Reviewer’s report

Title: The Predictability of Claim-Data-Based Comorbidity-Adjusted Models Could Be Improved by Using Medication Data

Version: 4 Date: 14 December 2012

Reviewer: Patricia Halfon

Reviewer’s report:

The predictability of claim based comorbidity -adjusted models could be improved by using medication data

1. Is the question posed by the authors well defined?

No; the main aim of the study is not clear. Is it to check the accuracy of detection of certain co morbidities (those defined by Charlson or Elixhauser) by an algorithm based on drug information? In this case the outcome would be a concordance coefficient or sensitivity /specificity against a gold standard (if it is possible to have a gold standard)? or prediction of in hospital death as additional data to diagnoses? ; If the aim is only prediction, why did the authors limit their study to a few specific conditions already detected by diagnoses?

Co morbidities defined by Charlson are not benign condition and generally are rather well collected in hospital claims. It is not surprising that drug information had little added value in predicting in hospital death

Some co morbidities defined by Elixhauser are more likely undercoded (hypertension, obesity, mental diseases): unfortunately, most drugs therapeutic groups which might detect these conditions were not considered by the authors. I think that drug information is mainly useful for outpatient because diagnoses data are not widely available in ambulatory care. There are several co morbidities scores which have been studied and found useful to predict certain outcome (especially total costs prediction). Inferring adequately health status from drugs is certainly an important research question. But I am not sure that the questions posed by the authors are important to the research status in this field.

2. Are the methods appropriate and well described?

The studied population is limited to few conditions and thus the generalizability of the study is problematic. There is a prediction model by condition. Why do the authors exclude “most responsible diagnoses that have a wide range of clinical spectrum: sepsis, respiratory failure, etc. What would be the use of such prediction models?

3. Are the data sound?

The authors do not give the model coefficients of the predictive categories. This is important to check their relevance (are there categories with a non significant coefficient or even a negative coefficient)

4. Does the manuscript adhere to the relevant standards for reporting and data
deposition?
Table 3 and table 5 are problematic; the large excess of cases detected by drugs in some conditions suggest a specificity problem; the fewer cases detected by drug in other conditions suggest a sensitivity problem. Consequently, I am concerned by the term fraction of additional co morbidities inferred by prescribing information.

5. Are the discussion and conclusions well balanced and adequately supported by the data?
No; the results suggest poor added value of drug information whereas the authors conclude that “prediction models are improved…

6. Are limitations of the work clearly stated?
Not enough from my point of view

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
yes

8. Do the title and abstract accurately convey what has been found?
No see point 5

9. Is the writing acceptable?
I can’t answer; not a native English speaker

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.