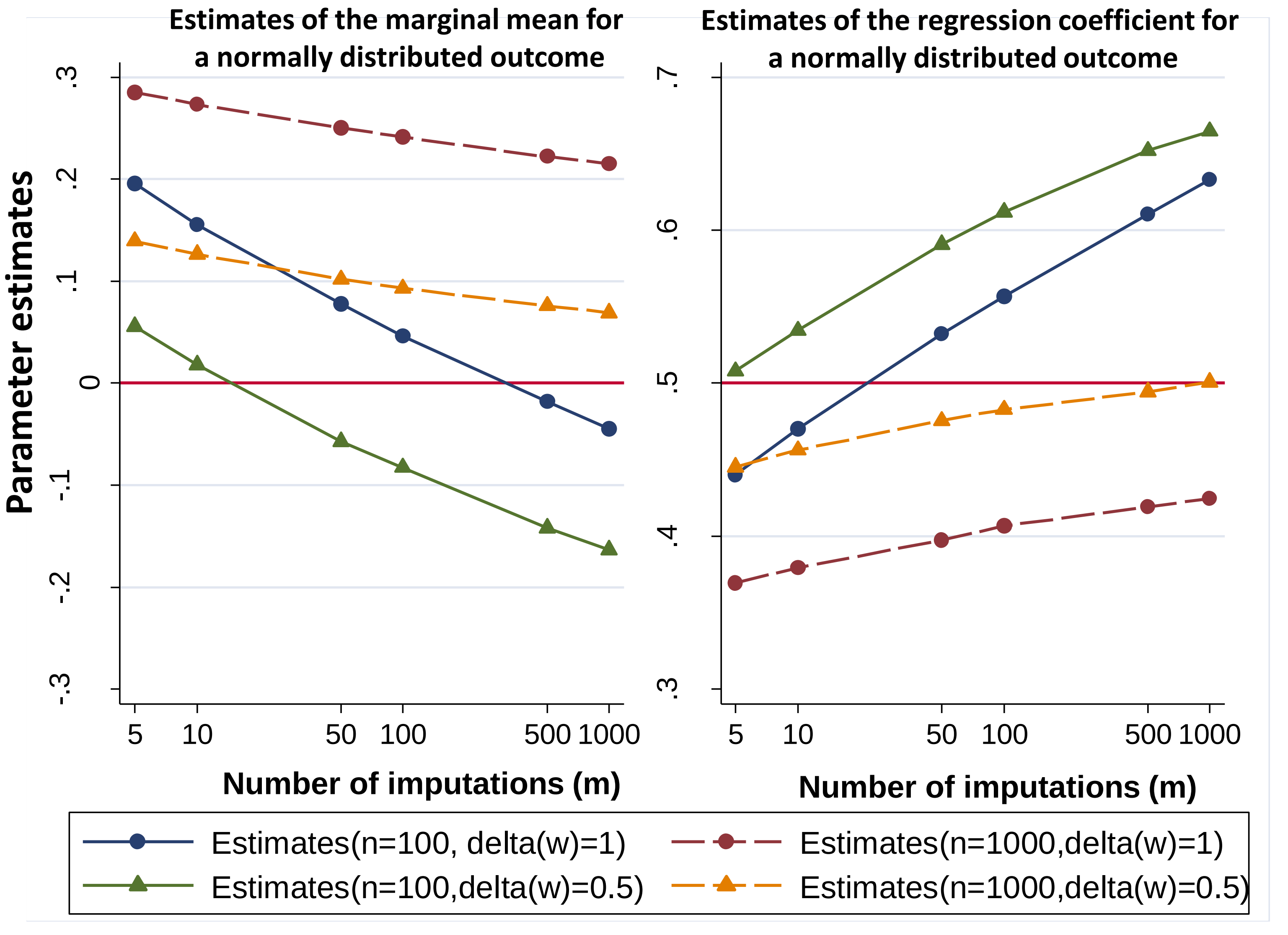
Results of simulations - Weighting approach

