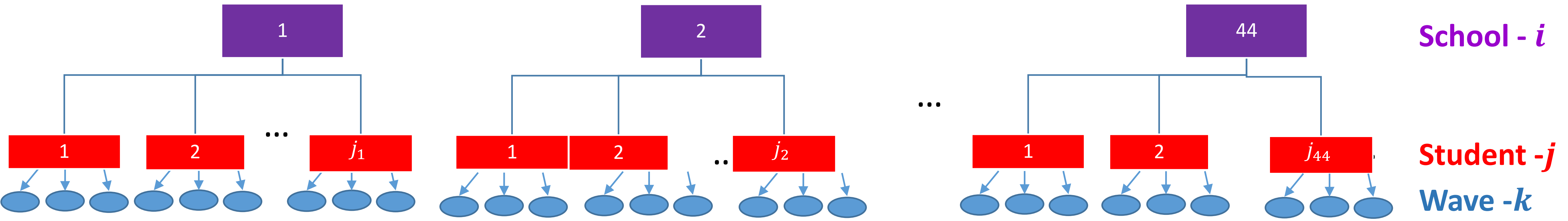
Evaluation of approaches for multiple imputation in three-level data structures

Rushani Wijesuriya

Supervisors :
A/Prof Katherine Lee, Dr. Margarita Moreno-Betancur, Prof John Carlin and
Dr. Anurika De Silva

24th of September 2019

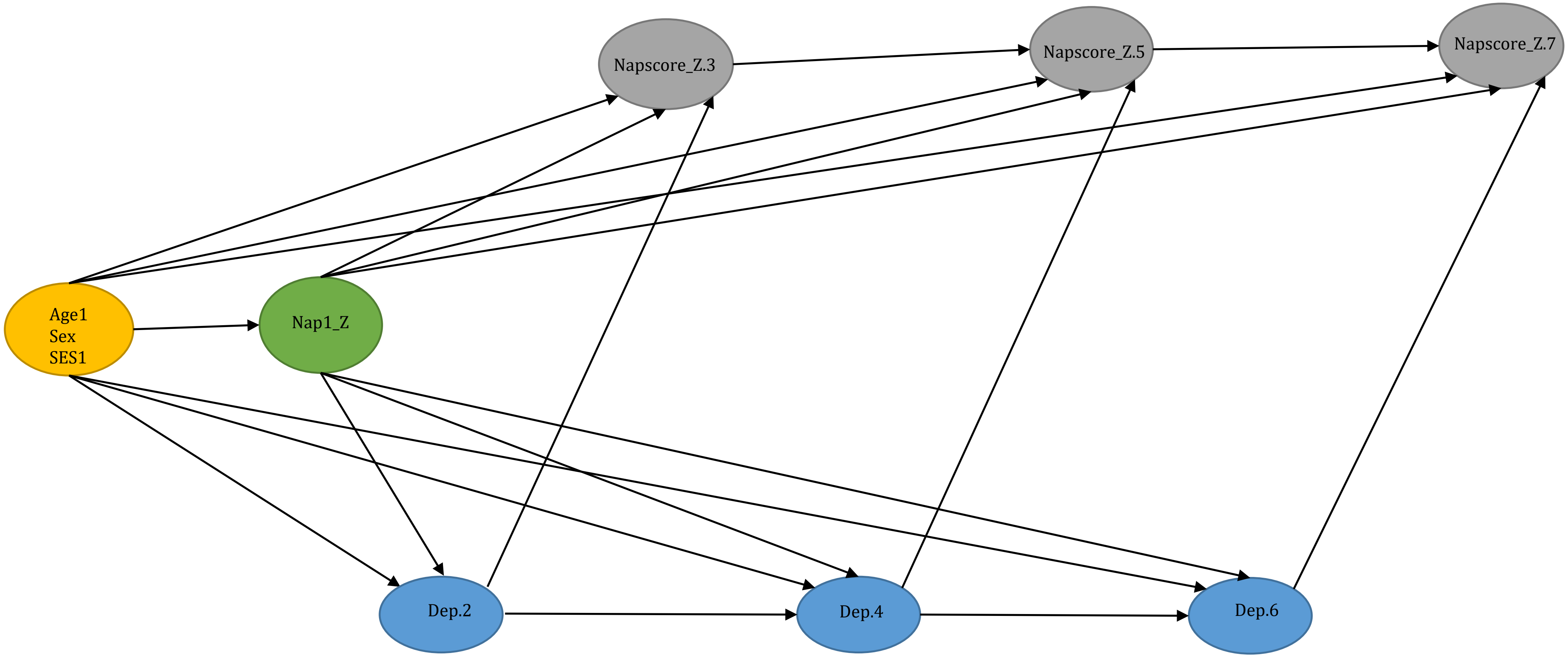
Case Study: Childhood to Adolescence Transition Study (CATS)



