Reviewer's report

Title: Assessing Correct Record Linkage between Administrative Data and Vital Statistics

Version: 2 Date: 22 November 2005

Reviewer: robert D pates

Reviewer's report:

General
1) This paper has undergone a significant improvement but is still a little "rough around the edges".
2) I don't like the hyphenated term "correct-linkage" which the authors have used throughout the work and I would either delete the term or alter it appropriately. For example in the first paragraph of "Background", perhaps the second question would read better as "What is the accuracy of the record linkages obtained?"
3) The background section is now much better and explains why the work is being done and what's new about it. However, it remains a little disjointed and I think it lacks a clear statement of methodological strategy. For example the authors compare 3 deterministic linkage approaches and use an unique identifier (Personal Health Number) to validate the linkages. I think this needs to be added to this introductory section.
4) Some of the wording needs some work and I would make the following suggestions:
Page 5: ... 1358 were also excluded because the individuals were not CHR residents, ...
I would delete or clarify the sentence:
No postal code is missing and 6762 were used.
Apart from my no understanding this, it contains a mixture of tenses, for which the authors might want to check the entire manuscript for the next draft.
Page 6:
I would change ... inpatient separations (by discharge or death) ... to inpatient discharges (dead or alive) ...
Page 7 (top): I don't understand the following ... The extracted data are stored in the region as well as ...
Page 8: Change ... Although this number is complete ... to ... Although this identifier is complete ...

5) The sections added to the discussion are helpful. On page 12, though, the paragraph on those VS records without PHN is a little tentative. Is there no way of taking those data and characterising them more definitely?

6) I must say I am surprised that the authors don't find significantly higher linkage rates for males than females when using the recommended combination of surname, sex and date of birth (since women more often change their surname). Perhaps this means that the data sources are of high quality and kept up to date?

7) The figures are still a little troublesome -- are they really necessary? They take up a lot of space and are repetitive. I would think about sharpening up the methods to describe the processes in detail, and delete them.

8) The same comments could be made about the tables. There are a lot of numbers to look at / interpret (!). Have the authors tried reducing the number of age categories? Could the same points be made by reducing the number of categories to four: (1-9) / (10-19) / (20-64) / (65+)
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
1) I would say that the title is inadequate and needs some work. I would suggest something along the lines of:
   "Assessing Record Linkage Between Health Care and Vital Statistics Databases Using Deterministic Methods".

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Accept after minor essential revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No

Declaration of competing interests:
I declare that I have no competing interests