Missing data in single variable, δ_w known



Panteha Hayati Rezvan et al. BMC Med Res Methodology 2015

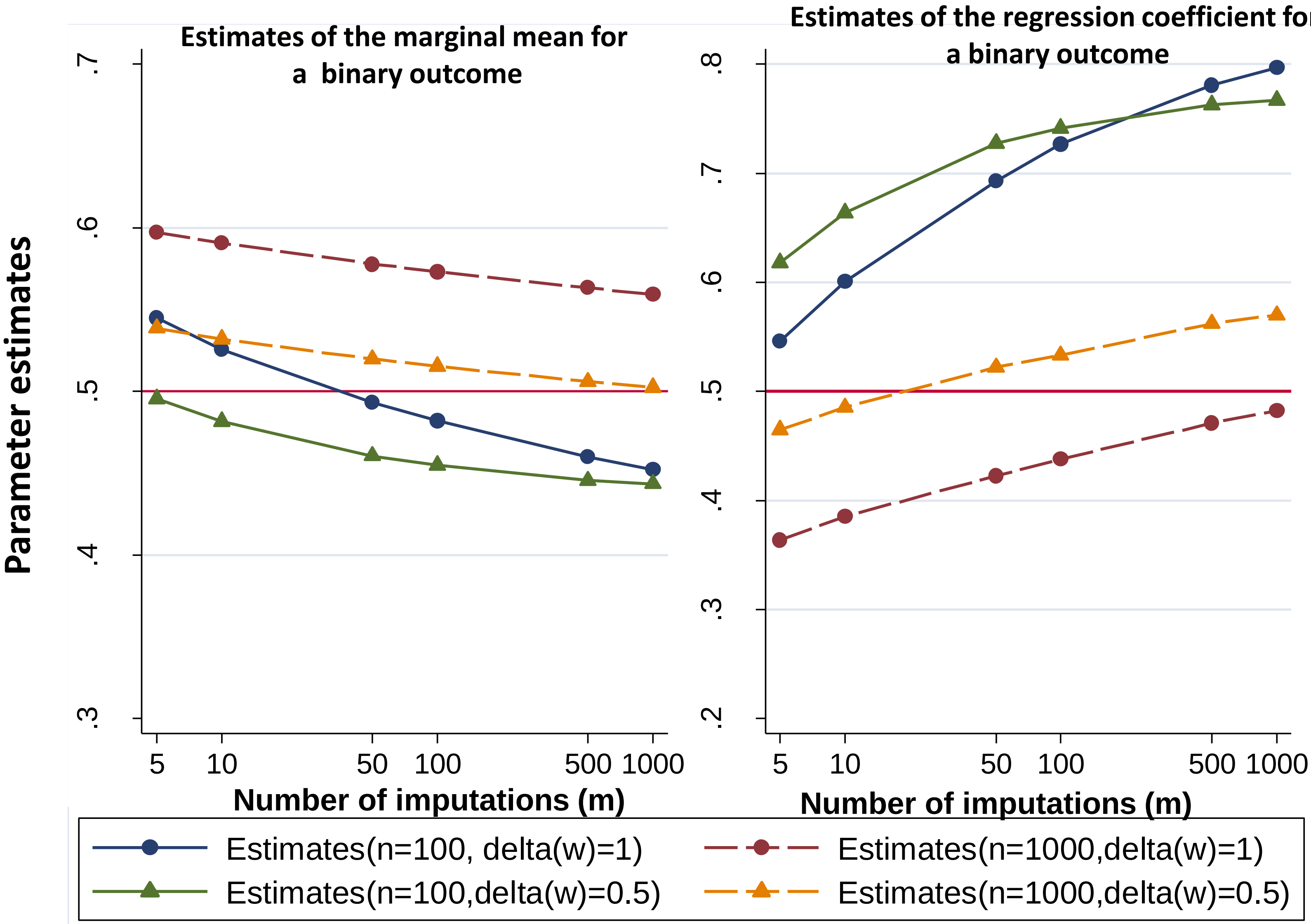
Results of simulations – Weighting approach
Missing data in single variable, δ_w known



