Author's response to reviews

Title: Predictive Performance of Comorbidity Measures in Administrative Databases for Diabetes Cohorts

Authors:

Lisa M. Lix (lisa.lix@usask.ca)
Jacqueline Quail (jquail@hqc.sk.ca)
Opeyemi Fadahunsi (fad_91@hotmail.com)
Gary F. Teare (gteare@hqc.sk.ca)

Version: 5 Date: 30 April 2013

Author's response to reviews: see over
Response to Reviewer's Comments
Note that all changes described below are marked in red font in the manuscript.

Reviewer 2
The conclusion sentence in the abstract is awkwardly worded. "Cohort age and the health outcome under investigation, but not the diagnosis coding system…” would be more clear.

This change has been made to the manuscript

The description of the basis for the categorizations in Table 1 is unclear; how can categorizations be based on two different distributions (i.e., the overall and age-specific cohorts)?

We needed to take into account the distributions in both the overall and age-specific cohorts to arrive at a single categorization that would work well for all models. We have modified the wording in Table 1 to make our statement less confusing: "Variable categories were defined from the frequency distributions of the comorbidity measures"

Dollar amounts should be included for the income quintiles in Table 2. Otherwise these quintiles are uninterpretable; it is important to know how wide (or narrow) these intervals are. If the income data are limited, that can be mentioned in the limitations.

We do not have available to us the upper and lower limits of the dollar amounts for the income quintiles. However, we do have the average dollar amounts for each quintile and these are now reported on page 11 of the manuscript.

Table 3 is cluttered and hard to read because the c-statistics, their CIs, and the % changes and % are reported to 3 decimal places. Many of these numbers are quite similar; reporting them to 3 decimal places is not meaningful.

We assume that the reviewer is referring to Tables 5 and 6, where the c-statistics and related statistics are reported. We now report all numbers in both tables to a maximum of two decimal places. As well, the Results section of the manuscript has been updated to reflect these changes.

Multiple imputation is never mentioned in the analysis section, or in the presentation of the results. Was a multiple imputation method used to create a single imputation? If not, how were the multiple imputations for each observation handled in the analysis?

Multiple imputation was not used in the predictive modeling of the data. It was used in the income quintile methodology, as described on page 11 of the manuscript: "A multiple imputation approach was then used to assign income quintile, taking the average of the multiple imputed values, [38] for all DAs that did not have missing information on one or more of these socio-demographic variables and for all individuals who did not have a missing postal code.”

We thank the reviewer for these final comments on the manuscript and hope that it is now acceptable for publication.