Reviewer's report

Title: Cancer diagnosed by emergency admission in England: an observational study using the General Practice Research Database

Version: 1 Date: 23 April 2013

Reviewer: Ileana Baldi

Reviewer's report:

The authors aim to estimate the incidence of first-ever cancers diagnosed by emergency admissions and the effects of some individual-level predictors for such events. This study is supposed to extend previous research by investigating diagnoses for all cancers, and not only for specific cancer types, in a longer study period.

Relying on these premises, the reader would expect some results by cancer type or a convincing reason why effects' modification due to cancer type do not come into play. Surprisingly, only an incomplete distribution of cancer type diagnosed during an emergency admissions (breast n=123, colon n=104, other n=193 and what about the remaining 482 diagnoses?) is given in the text and the lack of an analysis by cancer type is acknowledged as a main limitation of the work. As far as I understand possible explanations for this failure comprise data quality issues, imperfect matching between Read and ICD codes and a model fitting choice.

In order to exclude possible methodological flaws, all these issues need clarification.

Major compulsory revisions:

1. Even though GPRD is a well-established source for epidemiological research, I think that some additional information, especially on Read codes and matching to ICDs, should be given for readers who are not familiar with this UK specific data source. Please extend the description in “Cases of first-ever diagnosis” subsection by including a table with selected Read codes and matched ICDs.

2. Please build a distribution of %emergency diagnosis by cancer type and/or clearly explain which criticality prevented it.

3. As to model fitting, the sentence on page 15 line 15 [We performed….diagnoses] is unclear since either an analysis stratified or adjusted by cancer type, carried out at individual-level, is feasible even in presence of multiple events (i.e. dual cancers). This aspect must be dealt with.

Furthermore, when pooling all cancer diagnoses, the need to perform a selection of 8 covariates out of 12 does not seem justified. The trade-off between degrees of freedom and sample size appears satisfactory even including all predictors.

Please re-analyze data in the light of the previous considerations and convey results in a single table. In any case, I would recommend a model building
approach based on a sensitive choice of predictors taking into account collinearity issues, number of missing, etc… rather than using an automatic procedure. For example the inclusion of ethnicity, with 50% of missing in all routes, is likely to increase residual confounding rather than improve adjustment.

4. In the last section the authors draw attention to the importance of retrieving information on cancer staging and treatment to better understand why some groups are at higher-risk for delayed diagnosis. This is absolutely true but first, they should assess if some cancers are at higher-risk for delayed diagnosis and if these trends varied over time.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests