Reviewer’s report

Title: ‘Caveat emptor’: the cautionary tale of endocarditis and the potential pitfalls of clinical coding data: an electronic health records study

Version: 1 Date: 17 May 2019

Reviewer: Martin Thornhill

Reviewer's report:

Thank you for asking me to review this revised manuscript. The authors have clearly made a number of changes and have attempted to address some of the issues raised by both reviewers - generally improving the paper.

I have, however, had great difficulty evaluating the changes made in response to my comments. This is because, although they have attempted to address my comments one by one, the line number references they give for many of the changes do not seem to tally with the line numbers in the amended and tracked manuscript. Indeed, on a number of occasions their line references refer to deleted text of blank lines. Even when they refer to text, it is often unamended text or the changes they say they have made are not present. Notably, however, most of the references to the abstract line numbers appear correct. I wonder if the authors made further changes to the manuscript after writing their response to the reviewer's comments. Thereby, invalidating the line number references used.

In view of this I have not been able to properly re-review the paper but would be happy to do so if the authors can correct this problem. It would also aid the review process enormously to have a clean version of the manuscript available (without tracked changes).

To save time, however, I have noted a few other items while reading the manuscript:

1. The Title. I made clear my concern about the implications, pejorative tone and sensationalism of this title when I first reviewed this paper, but the authors have chosen to ignore this. It might be suitable for a novel or newspaper article headline. It might even be suitable for an opinion piece in a journal, but this study is trying to sell itself as a piece of science. The difficulty before was that the authors didn't seem able to decide if they were writing an opinion piece or reporting a scientific study. They do now seem to have shifted more towards publishing a scientific study, which given the data would seem where it most appropriately sits. However, the title still doesn't seem appropriate for that. As it stands, this title casts doubt on the appropriateness and value of all electronic records studies and implies that the reader of any such study is being sold a lie. In reality, if such studies are done appropriately, with proper choice of codes etc, and the limitations of the study are properly described, this is not the case and such studies are valuable contributions to our knowledge on the subject - as is this study.

Indeed, the real value of this study is to provide the data to enable researchers to design better studies and make better code choices in the future.
2. With all the changes to the manuscript it looks like some of the referencing may have been inadvertently changed - and therefore needs checking. The one that was most obvious to me was the use of reference 10 on line 589 (p29). Here you exonerate most studies from using the I38 code, particularly in the secondary position but note that two studies did make this error, at which point you cite reference 10 the Dayer et al Lancet paper. This paper did not use I38 at all to identify it's cases.

3. In my last response, I pointed out that the Dayer et al paper had used only I33.0 primary codes, had used superspells to avoid counting transfers more than once and had excluded elective admissions. I suggested that the nearest specificity and PPV values for the Dayer study, based on your data were likely therefore 0.97 and 0.88 respectively. The purpose of this was to point out that we had not been naïve in our choice of codes etc for the study, as was implied, but had put a lot of thought into it. Furthermore, we did not expect you to calculate these values for the Dayer study, as you suggest. Indeed, we did not wish to make a big issue of this we just wanted to point out that to lump the Dayer study, and similar studies, in with studies using a broad range of ICD-10 IE codes both primary and secondary, including I38, was not reasonable or fair.

We note, however, that in your response and in the changes to the text you have gone into the detail of waiting list and non-waiting list elective admissions, which seems unnecessary and of little interest to readers. You have also been at pains to point out that the Dayer study did not exclude readmissions and short stays, which was a factor in the 0.97 sensitivity and 0.88 PPV. In fact, we did evaluate both short admissions and readmissions. We performed sensitivity analyses on our data excluding all admissions shorter than 2 weeks, 1 week or 3 days. We also performed sensitivity analyses on all re-admissions, readmissions within 6 months (the criteria many use to distinguish between new IE cases and re-admission for the same episode of IE) and readmissions within 30 days. However, none of these altered the outcome of our study and our reviewers, therefore, felt it was unnecessary to include these analyses in the paper. Hence, I feel that the values cited above probably most accurately represent those that applied to the Dayer study.

