Reviewer's report

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

Version: 1 Date: 18 December 2012

Reviewer: Ileana Baldi

Reviewer's report:

This paper is an interesting and potentially important contribution to the literature for cancer surveillance using claims data.

Major Compulsory Revisions

1. The authors aim to establish the validity and completeness of hospital data for identifying incident cancers diagnosed during pregnancy or within 12 months of delivery.

Actually, this validation study misses “by design” the proportion of cancers associated with early pregnancy loss such as abortion and miscarriage. This aspect and related limitations must be clearly acknowledged throughout the text.

2. The case identification from hospital data is unclear. I expected the study strategy to start from the identification of women who gave birth from 2001 to 2008 using proper ICD-10-AM codes in the chapter “Pregnancy, childbirth and the puerperium” (please specify codes) and continue with the search of their hospital admissions with a cancer diagnosis code ICD-10-AM C00-C96 within -9/+12 months of delivery. Otherwise, the time frame of retrieval of cases in hospital data and cancer registry data could be markedly different and lead to biased validation result. Please clarify.

3. If the authors propose this methodology to estimate cancer incidence, they need to quantify the discrepancy between hospital and cancer registry data regarding the timing of the incidence of disease. Comparison of admission date and registry date should be performed.

4. Analysis of false positives and false negatives provides important insights into the algorithm used and suggests possible modifications to enhance algorithm’s performance. Finally, a critical comment on that should be part of a balanced discussion/conclusion section.

Unfortunately, the papers lacks of a detailed explanation for false positives (maybe in a table and stratified by year) and it is unclear why the authors were unable to investigate reasons of false negatives.

Level of interest: An article of importance in its field

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