**Reviewer’s report**

**Title:** An ontological approach to identifying cases of chronic kidney disease from routine primary care data: a cross-sectional study

**Version:** 1  **Date:** 14 Dec 2017

**Reviewer:** Shuchi Anand

**Reviewer’s report:**

Please include all comments for the authors in this box rather than uploading your report as an attachment. Please only upload as attachments annotated versions of manuscripts, graphs, supporting materials or other aspects of your report which cannot be included in a text format.

Please overwrite this text when adding your comments to the authors.

Cole et al. are attempting to identify patients with CKD using existing primary care data from the U.K., and to see if an ontological approach allows for more specific/sensitive capture of patients with CKD, particularly early stage CKD. Overall I find the manuscript is very well-written with the authors' rationale and approach very well described. However, I find that the ontological approach also has similar weakness as in previous analyses attempting to quantify CKD in primary care, rather than community based, data sets: 1, underdetection of early stage CKD is likely as proteinuria or albuminuria and underdetection of presence of albuminuria and proteinuria in patients with any CKD since these markers are not being routinely measured in primary care settings; and 2. underdetection of CKD in otherwise healthy people either not presenting to PCPs or not being recommended for kidney function testing. Comments by section:

**Introduction:**

--very well written, could be shortened a bit, especially the "importance of CKD" section since this is a nephrology audience well aware of implications of CKD diagnosis

--(lines 76, first line in identifying CKD using primary care data) I am not sure primary care data can be used to derive

**Methods:**

--again very well written with clear description of rationale

--I would push the authors to make stronger the point that Step 1 in the ontology approach (i.e., "domain ontology for CKD") is replicable to other countries and health systems (perhaps make it again in the ontology and coding section)
--in the same section, what were the criteria used to exclude/include codes from the 1250 candidate codes? this is important for replication as well (clearly the involvement of nephrologists and clinical informatics experts means at this step means that it cannot be automated). A supplementary appendix listing the codes used would potentially be useful as well.

Results:

Table 1 is not necessary for the audience of nephrologists.

Table 2 I advise taking out the studies that are community based (I believe Roderick and Fraser et al). since the population age structure and overall study design is vastly different.

Table 3 I think presents the results very clearly. Just to clarify: the patients with a CKD read code but without supporting two eGFRs or two proteinuria measurements, all of them fell either in the CKD stage 2 or in the RRT category? In the discussion, the authors say that 26% of individuals with CKD read code did not have supporting lab evidence: it is not clear to me where these patients fit within Table 3 or Figure 3.

In individuals identified in early stage CKD by CKD read code, do the authors have any way of validating that these patients indeed do have early stage CKD (e.g., nephrology visits or at least a single measure of urine consistent with proteinuria?)

Discussion:

Lays out the advantages and most limitations nicely. I do believe the authors could strengthen their analysis by attempting to validate the CKD read codes in a subset of patients without supporting eGFR or proteinuria evidence. They are absolutely correct that the proteinuria alone stage is hardest to detect, and if the authors are able to accurately capture more patients with their approach, this is useful. Also the authors could emphasize here that while their approach is not suitable to create estimates of CKD prevalence, it may still be relevant for clinical studies evaluating outcomes by certain clinical practice variations. For example, if we want to study the real world implications of tighter vs less strict BP control in a clinical population with CKD.

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 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