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 July 2010

Reviewer: P Camp

Reviewer's report:

Overall Comments
This is a well-written paper that addresses an important question, that of the validity of ICD9 codes to identify patients with COPD from administrative health services data. The major strength of this paper is the use of spirometry to ensure the accuracy of the diagnosis of COPD. The use of spirometry as the gold standard for diagnosis is more stringent than other studies which have relied on self-report or physician diagnosis. As such, we expect that the agreement between the ICD9 codes and the spirometry indication of COPD to be less than studies employing less rigorous indicators of COPD, as was demonstrated in this study.

My comments are related more to the completeness of the paper, and the additional information that would both further strengthen this study and answer a few additional key questions. If the authors are able to address these items I believe it would be a very good paper.

Major Compulsory Revisions
1. Age of study participants.
I assume that as the data are collected from VA databases that all the participants are adults. And the authors have dichotomized age in the analysis as > or < 65 years old. However, was there a minimum age in the inclusion/exclusion criteria? COPD with airflow obstruction is really a disease of older individuals with demonstrated airflow obstruction. However, chronic bronchitis is a clinical phenotype of chronic productive cough with no mention of airflow limitation. As such young adults with chronic productive cough could be coded as “491” but not have COPD. Many researchers including ourselves have required a minimum age of 40 or 45 years to prevent including young adults with chronic cough.

2. Many researchers have included ICD490 as one of the codes for COPD. However others, including our lab, have questioned the validity of this code for COPD. Did you consider ICD9 490 in your case definitions? Why or why not? How does this code change the sensitivity and specificity of the case definitions?

3. Could the authors please define the difference between a ‘primary’ code and a ‘secondary’ one? I’m assuming that a secondary code occurs in additional coding fields available. In Canadian studies we often identify patients from the
administrative hospital data if COPD occurs in any one of the 16 diagnostic fields, not just the first one. This is because for COPD exacerbations, the first code is often ‘influenza’ or ‘acute bronchitis’ while the second code is for COPD. We consider these patients to be COPD patients. Could the authors describe the primary and secondary coding issues and the implications of broadening the case definition to include secondary codes.

4. The authors selected the one prior to and one year after the index visit as the window for identifying COPD codes from the database. I’m a bit confused as to why the one year prior would provide strong results. Although it is possible that, for incident cases of COPD, a physician may provide a COPD code before the diagnosis has been confirmed with spirometry, in some cases the spirometry may rule out the diagnosis of COPD. In this situation, you have a COPD code for a visit prior to spirometry, then the spirometry does not show COPD so that would be a discordant case.

I believe it would make more sense to have the year or two after the index/spirometry visit as the window for identifying COPD. Gershon and To used different year windows for their case definitions, such as 1 year, 5 years, or ever. What is the implication of using different time windows after the index spirometry?

5. Although the authors mention it briefly in the limitations, there needs to be more discussion on how the very high prevalence seen in this study affected the positive predictive value. A prevalence of almost 50% is extremely high. I’m assuming this is due to the patient population – older men with substantial smoking exposure. Can the authors simulate what the PPV would be if the prevalence was more in line with population estimates of 10-20%?

Minor Revisions

1. Was tiotropium available in this patient population, and if yes, how would the inclusion of this anticholinergic affect the results?

2. In Table 1, could the authors include the GOLD stage breakdown instead of just mild/moderate/severe etc.

3. Table 1: BMI, for category one it isn’t 0 to 18.5 as your minimum cutoff was 15. Similarly, your 30+ category should be 30-55.

4. It seemed odd that the prevalence of congestive heart failure was the same in the two groups, as it is common comorbidity in COPD. Same comment with depression. Does the author have any comment on this – is it related to the VA population?

5. Figure 1 does not add much additional information and could be deleted. Perhaps a different figure regarding Item 1, next section below would be helpful.

6. Some of the references are typed in inconsistently ie capital letters in all words in some titles but not others.

7. Reference 19 – physician is spelled incorrectly.

8. Reference 10 – Ontario should be capitalized
9. Page 5, second to last line: postbronchodilator spelled incorrectly

Discretionary Revisions

1. It would be very informative to have a table with the frequency of the non-COPD codes that were used in the patients with COPD. If they were not being coded with COPD yet had airflow obstruction, what were they coded with? This is a question that is repeatedly raised – the authors have an opportunity to shed some light on this issue.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.