Panteha Hayati Rezvan et al. BMC Med Res Methodology 2015

Results of simulations - Weighting approach

Missing data in single variable, δ_w known



Panteha Hayati Rezvan et al. BMC Med Res Methodology 2015

Aims

- To evaluate the pattern-mixture method for missing data in one and two variables using simulation experiments

- To demonstrate the application of the pattern-mixture method using data from the Longitudinal Study of Australian Children (LSAC)
    1. Elicit sensitivity parameters from content experts for the outcome and exposure of interest.
    2. Implement the pattern-mixture method in the statistical software package, Stata.

Pattern-mixture method

Procedure

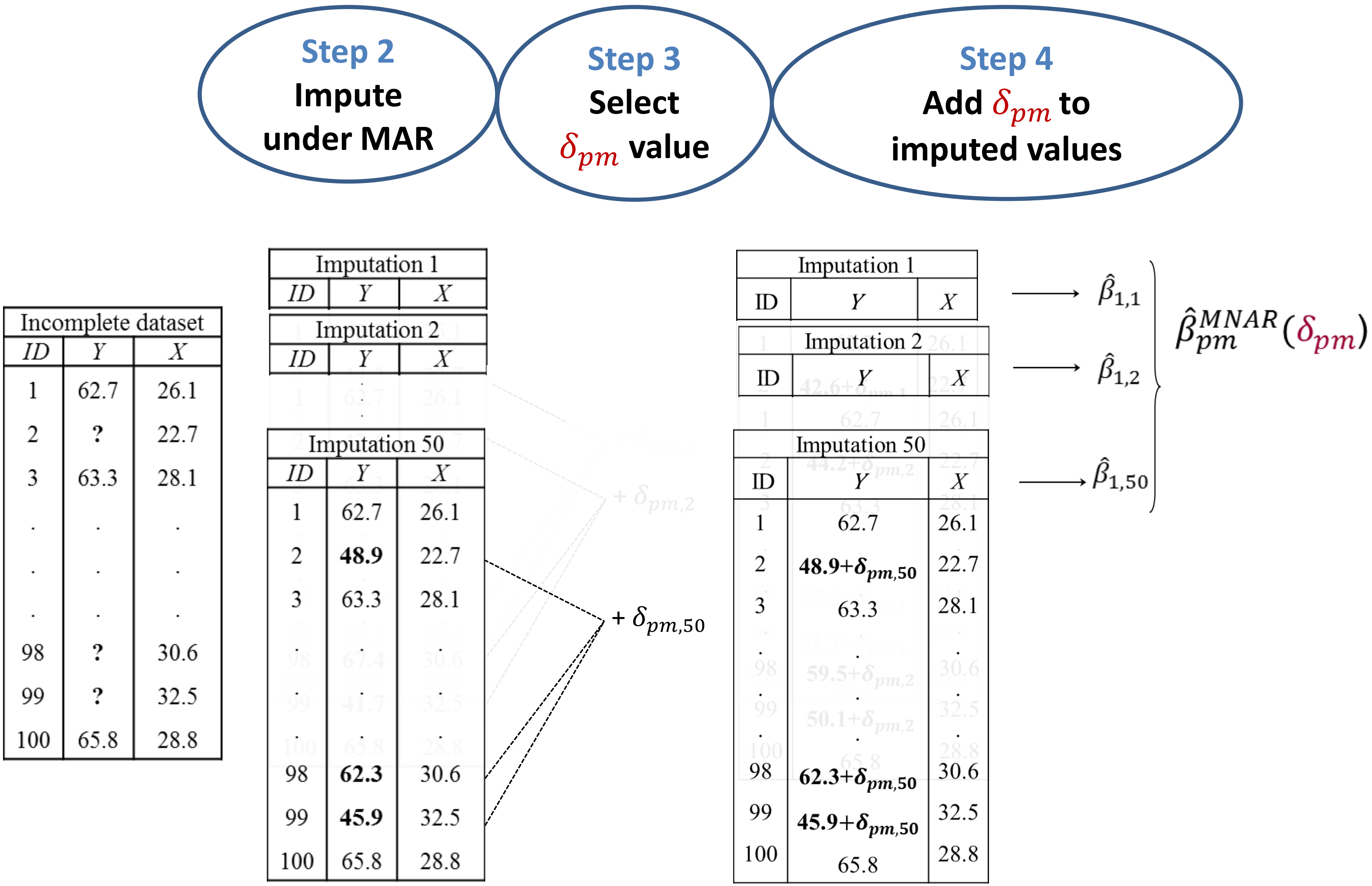
1. Define imputation and analysis models as usual
2. Impute under MAR
3. Select fixed value (or distribution) for δ_{pm}

