

Match quality of a linkage strategy based on the combined use of a statistical linkage key and the Levenshtein distance to link birth to death records in Brazil.

Guillen, Luis Carlos^{1*}, Domenico, Juliana¹, Kenneth, Camargo Jr.², Pinheiro, Rejane¹, and Coeli, Claudia¹

¹Universidade Federal do Rio de Janeiro (UFRJ - Brazil)

²Universidade do Estado do Rio de Janeiro (UERJ - Brazil)

Objectives

To assess the match quality of a linkage strategy based on the combined use of a statistical linkage key and the Levenshtein distance to link birth to death records in Brazil.

Approach

First we evaluated the discrimination power of a statistical linkage key adapted from the Australian SLK-581. The modified statistical linkage key (MSLK-781) was based on the concatenation of the 2nd, 3rd and 5th letters of the mother's family name, the 2nd and 3rd letters of the mother's given name, the 2nd and 3rd letters of the mother's middle name, the child's date of birth and sex. We calculated the proportion of records that have a unique value for the MSLK-781 within the 2013 live births (N=224,038 records) and mortality (N=132,646 records) databases for Rio de Janeiro state. We also calculated the joint unique proportion measure based on the product of these two proportions. Second we evaluated the match quality of a linkage strategy based on the combined use of the MSLK-781 and the Levenshtein distance of the mother's name to link the live births database to death records of singleton children younger than one year of age (N=1,488). To assess the match quality we calculated the sensitivity, the predictive positive value (PPV) and the F-measure.

Results

The proportion of records that have a unique value for the MSLK-781 within the live birth and the mortality databases were, respectively, 97.5% and 98.8%, which yields a joint unique proportion of 96.1%. The match quality measures of the linkage strategy based only on the MSLK-781 were: sensitivity=83.6%; PPV=98.3%; F-measure=90.4%. Combining the agreement on

the MSLK-781 with a Levenshtein distance of the mother's name of less than 4 for the record pairs classification eliminated the false-positive matches (PPV=100%) with a small decline in the sensitivity (81.7%) and the F-measure (89.9%).

Conclusion

Using the MSLK-781 combined with the Levenshtein distance can be used as a first pass for linking birth to death records in Brazil without having to send pairs of records to clerical review.

*Corresponding Author:

Email Address: lctguillen@gmail.com (L. C. Guillen)

