Reviewer's report

Title: The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease

Version: 1 Date: 23 October 2010

Reviewer: Reijo Sund

Reviewer's report:

This is very interesting and well written manuscript. I have only a few minor comments that should, however, be considered before publication:

Does the PFT test (that was performed for each patient in your data) "automatically" result in COPD inpatient or outpatient visit in your administrative data?

Exclusion of lung cancer patients should be feasible with administrative data, but how to do BMI-based exclusion with administrative data only? Should one try to take this into account while using administrative data to identify COPD?

Was the data exactly +/- one year from index date, or +/- one "register years" (i.e. full calendar year before the index calendar year and full calendar year after the index year)? This is important, because it may affect the initial database queries to administrative registers (at least if applied in non-VA context).

Justify more carefully why potentially COPD-related secondary diagnoses in outpatient visits were not considered as COPD visits.

Were comorbidities detected using only data prior to index date (and how many years "all previous outpatient visits" mean here)? Why only outpatient visit were used to detect comorbidities (and not inpatient admissions or pharmacy data)? Why didn’t you use (or refer to) widely applied techniques for detecting comorbidities such as Charlson or Elixhauser categories?

Please describe the structure of pharmacy data more precisely. Were those data describing prescriptions, purchases or reimbursements? What if there were more than one canister per prescription?

In this study you had a fixed index date (PFT test). Do you think that the detection would work as well with an arbitrary index date for this population?

Please discuss why you didn’t use (other) model selection criteria (such as AIC or MDL).

Please mention that AUC corresponds to C-statistics.

Would it be useful to include information similar to Table 3 and 4 for other relevant models also as an online appendix?
Level of interest: An article of importance in its field

Quality of written English: Acceptable

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

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