Reviewer's report

Title: Predictive Performance of Comorbidity Measures from ICD-9 and ICD-10-CA in Population-Based Diabetes Cohorts

Version: 1 Date: 23 June 2012

Reviewer: Roslyn A. Stone

Reviewer's report:

General comments:

The objectives of this paper are to (i) compare the predictive performance of five comorbidity measures prior to and following the introduction of the ICD-19-CA in Canada, and (ii) compare predictive performance for different age groups. The five comorbidity measures are: # of diagnoses, Charlson score, Elixhauser, #drugs, and CDS, with some variations depending on the ICD codes used. Two separate cohorts are defined from what appears to be essentially the same population in Saskatchewan, Canada, based on administrative data during 2 separate time frames. C-statistics and Brier scores are computed for several outcomes over a 2-year follow-up period (i.e., AMI, stroke, LEA, ESRD, hospitalization for any reason, hospitalization for diabetes, and death).

Major Compulsory Revisions:

1. Some sort of table to summarize the definitions of the various comorbidity measures should be included, to show explicitly what component variables are being included in each measure, the source from which they are obtained, and how the measure is defined. The current description is very confusing, particularly the first para. of the statistical analysis, where most of the scores were categorized in a manner that was not explained and 31 dummy variables were defined for the Elixhauser. Standard methods of reporting these measures should be used, no matter what ICD 9 codes were used.

2. The definition of the 2 cohorts appears to be arbitrary. It appears that the same person could be included in both (assuming that they did not die during the first time period), with different index dates according to the time frame over which the data were searched. One could consider identifying a single cohort, and defining the alternative measures involving the ICD-19-CA over time. This still is problematic, however, because the patient condition can (and likely does) change over time, i.e. the definitions are changing and the patients are changing. This confounding by time is an unacknowledged fundamental problem in the current paper. If both the ICD 9 and ICD-19-CA codes are available for the same person at the same time, perhaps the most useful piece of information would be to look at the agreement between the classifications, particularly with the Charlson and the Elixhauser. The point of this paper needs to be thought through logically.
3. Table 4 focuses on c-statistics and Brier scores. Recent work by Colin Begg and others (BMC Med Res Methodol, 2011) has demonstrated that the DeLong test (and related tests) is unacceptably conservative when comparing nested models (as was done in the present paper), and that Wald or likelihood ratio statistics for the extra term(s) in the model should be used. Some of the conclusions of “no difference” in the present paper could be due to this conservatism.

Minor Essential Revisions:

The Elixhauser was left out of Table 1.

The number of patient who were dropped because they did not have “uninterrupted health coverage” during the 2-year follow-up should be reported. This restriction could introduce bias.

The index date does not appear to be technically correct as defined (:Study Cohorts, first para.). The person does not meet eligibility criteria after a single physician visit, so how can that be their index date?

This paper really does not highlight the differences between the ICD 9 and ICD 19 CA, and the title suggests it does.

Discretionary Revisions:
None identified.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.