Reviewer's report

Title: Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey

Version: 4 Date: 26 July 2012

Reviewer: roberto gnagni

Reviewer's report:

This is a well written paper comparing the prevalence of four chronic diseases, estimated by record linkage of administrative databases, with the prevalence drawn from a General Practice database in different areas of Italy.

Given the large use of administrative database for purposes of monitoring population health and of outcome research, studies aimed to validate algorithms to identify population affected by chronic diseases are highly welcomed.

Results are convincing, and of interest for international readers of a public health journal.

I have the following comments to the authors that I hope could help to improve the quality of the paper:

1. Minor Essential Revisions
   1.1 Methods
      1.1.1 Case ascertainment: please, clarify the statement “…cases were ascertained by applying the algorithms to the data covering 6 consecutive years…”. Was the prevalence estimated over a 6 year span ? To improve clarity, I would also suggest moving the description of the algorithms from the title of table 1 to the “case ascertainment” section
      1.1.2 In the “comparison data” section the study population is aged 16-95, while in the “Data analysis” (and following) 16 +; please clarify.

   1.2 Results
      1.2.1 While additional Table 2 is clear, easy to read and informative, Figure 1, left column in particular, is not adequately described. To most readers it will come difficult to understand the meaning of box, bars and circles if not clearly introduced.
      1.2.2 “The prevalence estimates for heart failure were systematically lower in GP.….”; given the very low prevalence estimated from both administrative data and GPs, can these differences be considered as statistically significant and clinically relevant ?

   1.3 Discussion
      1.3.1 The authors state that nitrates therapy has good specificity in detecting
cohorts with IHD; from the data shown in the paper this conclusion cannot be drawn, as no information is given about the percentage of subjects identified only by nitrates.

1.3.2 Possible limitation of the study to be addressed:
The authors state the regions participating to the project provided only a sample of their populations, limited to specific geographical areas. Populations from GP participating in HSD are also a sample from the general populations. The question is: were the prevalences estimated over populations from the same geographical areas within regions or do these prevalences come from different populations? If so, could this have affected the results?

1.3.4 Given the selections introduced in the sampling process, is the external validity of the result granted? Authors support the use of administrative database to estimate prevalence, but are the areas included in the study adequately representative of Italy?

2 Discretionary Revisions

2.1 In the case ascertainment section, I would underline the fact that these algorithms are not the only ones in the literature, but they have been selected by the authors because…..

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