Reviewer's report

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

Version: 2 Date: 1 June 2007

Reviewer: Tim Churches

Reviewer's report:

General

The authors have adequately addressed all of my comments on their paper except for the last one. Unfortunately, my last objection is a fundamental one. What I said was that it makes no sense to consider false-match rates without simultaneously considering missed-match rates, in the same way that it makes no sense to consider, even theoretically, the specificity of a diagnostic test without also considering its sensitivity (and related metrics such as PPV and NPV).

The authors acknowledge this in their final paragraph, in which they state: "In this paper it has not been possible to examine the trade-off between false and missed matches - an issue faced by many linkage strategies. The relationship between these two in the context of movements of linking hospital to residential aged care data will be explored in a subsequent study that compares the event-based linkage strategy with a full names-based linkage."

My suggestion was to combine this paper with the planned subsequent study. As it stands, the current paper is of almost no utility. Why? Because, as the authors implicitly state in their formulae and methods, an arbitrarily low false-match rate can always be achieved simply by making the linkage criteria more strict. In the context of the hospital-to-residential-care event linkage they report, this means simply restricting the scope of the linkage to a smaller geographical region. However, such a strategy will also simultaneously *increase* the missed-match rate.

Thus, as it stands, the paper reports on a method to calculate the false-match rate assuming perfectly recorded data (an unrealistic assumption), but fails to consider the issue of the corresponding missed-match rate, which is reciprocally linked to the false-match rate. Thus an investigator using the proposed method may be tempted to lower the false-match rate as estimated by the proposed method by tightening the linkage criteria or restricting the geographical (or temporal in other scenarios) scope of the linkage, but without being able to consider simultaneously the effect of such actions on the missed-link rate. Therefore the utility of the proposed method is minimal. Given the authors' reluctance to state this fundamental limitation much earlier in their paper, as I suggested, I recommend that the paper be rejected because it too incomplete an advance as it stands - it needs to be combined with a consideration of the missed-link rate for any given linkage strategy, criteria and scope.

What next?: Reject because too small an advance to publish

Level of interest: An article of limited interest

Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.