I really don't think there is any value, however, going into this level of detail about our study or indeed, the level of detail you have already. I really don't see the point therefore of pedantic comments like "which also excluded "waiting list" elective admissions." In Lines 600-602. "PPV was only slightly higher, but sensitivity substantially lower using the I33.0 code in the primary position alone, as used in the previous study of English HES data which also excluded "waiting list" elective admissions." Such comments, that occur in several places in the paper, may be factually correct but are unnecessary, add nothing to the paper, will be meaningless to most readers and sound like you are trying to make some kind of pejorative political point.

4. In my comments, I pointed out that the aims of the study are important in determining which ICD-10 codes etc are most appropriate, and whether sensitivity, specificity, PPV or NPV should be prioritised. You responded by rejecting this and pointing out that, "As demonstrated by a ROC curve, 100% specificity is achieved at 0% sensitivity (and vice versa). As stated in our overall response above, an important of our study was to investigate how choices of different codes and exclusions affected these overall". However, this doesn't address the point I was trying to make. What I was trying to point out was that the aims/objectives of the study influence the trade-offs one is prepared to make and therefore the codes that are most appropriate to use. If the purpose of the study is to recruit every possible individual who may have had IE, regardless of the accuracy of that diagnosis (and we have recently performed such an electronic records study), then one would want to prioritise sensitivity over specificity. And it would be appropriate to use a much broader spectrum of ICD-10 codes, possibly
including secondary codes. In other studies, it may be important to, as closely as possible, identify just those IE cases that are Duke criteria positive, then it may be more appropriate to prioritise specificity over sensitivity - in which case a very different set of codes would be appropriate focusing more on primary I33.0.

Although you have been dismissive of this point, I still feel it is important. Indeed, you implicitly acknowledge this in lines 657-660 "However, studies that aim to maximise inclusion of possible cases should not automatically disregard them, as a substantial proportion of confirmed clinical cases may only receive a secondary code, as in our endocarditis examples."

In lines 593-596, you say " Whilst maximising PPV may appear attractive, a very strict rule can achieve 100% PPV whilst missing most true cases, underestimating incidence and with an uncertain impact on incidence trends. Overall we consider that using all codes except I38 secondary provides a good balance between PPV and sensitivity". On the basis of your data, the last sentence may well be true. But it doesn't mean that all studies should use those criteria. They may be appropriate for some studies but totally inappropriate for others depending on the specific aims and objectives of the study. For certain studies, maximizing PPV, even if true cases are lost, may be the most appropriate strategy. What is important is that studies make their aims clear, state what codes etc. they used and explain the limitations of their approach so that readers are fully informed and have the information with which to evaluate the appropriateness of the approach and value of the results.

I maintain, therefore, that there is no one approach that is best, the approach that is most appropriate will depend on the aims/objectives of the study.

5. In lines 108-109 you talk about the Dayer study and the more recent study on the impact of the AHA guidelines that was published in JACC. You then go on in lines 110-111 to say, "Their findings have prompted some to advocate that changes in the dental prophylaxis guidance should be reversed." But this isn't true and overstates the facts. While some may have used the Dayer data to advocate that the 2008 NICE guidelines are revised, because the data suggests the possibility of an increase in IE cases since they were introduced, the data from the JACC paper is highly supportive of the AHA guidelines. I am unaware that anyone has used the JACC data to suggest that the AHA guidelines should be revised - let alone reversed.

