Reviewer’s report

Title: Validation of diagnosis codes to identify side of colon in an electronic health record registry

Version: 0 Date: 27 Dec 2018

Reviewer: Peter Gibbs

Reviewer's report:

There is an increasing focus on the use of real-world data to understand and improve clinical practice and over time these analyses will have increasing impact. Therefore it is essential to demonstrate the accuracy and completeness of these datasets. The authors have generated a random patient sample from the Flatiron Health Database and compared ICD codes with unstructured data in the EHR. This is a simple but important study. The conclusions made by the authors are appropriate.

I am not fully reassured that biases are not introduced when limiting the sample to those patients with site-specific codes, but from the data presented there is no suggestion of this.

Does the result of the analysis change if the transverse colon cancers are included in the right colon category? (as has been done for many clinical series)

How was the random sample of patients generated?

Were there differences between patients with initial stage IV disease and those with an earlier colon cancer at initial presentation? Many in the former group do not undergo primary resection so I would presume less would have an ICD code.

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

No

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

Yes

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 am able to assess the statistics

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