   $\delta_{pm} = 0$ - imputation assuming MAR

   $\delta_{pm} \neq 0$ - sensitivity analysis, assessing plausible departures from MAR

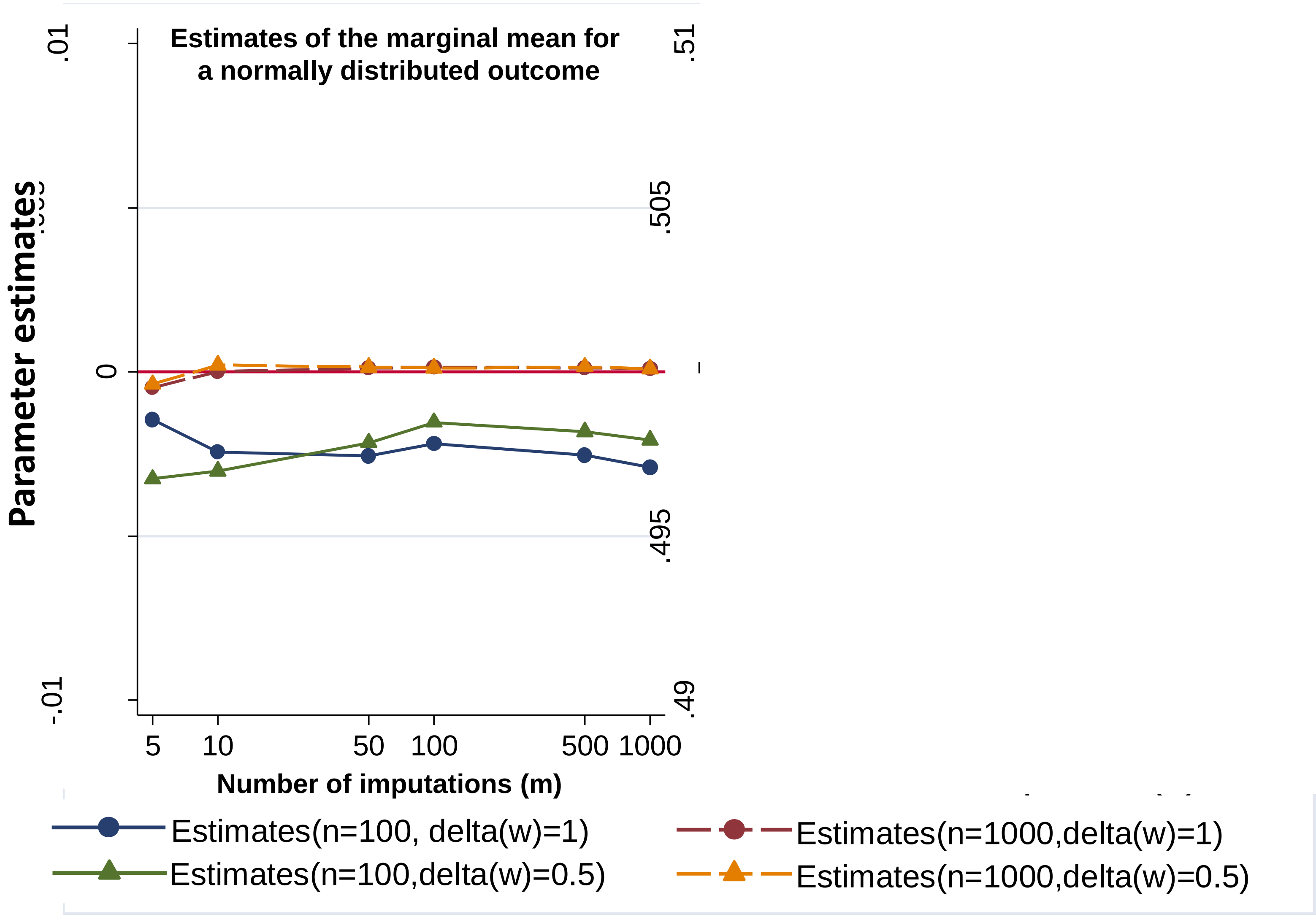
4. For continuous variables with missing data add δ_{pm} to imputed values of imputed dataset 1; repeat for each imputed dataset. For binary variables include an offset in the imputation model.

