Reviewer’s report

Title: The Impact of Standardizing the Definition of Visits on the Consistency of Multi-Database Observational Health Research

Version: 2  Date: 30 September 2014

Reviewer: Andrew Williams

Reviewer’s report:

Major Compulsory Revisions
None

Minor Essential Revisions
None

Discretionary Revisions

1. Describe the rationale for the standardized approach for defining inpatient visits, including the merits and drawbacks of alternative approaches. In particular, they might describe why, apart from consistency, the OMOP CDM algorithm is better than a place of service approach and a room and board code approach.

2. Elaborate on their assumptions about the expected degree of consistency in prevalence under perfect measurement. Ideally the results would be interpreted with reference to an empirically-supported quantitative estimate of the expected consistency and a threshold magnitude of insufficient consistency defined with reference to a practical impact on statistical inference. Lacking an empirical basis for these estimates, the authors still might guide readers’ understanding by describing these issues and illuminating them with hypothetical examples.

3. Specify what they mean by “consistent” when they write that the results for “2012 were consistent with the patterns observed in both databases from 2006 through 2011” on line 201. This seems to be contradicted by the variability in the Optum data described on line 209-211.

4. Offer their ideas about what determined whether a prevalence estimate became more consistent, remained essentially unchanged in consistency, or became less consistent across data sources.

5. Clarify whether the findings reported in lines 227-232 and Figure 1 are simply an expected consequence of the way the CDM algorithm excludes inpatient visits as defined in the CCAE raw data and includes inpatient visits as defined in the Optum raw data. I.e. whether this is a finding or a consistency check given the described algorithmic differences in classification approaches.

6. Give more detail in the methods and assumptions of the statistical analyses they performed and their choice of analytic approaches over other options such
as measures of inter-rater agreement like Kappa.

7. Situate the larger project of standardizing and illuminating definitions and assumptions in claims data that this analysis is an instance of within the context of similar efforts attempting to improve the reproducibility of findings, such as the Cancer Intervention and Surveillance Modeling Network (CISNET).

8. Address, in the conclusions section, the fact that the results are somewhat equivocal; that standardization decreased consistency for a non-trivial proportion of conditions.

**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