Reviewer’s report

Title: Positive predictive value of automated database records for diabetic ketoacidosis (DKA) in children and youth exposed to antipsychotic drugs or control medications: a Tennessee Medicaid study

Version: 1 Date: 16 September 2011

Reviewer: John Seeger

Reviewer’s report:

The authors of this manuscript have a well-defined objective (predictive value of automated identification of DKA) that is feasibly studied and their execution of the study provides a sound answer to the question. The findings of the manuscript will be useful for other investigators who have research questions involving incident DKA in relation to medication use. Comments for improvement:

Minor Essential Revisions

1) On page 4, in the last sentence of the first paragraph, the word “prospective” does not seem correct – I could imagine a “prospective” study (meaning one where the outcome has not yet occurred) conducted within an insurance claims database where cohorts of patients receiving different medications are formed and then followed for the occurrence of DKA, and these PPV results would be useful in such a prospective study.

2) Page 9, second full paragraph, last sentence: it seems too strong to assert that “a valid computer-based DKA endpoint definition is clearly requisite...”. The value of the algorithm is not in required use, but in improving efficiency.

3) Page 8, second paragraph, last line: why the focus on “significant” differences between the full sample and the abstracted cases? Since the sample size is small, significance will be difficult to achieve so that even substantial differences may not be significantly different. Suggest indicating that the full sample and abstracted cases are similar.

4) On page 8, sentence 2, the word “rates” for description of a prevalence is confusing. Perhaps simply indicating that the prevalence of mental health diagnoses is high would be less likely to confuse.

5) The methods are appropriate – predictive value comparing electronic data to medical record data. However, the PPV of such a claims algorithm might vary according to subgroups of patients. For example the PPV could be different for older vs. younger patients or patients without certain comorbidities. Although the overall PPV is most important, to address concerns relating to variability of PPV across identifiable subgroups, the PPV should also be presented in a table that lists its value within some of these subgroups. This table would provide a limited number of subgroups, such as by age (stratify at 10), gender, and ADHD and
provide the PPV within strata defined by the subgroups. Such a table would make a useful addition to the existing manuscript since another investigator may be working in a different population with a different patient mix to which the subgroup PPV may be more applicable.

Discretionary Revisions

1) The question posed by the authors is well defined, although I was left wondering how the algorithm would have performed for DKA in general (including among diabetics) or among adults. Perhaps this can be speculated on in the discussion.

2) Although sensitivity was not explicitly measured, an argument might be advanced that it is likely near 100% if DKA has near uniform likelihood of hospitalization.

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 declare that I have no competing interests.