Pattern-mixture method



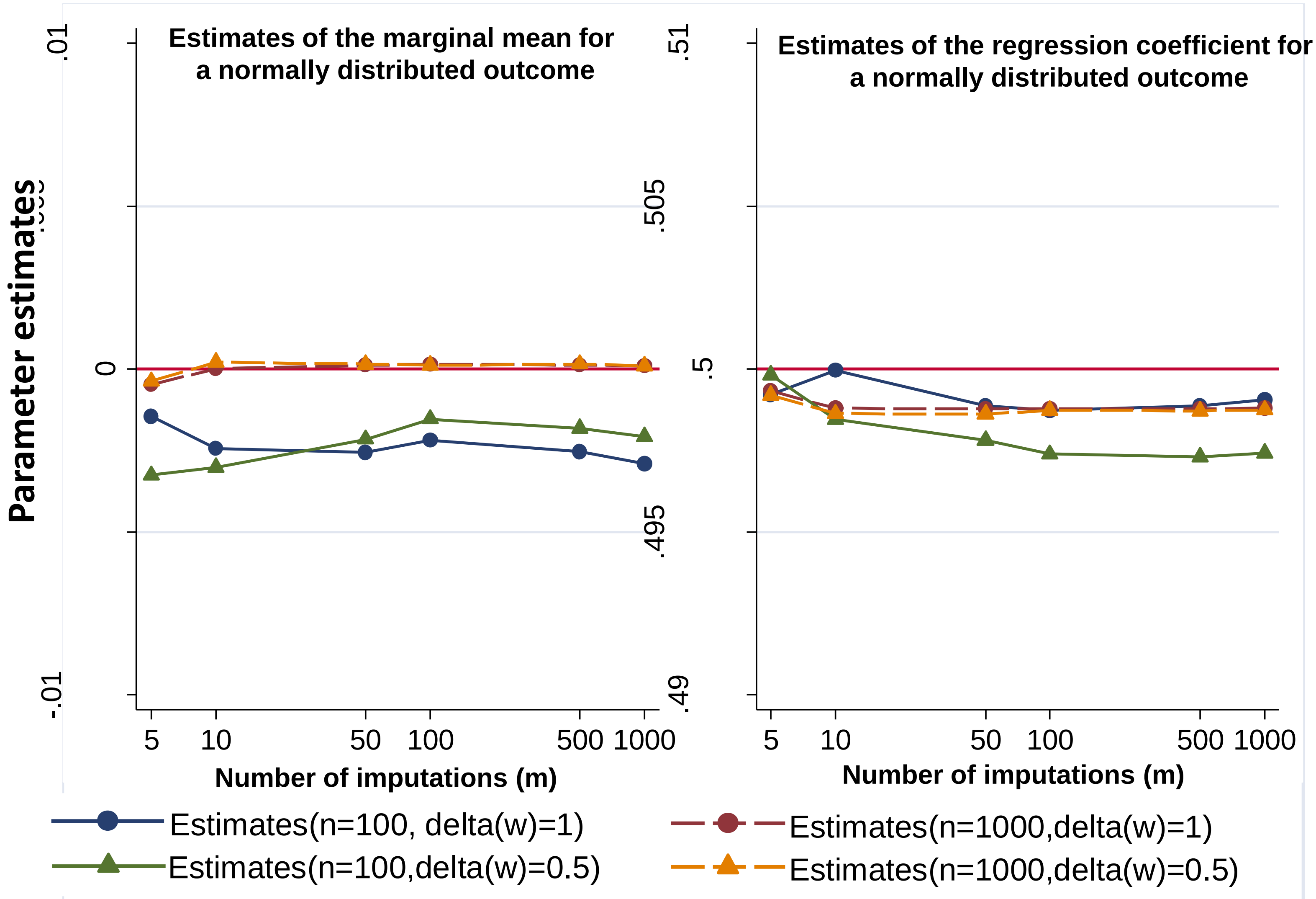
Results of simulations – Pattern-mixture method

