Reviewer's report

Title: Using hospital discharge data to identify incident pregnancy-associated cancers: a validation study

Version: 1 Date: 29 November 2012

Reviewer: Jonathan Snowden

Reviewer's report:

In this paper, the authors examine the validity of hospital discharge data in identifying incident pregnancy-related cancers in the obstetrical population of New South Wales between 2001 and 2008. They analyze linked NSW data including birth records, hospital discharges, and a cancer registry. They find that the hospital discharge data has moderate validity in identifying incident pregnancy-related cancers, against the gold standard (cancer registry). The authors clearly explain the gaps in the literature that this paper fills, and make a strong case for the importance of this work. This paper is strong in its conception, analysis, and presentation.

Discretionary revisions:
1. In an otherwise very well-written paper, the second sentence of the abstract results is confusing.
2. Stating that the majority of women did not have cancer is not the most precise explanation for the very high specificities and NPV; it's more that cancer was extremely rare. (Results, paragraph 2).
   a. At the end of this sentence, the authors could state “data not shown,” so the reader knows not to look for these findings in the tables. I wonder if these findings (specificity, NPV) should be presented, though, for completeness.
3. The discussion might be improved by discussing the generalizability of the study findings. For example, Australia’s unusually high melanoma incidence seems relevant given that melanoma (one of the more common cancers in this study, and a majority of the false negatives) may be driving the overall low sensitivity.
4. Table 2 is a bit hard to read, partly because the order of the columns (i.e., numbers are mixed in with validity statistics and CIs).

Minor essential revisions:
1. It was interesting to refer to the sensitivity as “completeness” and the PPV as “accuracy.” I think that completeness makes sense, but I’m not accustomed to seeing accuracy used to denote PPV specifically. “Accuracy” is often used as a synonym for validity, of which PPV is only one component. This may confuse some readers.
2. Be careful about the numerator and denominator when describing the
percentage differences. Results, paragraph 1 implies that the gold-standard melanoma incidence was 56% higher than the index hospitalization incidence, when it appears that the index hospitalization incidence/number was actually 56% lower than the gold standard.

3. The use of “reliability” throughout the paper seems inappropriate (e.g., abstract background and conclusion; last sentence of background; last sentence of conclusion). This analysis seemed to deal exclusively with validity and not at all with reliability.

4. Table 1: typo—register should be registry.

Major compulsory revisions:
None.

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

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