Reviewer's report

Title: A Cross-sectional Study of the Identification of Prevalent Asthma and Chronic Obstructive Pulmonary Disease among Initiators of Long-Acting Beta-Agonists in Health Insurance Claims Data

Version: 1 Date: 22 April 2013

Reviewer: Lucie Blais

Reviewer's report:

This article describes a study that validated asthma and COPD diagnoses recorded in a US administrative database. The article is well written, and the topic is of interest to researchers in the field of pharmacoepidemiology of asthma and COPD, and more specifically for future studies performed with administrative databases.

Major Compulsory Revisions

1. Please clarify and justify why this validation study was performed only among LABA users instead of among all patients who had a diagnosis of asthma or COPD recorded in the database? Also discuss the impact of performing the study among LABA users on the values of the PPV.

2. Why the authors did not estimate a NPV? If possible, please provide such estimate.

3. I would like the authors to justify why they took only a 6-month period prior to the index date to look for asthma and COPD diagnoses in the database, while I was expecting at least 12 months to cover the potential seasonality of the disease (at least for asthma).

4. In the Cohort Formation section, the authors said that eligible patients were initiators or switchers of LABA or LABA and ICS combinations, but did not define the switchers. Please do so.

5. The authors should describe with more details how they identified medical records with inadequate information, and should discuss the impact of this record selection on the study results.

6. Why the authors did not use regression models to identify the 3 covariates with the largest absolute difference in prevalence across patients with confirmed and unconfirmed asthma and/or COPD. I think that regression models would have provided a more valid answer. Also, clarify why the chosen approach was used among all patients except for those with only a claim for COPD, since the given explanation is not clear.

7. In the 5th paragraph of the discussion, the authors presume about the sensitivity of the claims data without presenting the results. If sensitivity is an important issue and they want to state on the sensitivity of the diagnosis codes on claims they should formally estimate it.
8. The authors should state clearly in the introduction if the aim of the study was to find algorithms able to identify off-label use of LABA, since they refer to the classification of on vs off-label use of LABA in the conclusion of the study.

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I have no competing interest.