Missing data in single variable, δ_{pm} known



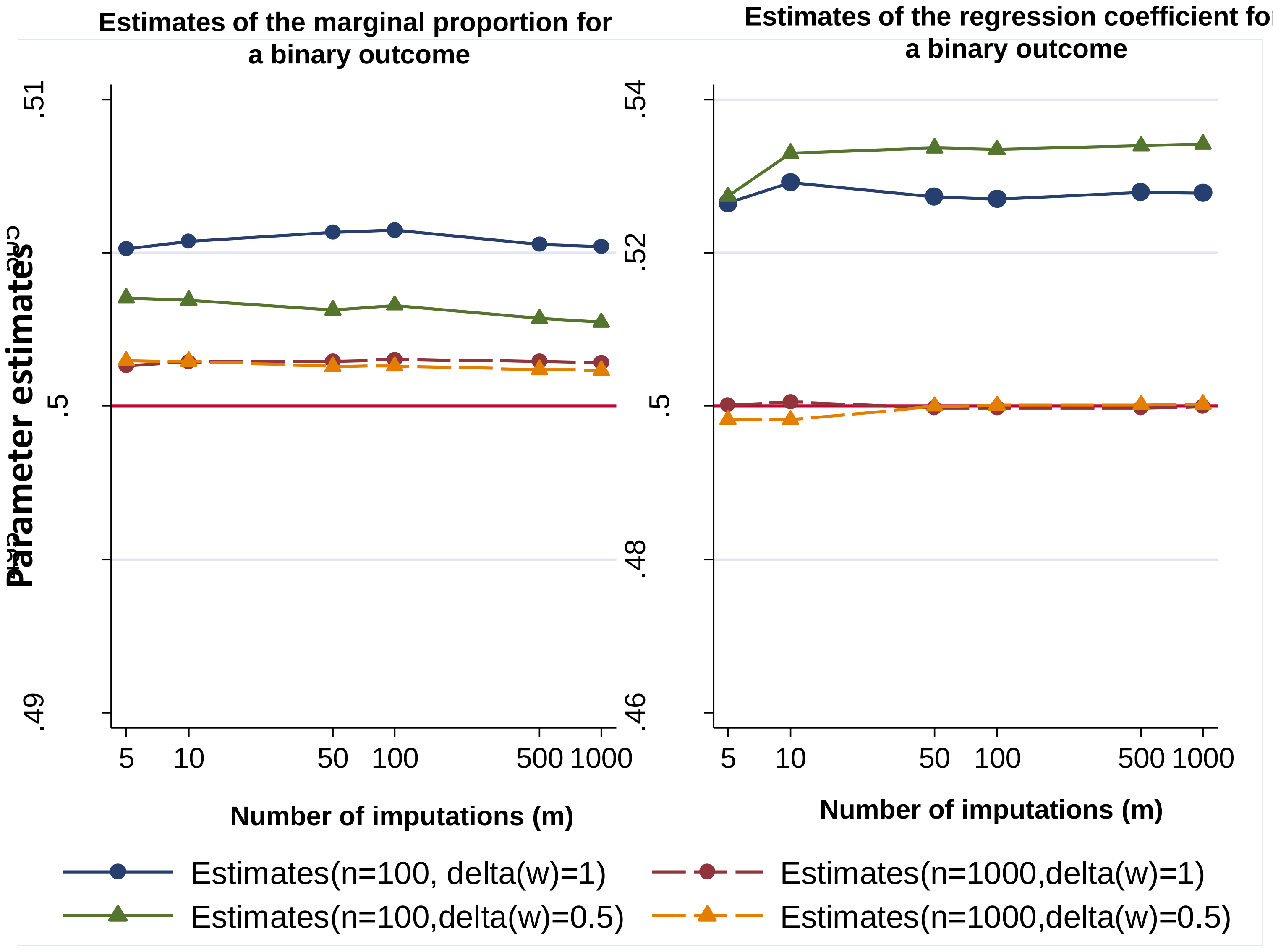
Results of simulations – Pattern-mixture method