<u>Two</u> levels of clustering or <u>Three</u> levels of hierarchy

- Repeated measures within an individual and also clustering by school

Case Study : Target Analysis and Missing Data



Multiple Imputation

- MI is a two stage approach with a separate imputation stage and an analysis stage
- A key consideration in MI : the imputation model needs to preserve all the features of the analysis
- Need to incorporate the clustered structure in the imputation model

Multiple Imputation for multilevel data

How to incorporate the multilevel structure in the imputation model?

Manipulate the standard (single-level) MI approaches

- The Dummy Indicator (DI) approach
- Just Another Variable (JAV) approach (if repeated measures are at fixed intervals of time)

Wide format
one row per individual

ID	Age	Sex	Dep_1	Dep_2	Dep_3
1	8	Male	0.4	1.9	0.2
2	7	Female	1.9	-	2.9
3	9	Male	1.0	3.1	-
4	8	Male	-	2.6	-
5	10	Female	1.5	0.5	1.5

MI approaches based on mixed effects /multilevel models

Long format
One row per wave per individual

ID	Age	Sex	Wave	Dep
1	8	Male	1	0.4
1	8	Male	2	1.9
1	8	Male	3	0.2
2	7	Female	1	1.9
2	7	Female	2	-
2	7	Female	3	2.9

Structure used in the analysis stage

Multiple Imputation for three-level data

How to impute incomplete three-level data?

Manipulate standard MI approaches to allow for both levels of clustering

School clusters :DI
Repeated measures: JAV

- JM-STD
- FCS-STD

One level of clustering : mixed model based MI (specialized for one level of clustering)

Remaining level of clustering : JAV or DI

School clusters :Mixed model based MI
Repeated measures: JAV

- ML-JM-JAV
- ML-FCS-JAV

School clusters :DI
Repeated measures: Mixed model based MI

- ML-JM-DI
- ML-FCS-DI

mixed model based MI for both levels of clustering

- Blimp (FCS)

Simulation of Complete Data

- 1000 datasets were simulated
- 40 school clusters $(i = 1, \ldots, 40)$ were generated
- Each school cluster was populated in two ways: Fixed, Varying
- Four different strengths of level-2 and level-3 intra-cluster correlations

