Author's response to reviews

Title: Event based record linkage in health and aged care services data: a methodological innovation

Authors:

   Rosemary Karmel (rosemary.karmel@aihw.gov.au)
   Diane Gibson (diane.gibson@aihw.gov.au)

Version: 4  Date: 7 August 2007

Author's response to reviews: see over
Dear Dr da-Silva

Thank you for the further comments on our article.

Following the continuing concerns of Referee 1 about the issue of missed matches versus false matches, we have re-examined our approach and extended our analysis to examine the trade-off between false and true matches when there are inconsistencies in the linkage data.

The prime purpose of the paper is to provide a tool for deciding whether event-based linkage should be considered for linking data sets before data linkage has been undertaken. When the linkage data are highly accurate, the sensitivity of the linkage should be high (that is, number of missed links should be small), and so the rate of false matches is the key issue. However, as Referee 1 points out, often data for variables used in linkage are inconsistent between the two data sets so that there is a trade-off between relaxing the linkage criteria to increase the sensitivity of the strategy (that is, reduce the number of missed matches) and tightening the criteria to increase its positive predictive value (that is, reduce the rate of false links).

If the theoretical analysis indicates to researchers that event-based linkage could provide a useful linkage approach, once linkage is being undertaken the theoretical approach can then be used to estimate the false match rate among achieved matches, and thus its converse the positive predictive value. When developing a preferred strategy, different match criteria are tested in turn (for example, by broadening match regions) and quite different numbers of links can be obtained due to data inconsistencies. By estimating the number of false matches (and consequently, true matches) for each of these test linkages, the trade-off between false and missed matches can be examined and a preferred strategy decided upon.

To address the above issues in the manuscript we have:

1. In the Background section, added a paragraph (paragraph 9) pointing to the use of the theoretical approach to examine the trade-off between false and missed matches.
2. In the Method section, added a paragraph at the end outlining how the approach can be used to examine the trade-off between false and missed matches.
3. In the Discussion section, after Figure 2, added an example (with a new figure Figure 3) illustrating the use of the method to help identify preferred linkage criteria when matching two specific data sets.
4. In the Conclusions section, adjusted the wording in the second and last paragraphs to reflect the new material.

Minor changes have been made to the abstract to reflect the above changes in the manuscript.

Thus, while the absolute number of false and missed matches cannot be determined without reference to a gold standard, the methods in this paper provide researchers with tools that can help them firstly to decide whether event-based linkage could be
useful, and secondly to identify preferred match criteria once matching between two specific data sets is being undertaken.

The authors feel that the additions to the manuscript instigated by the Referee’s comment have considerably strengthened the paper, and would like to thank the Referee for insisting on the importance of addressing this issue.

We look forward to hearing your response to our amended paper.

Yours sincerely

Rosemary Karmel