Missing data in single variable, δ_{pm} known



Results of simulations – Pattern-mixture method

Missing data in single variable, δ_{pm} known



Results: Simulation experiment

Missing data in two variables – Y (continuous) and X (binary)

Missingness mechanism: $R_x \sim X$; $R_y \sim Y$

Pattern-mixture model:

$$E[(Y|R_y, R_x, X)] = \beta_0 + \beta_1 X + \delta_{pm(x)} R_x + \delta_{pm(y)} R_y$$

	True values	Complete-case analysis	MI under MAR	MI under MNAR
Marginal mean of Y	0.0010	-0.2532	-0.2434	0.0010
Regression coefficient of Y|X	0.4706	0.4072	0.4051	0.4726
Marginal proportion for X	0.5003	0.5467	0.5451	0.5006

Longitudinal Study of Australian Children: Case Study example

Research Question:

To estimate the association between maternal emotional distress at pre school aged children (4-5 years) and the middle childhood total (emotional and behavioural) difficulties (8-9 years)

Exposure variable:

Maternal emotional distress (binary); 16.4% missing

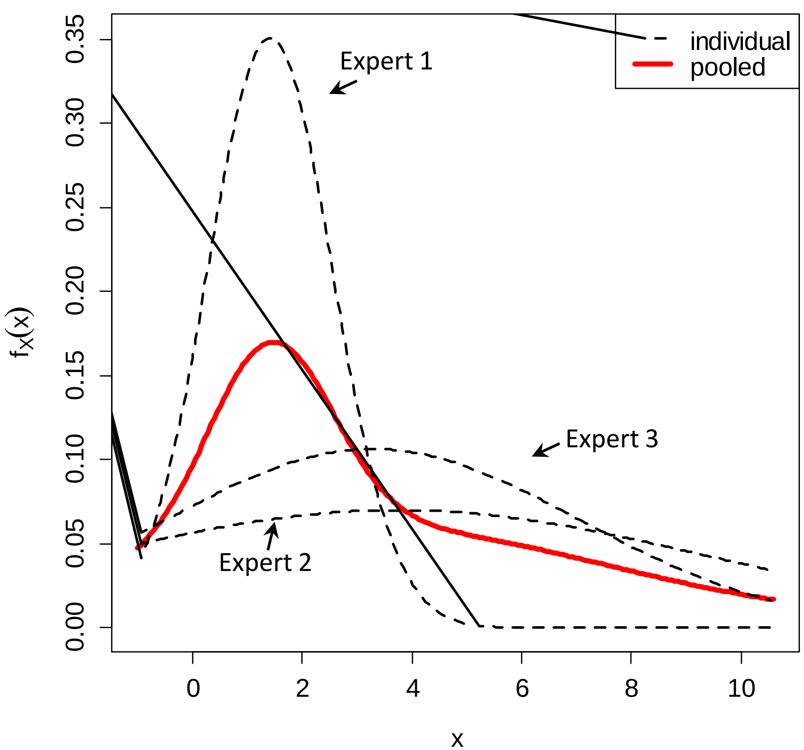
Outcome variable:

SDQ total score (continuous); 23.8% missing

Elicitation of $\delta_{pm(y)}$ from content experts

	Difference in mean SDQ Total score Non-respondents minus respondents				
	Minimum	**Lower quartile**	**Median**	**Upper quartile**	**Maximum**
Hypothetical example	-1	1	3	7	10
Expert 1 response	0.5	0.75	1.3	2.25	2.5
Expert 2 response	-1	1	2.6	8	10.6
Expert 3 response	-1	1	3	6	9

Distribution of $\delta_{pm(y)}$ for SDQ score pooled across 3 investigators



Results:- LSAC Case Study

Association between mother's emotional distress and SDQ total score

	Missing %	Complete-case analysis		MI under MAR		MI under MNAR	
