Author's response to reviews

Title: Creating the linked data set for the PIAC cohort study: record linkage between several data sets using statistical linkage keys

Authors:

Rosemary Karmel (rosemary.karmel@aihw.gov.au)
Phil Anderson (phil.anderson@aihw.gov.au)
Diane Gibson (diane.gibson@canberra.edu.au)
Ann Peut (ampeut@gmx.com)
Stephen Duckett (Stephen.Duckett@albertahealthservices.ca)
Yvonne Wells (Y.Wells@latrobe.edu.au)

Version: 5 Date: 16 November 2009

Author's response to reviews:

Dear editors,

Reviewer 1 (Roos) now considers the manuscript to be acceptable for publication while Reviewer 2 (Ravelli) has stated that her comments on the previous revision have not been solved and she does not believe that another revision would make a difference.

Throughout the refereeing process the two referees have generally taken opposing views on the direction in which the manuscript should be revised. From the start, Roos in general wanted the authors to take a less technical and more strategic approach while Ravelli’s comments pointed to focusing on a technical discussion of data linkage methods in the paper. These contradictions between the two referees’ suggestions led us to consult the BMC editors on their preferred direction. As a result we adopted the strategic approach proposed by Roos, while still addressing specific concerns raised by Ravelli’s.

Significant changes have been made to address both referees’ comments on the original and resubmitted manuscripts. Roos – a leader in the field of record linkage of health data sets - now rates the manuscript as ready for publication, and that it is ‘an article of importance in its field’. However, Ravelli regards the manuscript as requiring significant amendment.

Ravelli’s comments on the two earlier versions of the manuscript do in part stem from a difference in scholarly perspectives concerning the relative value and roles of competing approaches to record linkage (see below, under Comment 1). On the positive side, we believe that her comments have led to a version of the paper with a clearer exposition of our approach, its benefits and its constraints. However, because of this fundamental difference in perspective, the authors agree with Ravelli that further revision is unlikely to meet her requirements. For this reason, and because Roos is satisfied with the current manuscript, the authors have opted to make what limited changes we can to address Ravelli’s
comments (see below), and leave it to the Journal’s editorial board to make a final decision on the competing views of the two referees.

Response to Ravelli’s comments:
Comment 1: This comment, and a range of comments in the earlier reviews, indicates that Ravelli is at odds with the direction taken by the authors of the manuscript in their approach to record linkage. Ravelli continues to express a strong preference for probabilistic linkage over deterministic linkage; and this paper relates to a project in which we have used deterministic linkage. Ravelli is disinclined to accept that there could be situations where probabilistic linkage cannot be used or where the researchers may not want to use it. The authors are presenting a step-wise deterministic approach which, like probabilistic linkage, allows for variation in reporting of personal details across data sets and which could be used when, for whatever reason, probabilistic linkage is not preferred by a researcher. The authors are not trying to prove that their way is the only way, but rather that it is a useful linkage strategy, as is stated in the Abstract:

“our deterministic algorithm, based on statistical properties of match keys, is a useful addition to the linker’s toolkit. In particular, it may prove attractive when insufficient data are available for the clerical review or follow-up, and the researcher has fewer options in relation to probabilistic linkage.”

Comment 2:
Reason behind criteria choice: Under the heading ‘Phase 3: identification of keys to use in matching’ a sentence has been added before and after the three criteria to explain the reasoning behind the three criteria chosen for identifying suitable match keys.

The sensitivity analysis suggested by the remainder of this comment is beyond the scope of this paper.

Minor corrections
Cover page: revised two email addresses

Abstract conclusion, last sentence: dropped ‘the’ from ‘for the clerical review or..’

Basic strategy, 3rd sentence: dropped ‘for’ from ‘or address data for – information...’

Linkage validation, 2nd paragraph last sentence: replaced ‘this’ with ‘the RCCP to cohort linkage’ in ‘Lack of full name on the PIAC cohort dataset prevented this the RCCP to cohort linkage from being used for the comparisons.’

Appendix A, top of third page: dropped ‘a’ from ‘(just 15 matches with a an estimated ‘worst case’ FMR of more than 20%)’

Table 2: capitalised program names in column ‘Dataset 2’