

Maternal age and child development outcomes at age five in Australian Aboriginal and non-Aboriginal children: a population data linkage study

Kindermann, Bindi^{1*}, Chipperfield, James¹, Hansen, Noel¹, Rossiter, Peter¹, and Wright, Jeffrey¹

¹Australian Bureau of Statistics

Objectives

Various organisations are increasingly linking administrative, survey, and census data to enhance dimensions such as time and breadth or depth of detail. Because a unique person identifier is often not available, records belonging to two different people may be incorrectly linked. Estimating the proportion of links that are correct, called precision, is difficult because, even after clerical review, there will remain some uncertainty about whether a link is in fact correct or incorrect.

This presentation proposes some methods for estimating precision when using either deterministic (rules-based) or probabilistic linkage. These methods are model-based and do not require clerical review. The main uses of these methods are to estimate:

1. Precision during the linking process. This is useful to refine how linkage is carried out, such as the choice of linking variables and weight thresholds.
2. Precision after the files are linked. This provides a useful "quality indicator" of the linked data.

Approach

Two methods of estimating precision are described:

1. Simulation - the linking process is simulated many times, whether it is probabilistic or deterministic. The key step being the simulation of the agreement pattern between data sets, based on underlying probabilities.
2. An algebraic estimator - this is applicable for deterministic linking only, and provides a quicker way of estimating precision.

Both methods are investigated using two studies: (i) synthetic data (ii) real data (death registrations linked to census data).

*Corresponding Author:

Email Address: bindi.kindermann@abs.gov.au (B. Kindermann)

Results

The estimators perform very well using both the synthetic and real data, even when assumptions about the independence of linking variables are violated. This suggests that the estimators are robust against moderate violations of these assumptions.

Conclusion

The proposed estimators of precision are a very useful addition to the record linkage tool kit, providing methodical, faster, and cheaper alternatives to many present strategies that rely on clerical review. Estimates of precision are useful in the planning, process, and analysis of record linkage activities.