You go on to say in lines 111-112 "it is therefore important that the validity and accuracy of the data on which these studies are based is assessed." In your response to the reviewers comments, however, you were at pains to point out that: "Our primary aims were to use the case study of endocarditis to explore the relationship between clinical cases and diagnostic codes, quantify discrepancies, and identify reasons for these discrepancies, to enable better design of electronic health records studies in the future." Which seems entirely appropriate. And you were also at pains to point out that it was not the aim of the paper to evaluate the Dayer study or call into question it's findings. This doesn't seem to be reflected in what you say in the manuscript, however, about assessing the validity and accuracy of the data in these studies (see above). So, which is it? Is your paper attempting to evaluate the relationship between clinical cases and diagnostic codes or is its main purpose to assess the validity of the Dayer and JACC papers (references 10 and 20)? The former would seem more scientific and less political.

6. I think we are agree, that the absence of specific ICD-9/10 codes for oral viridans group streptococci (OVGS) means that it is not possible to accurately perform studies of the part played by these organisms in IE using electronic health record data. This is a pity since OVGS IE is what we would really like to study when it comes to investigations into the use of AP to prevent IE in those
undergoing invasive dental procedures. Because your study does provide data about streptococcal IE (for which codes do exist), but which incorporates OVGS, I think it important that you make this distinction clearer than is currently the case. It is particularly important to make clear to those less knowledgeable about IE bacteriology, that the data on Streptococcal IE is not equivalent to OVGS IE and does not necessarily identify those cases for which AP is relevant.

7. You say that there is reasonable agreement between organism codes and microbiology and that they can therefore be used to look at changes in the proportions of specific organisms out of all organisms over time, and point out the reviewer 2 also suggested this. However, that is not necessarily the case. Although we gathered causal organism data for the Dayer study, one of the many problems we experienced with the data was that changes in policy and coding practice affected the proportions of some organisms and not others. For instance, when there was a big focus on the problem of MRSA in hospitals and they were under a lot of pressure to record and deal with this, there was an obvious and significant increase in the recording of Staphylococcal supplementary codes nationally that was not seen for other organisms. I remain unconvinced, therefore, that coding data can be relied upon to accurately track the proportion of certain organisms among all organisms. There are big issues about relying on supplementary/secondary organism codes for identifying epidemiological aspects of IE and other disease pathology. These include, lack of codes for important organisms, the fact that such codes are only recorded for a proportion of cases and that proportion changes over time due to political imperatives, administrative, resourcing and other factors (and varies enormously between hospitals) completely unrelated to epidemiological or disease related factor. Such factors can even result in the changes affecting certain species and not others. Furthermore, the use of supplemental codes means that organisms identified may have no relationship to the disease being studied but could instead relate to intercurrent infections and other pathologies and co-morbidities. Although individually, each may only have a small effect, together they have a large effect on the reliability of causal organism data.

You have made a big point in this paper of highlighting the inadequacies and limitations of coding data for identifying cases of IE. It is important, therefore, that you also highlight the even greater limitations of using coding data in electronic health record studies to identify causal organism data. All the limitations mentioned above deserve as much, if not more prominence in your discussion of the limitations of using coding data in electronic health record studies. Looking at causal organism data, is after all, plagued by both problems i.e. the limitations of data in defining IE cases plus the limitations of supplemental codes to identify the causal organisms of these cases.

What is needed, is a plea that future iterations of the ICD coding include a revision and expansion of the coding of causal organisms and particularly for IE it would help if there were specific diagnostic, rather than supplemental codes, to identify IE caused by specific organisms.

I still think this important data and the aim stated in the response to the reviewers i.e. "Our primary aims were to use the case study of endocarditis to explore the relationship between clinical cases and diagnostic codes, quantify discrepancies, and identify reasons for these discrepancies, to enable better design of electronic health records studies in the future.", is where it needs to be targeted. The previous version seemed to be focused on making all sorts of points unrelated to the aims of the study. This version is considerably better from that perspective, but it is still not entirely focused on those aims. Indeed, it would help if those aims were more clearly identified in the 'Background' to the paper.
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.

Not applicable

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

No

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

**Statement on potential review bias**

Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?

If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria'. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria

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/](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