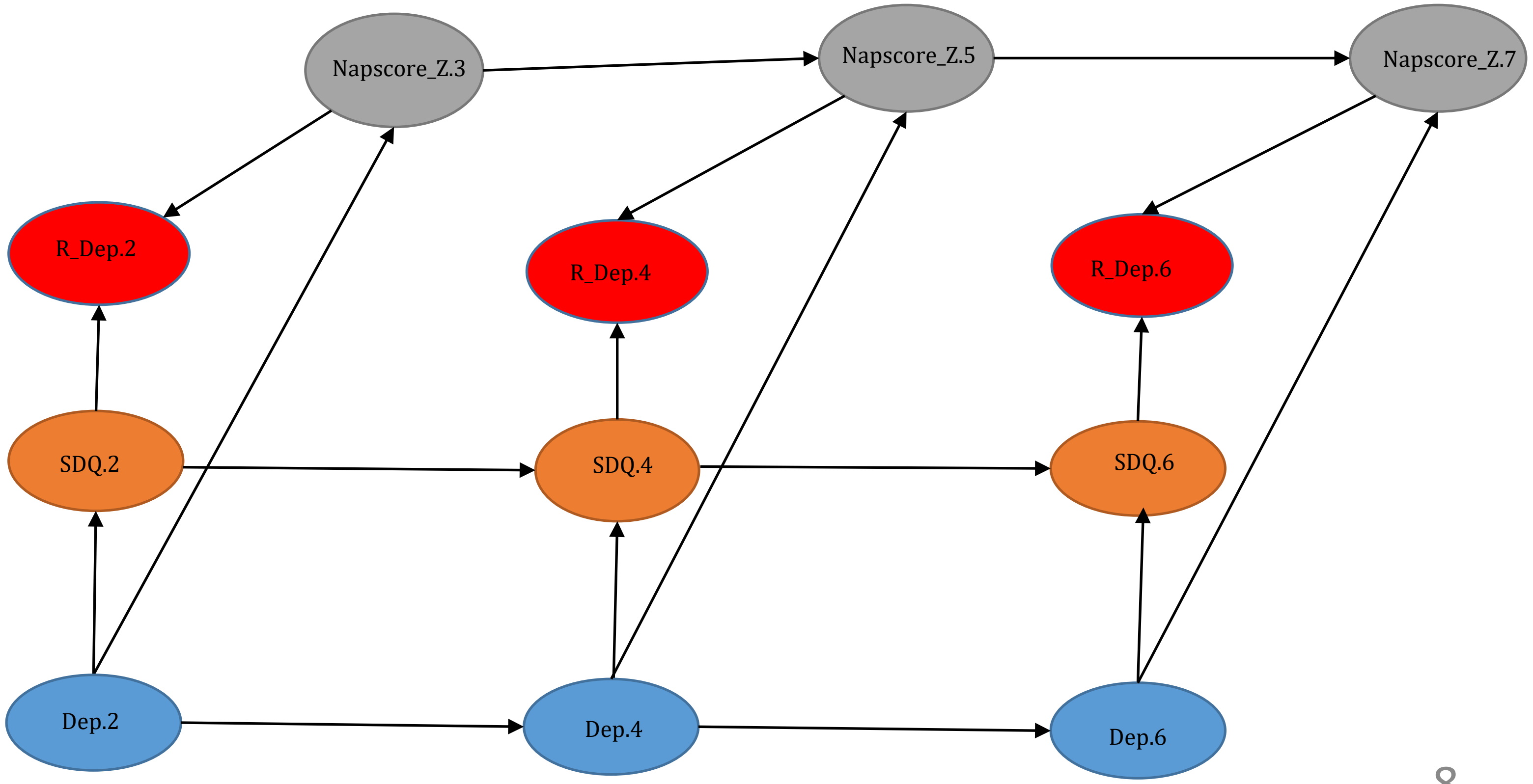
	ICC	
	level 3 (within school)	**level 2 (within individual)**
High-high	0.15	0.5
High-low	0.15	0.2
Low-high	0.05	0.5
Low-low	0.05	0.2

