Reviewer’s report

Title: Are hospital discharge administrative data an appropriate source of information to identify incident cases of cancer? Some insights from prostate cancer

Version: 2 Date: 26 February 2009

Reviewer: Ileana Baldi

Reviewer’s report:

Major Compulsory Revisions

This paper is an interesting contribution on a relevant topic based on appropriate methods and measures. However, I have one major consideration.

Since the introduction of prostate-specific antigen testing, in the late 1980s in US and later in Europe, prostate cancer incidence has increased dramatically. Many of these diagnoses represent clinically irrelevant tumours, and the risk of over-diagnosis and over-treatment poses a clinical and ethical dilemma. For this reason, selected individuals with early-stage prostate cancer may be candidates for watchful waiting or active surveillance followed by active treatment only if and when disease progresses.

In recent years, many changes in prostate cancer treatment have occurred, including new surgical approaches for localised disease, improved irradiation techniques and antiandrogenic therapy, increasingly administered in outpatient care.

Therefore, if the aim of this paper is to assess the potential of hospital inpatient data for detecting newly diagnosed cancer cases (as the authors state), a case study on prostate cancer is the worst setting. The study cannot but confirm what is already known: a large proportion of incident cases is expected to be missed as a consequence of diagnostic and therapeutic management strategies that do not require an inpatient stay. So, regardless of the algorithms used and of HDAD quality and accuracy, sensitivity could not be improved unless outpatient settings data and pharmaceutical claims come into play.

In order to propose an appropriate timeframe of retrieval, I would add a distinction of unrecognized prevalent cases between those diagnosed in the “active search period” (within 6 months or 1 year from PCR incidence date,…) and those diagnosed before. I hypothesize that a delay from the formal incidence date as defined by PCR to first hospitalization could be a possible explanation for unrecognized prevalent cases in 1999.

The considerable economic and societal burden of prostate cancer and its treatment, together with the expected increase in the number of new prostate cancer cases as the population ages, make this disease a very important public health issue. I think that this is the underlying motivation for the present study and the manuscript should be modified accordingly.
In detail:
1. making inference on the potential of HDAD for case ascertainment on the basis of a case study on prostate cancer is misleading and prejudicial. Therefore, the title must be reworded;
2. the introduction must include some background knowledge on the disease and possible therapeutical paths. Last paragraph should be modified [The aim of the study was… Prostate cancer was chosen to explore both objectives];
3. the discussion should take into account all the above considerations and distinguish between “design” and HDAD algorithm limitations. I do not agree with the generalization the authors make as [Nevertheless, the analysis of false positive and false negative underlying causes have provided very useful insights, for a better use of HDAD in the detection of incident cases of prostate and other cancer sites], [Further knowledge is needed…once the new insights we have learned are implemented];
4. the abstract must reflect the changes made in the manuscript. The sentence [Prostate cancer was selected in this study to illustrate the validity of this approach for the Spanish case] should be removed.

Nevertheless, the study may give important indications if some investigations on aspects of quality and effectiveness of prostate cancer care and other health-related objectives are carried out on the basis of HDAD data alone.

Discretionary Revisions.
1. Figure 1 may be removed.
2. Since age is an important determinant of the therapeutical strategy, the authors should stratify the analysis by age (maybe <65 versus >65).

**Level of interest:** An article of importance in its field

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

**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.