		Coefficient	SE	Coefficient	SE	Coefficient	SE
Mother's emotional distress	**16**	0.59	0.2	0.67	0.2	0.79	0.2
SDQ total score at baseline	0.3	0.5	0.02	0.5	0.02	0.5	0.02
Mother's age	0.8	-0.02	0.02	-0.02	0.01	-0.04	0.01
Sex of study child	0	1.12	0.15	1.12	0.14	1.11	0.14
Study child sibling	0	-0.78	0.24	-0.83	0.22	-0.88	0.23
Mother completed high school	0.9	-0.49	0.16	-0.61	0.15	-0.77	0.16
Mother's current smoker	**17**	0.34	0.2	0.29	0.21	0.4	0.2
Mother's alcohol consumption	**19**	-0.32	0.37	-0.29	0.39	-0.23	0.4
Consistent parenting	2	-0.12	0.12	-0.12	0.12	-0.3	0.12
Child physical health	**16**	-0.03	0.01	-0.03	0.01	-0.03	0.01
Family financial hardship	0.3	0.51	0.09	0.39	0.08	0.79	0.2

Summary

- MAR analysis assumes $\delta = 0$ for some unidentified parameter. <u>This cannot be estimated from the data.</u>
- Sensitivity analysis needed to explore a range of plausible values for δ elicited from content experts (recommend explaining to experts face-to-face).
- Many journals now request these sensitivity analyses are performed following MI.
- Pattern-mixture method
  - Intuitive and performs well (better than the weighting approach)
  - Can be implement in standard statistical software
  - For multiple variables with missing data, more assumptions are required, e.g. independence between R's.

Acknowledgements

Melbourne:

Panteha Hayati Rezvan

Katherine Lee

John Carlin

Emily Karahalios

Margarita Moreno-Betancur

Cattram Nguyen

Helena Romaniuk

Alysha De Livera

Laura Rodwell

Jemisha Apagee

Anurika De Silva

Tom Sullivan

U.K. (Cambridge):

Ian White

NHMRC CRE Grant 1035261 (2012-16)

VicBiostat