Reviewer’s report

Title: Predictive Modeling of Disease Burden in Adult-Onset Ulcerative Colitis Using Health Administrative Data

Version: 0 Date: 28 Nov 2018

Reviewer: Ammar Hassanzadeh Keshteli

Reviewer’s report:

In this study, Murthy et al. used health administrative data to predict disease burden in adult ulcerative colitis patients living in Ontario, Canada. The authors found that a model based on a group of health administrative data could discriminate levels of UC burden at diagnosis. The study is well designed and the results are interesting. However, there are some concerns that need to be addressed by the authors.

It seems that data from all 587 participants were used to generate the models and performance of the models were assessed in these 587 patients. I am wondering why the authors did not split their participants into two sets: a training set to develop the models and a validation set to verify the performance of the generated models. This way they could potentially investigate the generalizability of their models to other populations.

Introduction: The study rationale and objectives should be explained in more details.

It is not clear why the performance of the models to predict disease burden was superior to their performance in predicting UC extent and UC activity. Please discuss about this finding and its importance in more details.

Lines 181 and 182 should be rewritten (as it is not easy to understand in its current format). Apparently, 1-year definitions were more associated with outcomes in comparison to 2- and 3-year definitions. Please discuss about this finding in the discussion section.

Among the 20 variables listed in Table 1, which ones were used in the final models? Were their role in the final models equal? Please clarify.

Please define "disease burden" in the methods section of abstract.

Line 80: linkagelinkage -> linkage

Line 81: toto -> to

Line 101: Physician'sphysician's -> physician's

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.
Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.
I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal