Reviewer’s report

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

Version: 1 Date: 8 May 2012

Reviewer: Chuan-Fen Liu

Reviewer’s report:

Major Compulsory Revisions:
1. For the comparison of ICD-9 and ICD-10-CA, it is unclear why one would expect the comorbidity measures will be different under the two coding systems, especially for the Charlson and Elixhauser indices, which are classifications of major diseases. Similarly, one would not expect the rates of patient outcomes, which are major complications of diabetes, will change because of the ICD-10 system. The significance and contributions of this study is unclear.

2. It is unclear why these specific comorbidity measures are used in this study. Since the outcomes measures are complications related to diabetes, diabetes-specific measures should be more pertinent these outcomes.

3. The outcomes for hospitalization are extremely high, ranged from 30% to 65% for any hospitalization and from 15% to 30% for diabetes-related hospitalization. Based on Tables 2 and 3, the most common comorbidity conditions are uncomplicated diabetes and uncomplicated hypertension, which do not have the evidence to support these high hospitalization rates. For example, 34% among the age group 20-44 for Cohort 1. Are these numbers correct?

4. One critical issue of this paper is modeling logistics regressions with rare events (<5%) with small sample sizes. The patient outcomes are very rare in the study except for hospitalization rates, which don’t seem to be correct. The logistic regression models may be biased because of outcomes of rare events and the sample sizes are really small. The general rule of thumb for one predictor, there should be 10 events. This is especially problematic for age-specific analyses. For example, there were only 20 AMI events, 12 stroke events, 12 LEA events for the age group 20-24 of cohort 1. See the references below regarding logistics regressions with small sample sizes.


King G and Zeng L. "Logistic Regression in Rare Events Data." Political Analysis 9 (2001): 137-163.

5. Additional information is needed for the use of Brier score to measure model performance. It unclear whether the use of Brier score is appropriate in this context of predicting rare events. Based on the equation to calculate the Brier score in the cited reference, these scores would be extremely small from the
study outcomes because of extremely small probabilities.

6. The authors had detailed discussions about significant changes in c-statistics for specific models. However, according to Tables 4 and 5, none of the change in c-statistics was statistically significant. In addition, the % change in c-statistics is not very meaningful if the c-statistics of the full model itself does not achieve the acceptable range.

Minor Essential Revisions
1. Need additional information on the definitions of AMI, stoke, and ESRD. While the authors cited prior studies, a concise description of the definitions will be helpful.
2. Are Cohorts 1 and 2 overlapping cohorts?
3. How the missing data for the urban variable were handled is unclear.
4. Need additional information how to classified comorbidity into categorical variables (page 10).
5. Appropriate comorbidity measures sufficiently capture the comorbidity of the study cohorts. Tables 2 and 3 should include the proportion of patients who were classified into the any one of the categories.

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

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