

STATISTICAL INFERENCE FOR PROBABILITY-LINKED LONGITUDINAL DATA

Linked datasets are critical for scientific research. Recent research on the use of linked data has focused on confidentiality, essentially ignoring linkage errors. These errors arise when distinct population units are linked or when a population unit is not linked. They can lead to incorrect and inefficient inferences if ignored, reducing the power to detect important changes and relationships. This project will lead to new and efficient methods for statistical analysis of linked datasets, particularly linked longitudinal datasets. It will develop measures for information loss caused by linkage, allowing assessment of the cost-benefit trade-off in the creation of linked datasets.

The Strategic Roadmap for the Australian Government's National Collaborative Research Infrastructure Strategy states that analysis of linked data, and particularly linked longitudinal data, has the potential to revolutionise Australian public health research. Similar benefits should flow from analysis of linked datasets in other areas, e.g. the Statistical Longitudinal Census Dataset that the Australian Bureau of Statistics intends to create by linking individual records across censuses. These benefits will be maximised by controlling the impact of linkage error when analysing these datasets. This project will develop the statistical theory and related methodology to solve this problem in a statistically efficient manner.

For further information, contact [Professor Ray Chambers](#).