Author’s response to reviews

Title: Linkage of health and aged care service events: comparing linkage and event selection methods

Authors:

Rosemary Karmel (rosemary.karmel@aihw.gov.au)
Diana L Rosman (Diana.Rosman@health.wa.gov.au)

Version: 3 Date: 4 April 2008

Author’s response to reviews:

Dear Dr da-Silva

Thank you for the comments on our article. We have considered the referees comments carefully, and as a result have considerably revised the manuscript. Please find our response to the referees' particular comments below. We have also included the names of the bodies that granted ethical approval for the study, including reference numbers, as requested (see paragraph 2 under the sub-heading The linkage strategies in the Methods section).

Major Compulsory Revisions

(a) MJG comment 1 and CQ general comment

There was some conflict in the comments by the two referees, with MJG finding the paper contained too much technical detail so that it “is unlikely to appeal to anyone other than those involved in the technicalities of linkage using these Australian datasets” while CQ seemed to want more detail “to define with precision the main steps of the event-based linkage”. After considering these comments in conjunction with all the other comments, we have decided to re-direct the focus of the paper from being a detailed description of a particular example of event linkage to being a more general discussion of event linkage using our Australian example to illustrate issues that can arise and how they can be considered and addressed. That is we have aimed to “give it appeal both in terms of attractiveness of presentation and in terms of ‘generalisability’ of findings” (MJG).

To achieve this we have:

• revised the Background section to provide more context for the desirability of being able to undertake event linkage.

• considerably revised the section under the heading Event-based (E) linkage to describe the issues considered when developing the strategy rather than trying to
describe in detail the specifics of the matching process for our example. This included dropping original Table 1 as being too detailed and application specific.

- considerably revised the Discussion section.
- revised the Conclusion to reflect changes in the other sections.
- changed the title of the paper to reflect the new focus.

(b) CQ comment 1: Methods section paragraph 1

The first two paragraphs of the Methods section have been revised to clarify the choice and scope of the data.

(c) CQ comment 2: description of N linkage

This description has been revised to provide more details on the N linkage method.

(d) CQ comments 3, 4, 5 and 6: description of E linkage

See (a) above. In the new approach some of the detail on the specifics of the data linkage has been dropped rather than increased. We hope that the new version provides a useful description of the approach taken and why two E strategies were trialled. Those interested in the technical details of the linkage are referred to the detailed project report.

Where appropriate, classical terms have been used. However, the authors were concerned that the terms ‘collision’ and ‘doubloons’ are not be meaningful to those not intimately involved with the technicalities of data linkage.

Table 2 has been revised (see new Table 1), and an overview of the matching process using blocking and match passes has been included.

(e) CQ comment 7: measures for comparing methods

To help non-epidemiologists, a new table (now Table 2) showing how to derive PPV, NPV, sensitivity and specificity has been added to aid the discussion under the heading ‘Comparing linkage strategies’. Also, values for specificity and negative predictive value have been added to the detailed results table for completeness (see new Table 6).

(f) CQ comment 8: relationship between Table 2 and Table 3

This issue has been addressed through the revision of Table 2 (now Table 1) in conjunction with the revised description of E linkage (see (a)). Also, additional text explaining the cause of the different numbers of links under the different E
strategies has been included—see paragraph 5 under the heading Accuracy of E linkage

(g) CQ comment 9 and MJG small point 5g: moving information from Discussion to Results

The authors acknowledge that the Discussion section mainly contained results. Therefore, a large part of this section has been moved into the Results section and the Discussion section has been re-worked.

Minor Essential Revisions

(h) CQ comment 1 and 3: terminology

The terminology has been revised to use false positive and false negative terms instead of false matches or mismatches and missed matches.

The term ‘separation’ (for a hospital event) has been replaced by ‘discharge’. The authors think that this latter term will be more readily understood.

(i) CQ comment 2: missing abbreviation

The acronym RAC is now introduced in the Background section.

(j) CQ comment 4: RAC data

This was addressed as part of (c) above.

(k) CQ comment 5: poor sentence

The sentence indicated has been re-written.

(l) CQ comment 6: terminology

In revising the description and results, the expression “SLA-based constrained strategy” is no longer used. We trust that the new way of referring to the different strategies trialled is clearer.

(m) CQ comment 7: 60% rule

Use of the 60% rule to drop poorly performing passes is now explicitly mentioned at the end of the paragraph indicated. It is also referred to in a couple of other places to remind readers of its use in deriving E linkage strategies.

(n) CQ comment 8: conclusion about region size
The Conclusion has been considerably revised in light of the new focus, and consequently the sentence referred to has been dropped.

(o) MJG comment 2: region size

In the new description of E linkage (see (a)) we have added some reasoning about why address information could be poorly reported in our datasets (that is, when people could be in the process of changing address) and therefore why larger region definitions (or alternative address information) may be required. As MJG points out, if address information is reliable then there is no need to use broader regions.

MJG is right in saying that the local geography is important as when using E linkage it is both the size of the population (in the match regions) involved and the rate of events that determines whether the method is appropriate. The example in this paper illustrates one area where the method provides useful results for linking events. However, the importance of local geography does not mean that the method can only be used in the Australian situation; rather it implies that the local geography will need to be considered before applying the method.

The authors trust that the revisions indicated in (a) above, in particular in the description of E linkage and in the Discussion, now make this point come across more clearly.

(p) MJG comment 3: context

The authors agree with MJG that the required accuracy of a linkage strategy as expressed in the PPV varies with the context of the linkage. Thus, while it is very important for clinical trials that identification of death through linking to death data needs to be highly accurate (PPV over 99%), when undertaking linking to allow examination of service patterns and transitions such high accuracy is not required, and PPVs of over, say, 90% will lead to usable data sets for the analysis of population patterns and themes.

The importance of context when deciding sufficiency of linkage accuracy is now explicitly covered in the Discussion (paragraph 2). The authors trust that this, along with the general revisions to the focus of the paper, makes this point come across more clearly.

(q) MJG comment 4: length of stay example

The authors disagree with this comment and CQ made no reference to this issue. The point of this and the other two example analyses is that the degree of error in the E matching does not affect the results when looking at post-hospital destination. Using only the hospital dataset (as implied by the suggested “splitting” of the hospital dataset into two random samples) could not give
meaningful results as the hospital data does not contain the required information on post-hospital destination that is obtained through the data linkage.

The results of the example analyses are now in a section called Utility of E linked datasets in the Results section. The authors have tried to clarify this issue by this rearrangement in conjunction with some minor re-wording in the description of the examples.

(r) MJG comment 5: a few small points

• The authors have addressed the language issues in points 5a, b, c and e.
• Point 5d has been addressed in (a).
• Point 5f has been addressed in (c).
• Point 5g has been addressed in (g).

We look forward to hearing your response to our amended paper

Yours sincerely

Rosemary Karmel