Generation of Missing Data

10% ⟸ Dep.2 ⟹ 20%
15% ⟸ Dep.4 ⟹ 30%
20% ⟸ Dep.6 ⟹ 40%

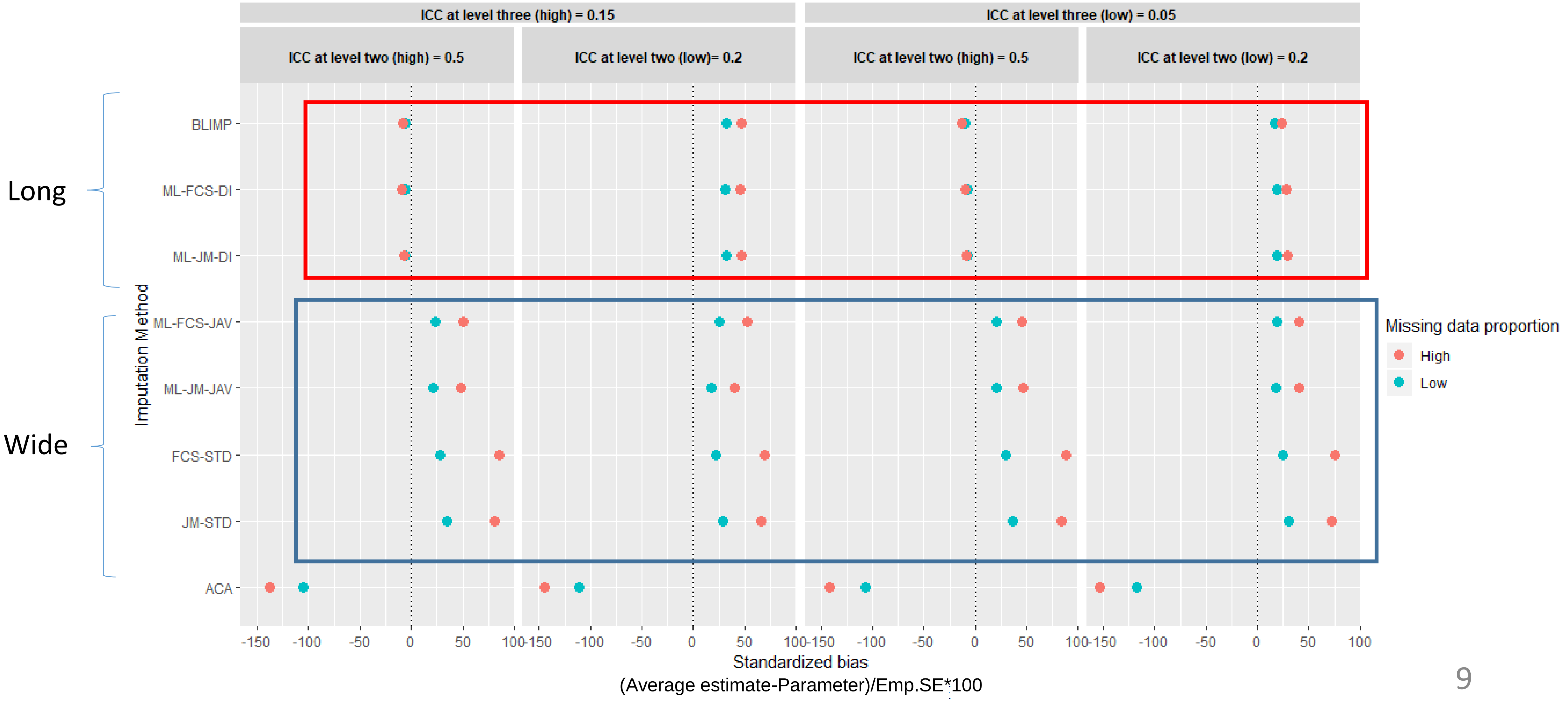
MCAR
Missing values assigned completely at random

MAR- Strong
MAR-Weak



Simulation Study-Results

Standardized biases for the regression coefficient β= (-0.5) - MAR (strong)



Key findings

- Approaches which imputes in long format (BLIMP, ML-JM-DI, ML-FCS-DI) were the best in estimating the effect estimate
- These approaches are also less sensitive to the missing data proportion
- However, ML-JM-DI and ML-FCS-DI can be problematic when the number of clusters is high

Acknowledgements

- Statistical Society of Australia, Victorian Branch
- Supervisors
- VicBiostat

Thank You

You can download the slides at :

https://www.slideshare.net/secret/svP7lOLLC0OzzS

You can contact me anytime at : rushani.wijesuriya@mcri.edu.au

VicBiostat

murdoch children's research institute

THE UNIVERSITY OF MELBOURNE

Statistical Society of Australia