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:** 12 August 2012

**Reviewer:** Lisa Lix

**Reviewer’s report:**

This is a very well written and interesting article that describes chronic disease case ascertainment methods in 5 regions in Italy using administrative health databases.

**Major Compulsory Revisions**

In Table 2, the percentage of the population who responded to the National Health Survey should be reported for each of the regions.

Figure 2 requires additional clarification. Please provide information about the meaning of the squares, diamonds, and circles in the left-hand figures in a figure note.

The method of data analysis requires additional detail. The authors should explain whether the overall prevalence estimates are adjusted for age and sex. Also, it is not clear why the authors chose to produce separate prevalence estimates for each GP and then report the average and interquartile range for these GP-specific estimates. The rationale for this choice of analyses should be provided. Confidence intervals should be provided for the prevalence estimates so that the authors can comment on whether the estimates from different sources are significantly different.

The discussion section should address the limitations of the study. In particular, agreement between the data sources at the level of the individual patient cannot be investigated with the method of analysis, because the data were not linked at the patient level. In particular, what are the limitations of using an ecological study to assess the validity of administrative health databases for chronic disease case ascertainment?

**Minor Essential Revisions**

Please provide a brief description of the ATC drug classification system and an appropriate reference.

In the background section on page 3, the acronym COPD should be reported the first time that the phrase “chronic obstructive pulmonary disease” is written.

Explain the following sentence in the “case ascertainment” section: “This time span of data that was available for the 93% of the population.”
Discretionary Revisions
The choice of shades of gray in the figures might be changed to produce a
clearer distinction amongst the frequency bars.

Level of interest: An article of importance in its field

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

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

Declaration of competing interests:
I declare that I have no competing interests.