Reviewer's report

**Title:** Positive Predictive Value of a Case Definition for Diabetes Mellitus Using Automated Administrative Health Data in Children and Youth Exposed to Antipsychotic Drugs or Control Medications: a Tennessee Medicaid Study

**Version:** 1  **Date:** 21 March 2012

**Reviewer:** Jess Fiedorowicz

**Reviewer's report:**

In the submitted manuscript, Bobo WV et al. estimate the positive predictive value for an algorithm to define diabetes mellitus using administrative claims data in Tennessee Medicaid. The manuscript is well-written and provides important background for planned pharmacoepidemiological study. Strengths of the study include the scope of Tennessee Medicaid data and the clearly defined definition and outlined process for validation and adjudication of outcome.

I appreciate the authors making the effort to publish the background on their outcome validation which will be useful in the interpretation of subsequent work and of interest to others addressing similar problems with observational data. I also appreciated the data on the secondary definition presented in the appendix. Overall, the manuscript is very well-written and organized.

Major Compulsory Revisions:

1) The sample to determine of other diabetes encounters for which records could be adjudicated is quite small -- only 30/113 (27%) samples cases could be adjudicated from other diabetes encounters when 46/64 (72%) of those meeting the case definition could be adjudicated. The authors appropriately provide some explanation for this in the discussion. While greatest interest is in the positive predictive value from this data, there is still interest in the sensitivity of the definition for which this latter sample is needed and somewhat lacking. I suspect the authors did not anticipate this difference would be as great, hence the seeking of records for only 60% of those with other diabetes encounters. Given the low yield, I’d suggest expanding this to 75-80% (140-150 records sought). Based on prior yield, this would be expected to provide an additional 7-10 records for adjudication, increasing this adjudication sample from 30 to 37-40.

Minor Essential Revisions:

1) Abstract, Results. The positive predictive values for Type 1 and Type 2 diabetes appear to be transposed. Also, the Type 2 figure of 83.9 is obtained when combining adjudicated Type 2 and unspecified diabetes mellitus. This is reasonable but should be specified in the abstract as well (this is noted in the results).

2) Apart from the discussion, the authors report their positive predictive values to
the tenth of a percent. This may encourage someone citing these findings to overestimate the precision of their estimates and I’d encourage the authors to drop the decimal place and simply round to the nearest percent throughout the manuscript and tables. In the case of the reported sensitivity of the case definition, this is problematic as the numbers for a and c are rounded to the nearest integer (which is 2 significant figures) prior to calculation of sensitivity, which is then reported to three significant figures. Without the interval rounding, the estimate comes to 64.6% instead of 64.8%. While this difference is meaningless from the numbers reported, it would be preferable to simply as 65%. If we were to construct a 95% confidence interval around the proportions reported using a binomial distribution, it would be clear that these are rough estimates, albeit important estimates to establish for future study. The extra decimal place also unnecessarily clutters the tables.

3) Even though the statistical procedures used for this study were basic, brief description of the analytic methods and software used appears warranted under methods.

4) Results, paragraph 1. Please specify how commonly medical records could not be located here as well.

5) The Figure 1 box “diabetes confirmation within 120 days” is not clear and referral to the appendix could frustrate the reader. Could the authors consider expanding the legend or clarifying the figure to more clearly portray?

Discretionary Revisions:

1) Methods, Sources of Data, Paragraph 2, Sentence 2. Please consider replacing “assured” with a word such as “maximized” or “improved” that implies less certainty in the presence or integrity of data.

2) Methods, Automated Database Definition of Diabetes, last paragraph. Consider replacing glycated hemoglobin with the more common glycosylated hemoglobin.

3) Consider making some mention regarding what portion of the sample had full pharmacy benefits.

4) Overall, the manuscript is remarkably well-written. However the first two paragraphs of the results section are somewhat more difficult to read and could perhaps benefit from even minor editing.

5) In Tables 1 and 2 please indicate somewhere Number (Percent) or include a percent sign in the parentheses. In table 1 Please change the third heading in the first column to read “Adjudicated: Not Incident Diabetes.”

6) Consider moving Appendix Tables 1 and 2 into the manuscript.

Minor Issues Not for Publication:

1) Discussion, paragraph 5, last sentence. Period is missing.
**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests.