Author's response to reviews

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

Authors:

Rosa Gini (rosa.gini@ars.toscana.it)
Paolo Francesconi (paolo.francesconi@ars.toscana.it)
Giampiero Mazzaglia (mazzaglia.giampiero@simg.it)
Iacopo Cricelli (iacopo.cricelli@genomedics.it)
Alessandro Pasqua (alessandro.pasqua@genomedics.it)
Pietro Gallina (pietro.gallina@sanita.padova.it)
Salvatore Brugaletta (salvatorebrugaletta@virgilio.it)
Daniele Donato (daniele.donato@salute.padova.it)
Andrea Donatini (adonatini@regione.emilia-romagna.it)
Alessandro Marini (alessandro.marini@sanita.marche.it)
Carlo Zocchetti (carlo_zocchetti@regione.lombardia.it)
Claudio Cricelli (cricelli@dada.it)
Gianfranco Damiani (gdamiani@rm.unicatt.it)
Mariadonata Bellentani (bellentani@agenas.it)
Miriam CJM Sturkenboom (m.sturkenboom@erasmusmc.nl)
Martijn J Schuemie (m.schuemie@erasmusmc.nl)

Version: 5 Date: 30 September 2012

Author's response to reviews: see over
Reply to reviewers’ comments

We would like to thank the reviewers for their insight and their constructive comments. We hope we have addressed their concerns in the comments below, that summarize the revisions that have been introduced into the manuscript. We grouped the comments of the reviewers in sections, according to the issues they addressed and the type of revisions they implied.

1 Rationale for the original data analysis

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?

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?

Add in the table and figure the overall data (union data of five regions). This can also give an overview of the Italian context given that there are region of north central and southern Italy.

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 to comment on whether the estimates from different sources are significantly different.

We see now that our rationale for analysis was not clearly specified and we thank the reviewers for the constructive comments. The rationale is now more explicitly described in the first paragraph of the subsection Data analysis of the Methods section. In the Discussion a paragraph (starting with In light of the limitations of the sampling design...) is dedicated to consequences of the sampling design and to external validity. Confidence intervals for estimates were added in Table 3 as well.

2 Add more data and more analysis

For administrative database the author had for patients a unique coded personal identifier for a single regions. Why were used capture-recapture techniques (only administrative database) to estimate the prevalence for any regions? These methods can give estimate prevalence more consistent.

We thank the reviewer for this very interesting advice, which improved our study. We applied capture-recapture analysis to adjust for underascertainment of diabetes cases, and incorporated the results in Table 3 and Figure 2. In order to implement the capture-recapture analysis we carefully revised the coding from data sources and new data became available, especially from regions A and C: this improved slightly the crude estimates as well.
Stratify prevalence by age for diabetes mellitus, ischaemic hearth disease an COPD and com-
ments any difference like you did for hearth failure.

Unfortunately when data collection was performed from the GP database only heart failure disaggregated
data were considered to be of interest for the study, and a new data collection is unfeasible.

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.

The percentage of subjects with IHD identified by nitrates dispensings is now mentioned in the Discuss-
ion in the paragraph starting with Ischaemic heart disease being congruently estimated... and the statement is fully elaborated.

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

This information has been added to Table 2

In the “comparison data” section the study population is aged 16-95, while in the “Data
analysis” (and following) 16 +; please clarify.

We thank the reviewer for this very insightful comment. Not only this inaccuracy has now been amended
in the whole manuscript, but for the sake of discussion in the case of heart failure we stratified the 85+
age band as 85-95 and 96+ so that a more accurate comparison with GP data is performed.

3 Discussion and motivation

Broaden the discussion with a paragraph on the limitations of the study

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?

The limitations of the study are now addressed in a specific subsection of the Discussion section.

“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?

We thank the reviewer for this remark. Statistical significance of observed differences is now discussed
as well.

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...

We modified the sentence that now reads as follows Diverse algorithms for diabetes, COPD and
ischaemic heart disease case ascertainment from Italian administrative databases have been pre-
viously described in the literature. Those published in Simonato et al. [1] were the results of
a workgoup involving two Italian scientific associations of epidemiology and of biostatistics, and
were therefore adopted. However, to deal with the previously reported issue of lack of sensitivity,
the algorithm for COPD was enriched with drug dispensing data [2].
4 Clarity

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?

Explain the following sentence in the “case ascertainment” section: “This time span of data that was available for the 93% of the population.”

The sentence is now reformulated in a hopefully clearer way after revision by a native language speaker.

To improve clarity, I would also suggest moving the description of the algorithms from the title of table 1 to the “case ascertainment” section.

The description of the algorithms is now in the “case ascertainment” section.

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.

This acronym is now added to the phrase.

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

The WHO Collaborating Centre for Drug Statistics Methodology’s website is now referenced and the system is described as the official WHO drug classification system.

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.

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.

It was considered that the first reviewer’s reference to Figure 1 as a figure “with columns” means that in fact the comment is to Figure 2, and coincides with a very similar remark form the third reviewer. A more descriptive caption was added to Figure 2.

References

