**Author’s response to reviews**

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

**Authors:**
- Patricia Luhn (luhn.patricia@gene.com)
- Deborah Kuk (dkuk@flatiron.com)
- Gillis Carrigan (carrigan.gillis@gene.com)
- Nathan Nussbaum (nathan@flatiron.com)
- Rachael Sorg (rsorg@flatiron.com)
- Rebecca Rohrer (rrohrer@flatiron.com)
- Melisa Tucker (mtucker@flatiron.com)
- Brandon Arnieri (arnieri.brandon@gene.com)
- Michael Taylor (taylor.michael@gene.com)
- Neal Meropol (nmeropol@flatiron.com)

**Version:** 1  **Date:** 24 May 2019

**Author’s response to reviews:**

Reviewer reports:

Peter Gibbs (Reviewer 1): 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)

Indeed, transverse and right colon tumors are commonly grouped together in clinical studies. When we collapse these two groups, the results do not meaningfully change; the results mirror those of the right colon category due to the larger sample size of this group. Table 3 has been updated to include this information (pg 19).

How was the random sample of patients generated?

We have now included additional information on how the random sample was generated in the methods section (pg 6)

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.

We have stratified our analyses by stage at initial diagnosis (stage IV vs. stage I-III). The results observed were generally similar between the two groups, though some measures were lower in the non-metastatic group (stage I-III) compared with the metastatic group (stage IV). A description of the results has been included in the results section (pg 9) and the results are included as supplementary data.

Reviewer 2 (Reviewer 2): PEER REVIEWER ASSESSMENTS:

OBJECTIVE - Full research articles: is there a clear objective that addresses a testable research question(s) (brief or other article types: is there a clear objective)?

Yes - there is a clear objective

DESIGN - Is the current approach (including controls and analysis protocols) appropriate for the objective?

Not sure - key details are missing from the manuscript
EXECUTION - Are the experiments and analyses performed with technical rigor to allow confidence in the results?

No - there are minor issues

INTERPRETATION - Is the current interpretation/discussion of the results reasonable and not overstated?

No - there are minor issues

OVERALL MANUSCRIPT POTENTIAL - Could an appropriately REVISED version of this work represent a technically sound contribution?

Maybe - with major revisions

PEER REVIEWER COMMENTS:

GENERAL COMMENTS: This paper examines the feasibility of using structured diagnostic codes in EHRs to determine tumour location for patients with mCRC. This study is well written and potentially relevant to other researches that may wish to leverage coding in EHRs for secondary uses. As the authors stated, it is indeed important to understand the reliability and completeness of EHR-based variables, although this would not be a novel assertion.

REQUESTED REVISIONS:

My major concern with this study design is the validation sample size of 100 cases. Even though demographic and clinical characteristics seem to be comparable, could these have happened by chance? If another random sample of 100 patients was drawn, would the same characteristics be observed? Did the authors draw any other samples before selecting this specific one? Ideally if checks were not carried out manually, a larger sample size (thousands) would be preferred.

The abstraction process is carried out manually, making a sample size of thousands not practical. However, we have included an additional 100 patients in the revised manuscript and the results did not change meaningfully, suggesting that the results are robust. We did not draw any other samples before selecting the original or the additional sample. Additional information on how the random sample of patients was identified has now been included in the methods section (pg 6).
Still related to the above comment, in Table 1, could the authors also compare the age distribution across both groups using averages with standard deviations (another metric of skewness could also be helpful)?

The median/mean and IQR/SD are now included in both Table 1 (pg 17) and Table 4 (pg 19) and show that there is no difference in age between the sample (n=200) and the parent cohort, nor if you compare patients who have the specific ICD codes in the EHR to those who do not in the parent cohort.

Secondly, could the authors please expand the discussion section to focus on the reproducibility of this work both in statistical terms (i.e. validation sample size, already mentioned above) but also by describing coding practices at their institution and how this may differ across different institutions/countries. For instance, if the primary purpose of coding diagnoses is billing, would there be any particular bias towards specific codes if they have similar costs or if some are considerably more expensive? A discussion on the origins and purpose of this data would be most helpful - this could also be done in the introduction section.

Flatiron is a network of oncology practices across the US, not a single institution, and thus there may be differing coding practices based on the center. To our knowledge, there is no published literature to suggest that there are differences in cost associated with the treatment of colon cancers by site. However, it is important to acknowledge this potential limitation. In addition, we acknowledge that the results from this study may not be generalizable outside of the US, where billing and coding practices may differ. The discussion section has been expanded to include these points (pg 10).

ADDITIONAL REQUESTS/SUGGESTIONS:

None