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: 20 Nov 2017

Reviewer: Lili Chan

Reviewer’s report:

Cole et al. utilized Ontology Web Language to identify CKD patients and compared this method with traditional laboratory values (eGFR and proteinuria) for identification. 1.6% of patients without laboratory findings were identified to have a clinical code consistent with CKD. I believe that this topic is novel to nephrology and of general interest given the increasing number of researchers interested in utilizing EHR as data. I did identify several areas that needed clarification.

Minor:

* Page 4 Line 101: "In the main" would suggest change to "In general", or "for the most part", or "a majority of studies"

* Page 4: I would be careful with eGFR and GFR, as they aren't necessary interchangeable.

* Page 4 Line 105-107: eGFR from a kidney transplant would be accurate.

* Page 6: consider also including proteinuria definition in mg/g for US audience.

* Page 8 line 212, the percentage should be next to 48,681 not 78,153.

* Page 8 line 214, the percentage should be after the 51,526.

* The submission states that the manuscript included corrections as advised. I wasn't provided with the original version or a document with tracked changes. Not sure if it is essential.

Major:

* Can the researchers clarify if ICD codes were also available for the associated visits? If so, how did this compare to the Read method?

* Page 8, Line 206: unclear where the percentages for the CKD stages with proteinuria are calculated from?
* Question regarding the utility of identifying CKD stage 1/2 given the low likelihood of progression. What were the common Read codes that actually identified CKD stage ½?

* I would also note that that there is no way that the creat based method could identif CKD stage 1/2 as by definition the eGFR is >60.

* For the proteinuria method, I would be interested in knowing what percentage of people actually had proteinuria checked, and what proportion of those patients had proteinuria.

* Comment on the ability to manually review a random subset of charts? If this is possible I suggest performing to give an idea of the presence of terms in the 26% of individuals with CKD read code without lab evidence.

* It is unfortunately that the researchers developed an ontological method of identifying CKD, however the method will not be useable for future research as NHS is being transitioned to SNOMED CT.

* I would include that only 25% of the notes are in Read as a limitation. Could this also partly be the reason that their prevalence of CKD is lower than the Health Survey for England.

* The researchers highlight the ability to identify ESRD receiving RRT and transplant patients using their method. In the U.S. we can link most databases to the USRDs which significantly lowers the utility of identifying ESRD on RRT using Read codes. I am not clear if there is an equivalent in other countries.

**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.

Not relevant to this manuscript

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 have nothing to disclose.

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