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: 0 Date: 07 Feb 2019

Reviewer: Martin Thornhill

Reviewer's report:

From the outset, I should declare that I am one of the authors of the Dayer et al paper cited several times in this study.

This study into the accuracy of ICD-10 codes for identifying case of infective endocarditis (IE) has been well done and contains some interesting data. Unfortunately, the importance of the findings and the positive value that it could contribute to research in this area has been submerged beneath negative and highly selective reporting of the data to make some preconceived pejorative value judgement, not supported by the reported data, about the work of others in the field.

Infective endocarditis (IE) is a rare condition. This makes it difficult to identify and recruit cases to clinical trials. In particular, the numbers of individuals that need to be recruited to achieve an adequate sample size for randomised controlled clinical trials of preventative measures such as antibiotic prophylaxis (AP) is so large as to be prohibitive both organisationally, and for grant funding bodies. This means that funding bodies have re-directed researchers to large observational studies using national data sets in order to study some of these things - so called 'big data' studies.

'Big data' studies have the advantage of size but also numerous disadvantages which are well known, some of which are mentioned in this paper and which have also been elucidated in the various 'big data' studies cited in this paper. Virtually all 'big data' studies depend on administrative coding data, mainly ICD-9 and -10 codes to identify cases and the issues around using ICD codes for identifying cases are well known and have been explored in a number of other studies, including the studies cited in this paper. With regard to IE, we are fortunate that there is a relatively limited number of codes that describe, or might be associated, with this condition - compared to some other conditions. Some are relatively precise descriptors for IE and others much less so. Much work has already been done by researchers in trying to determine which are the most precise descriptors and which are less so. Much of this work has been theoretical, often taking place when researchers design the protocol of their study (including most of those 'big data' studies described in the first part of Table S1 and several others not described there), while other studies (including those described in the 2nd part of Table S1, and some not described there) have tried to validate codes against actual patients - usually using the Duke criteria to define cases. And this study falls into that category.

ICD codes for IE can be used for different things in different studies. The aims of the study are therefore as important in determining which codes are appropriate to use as is the accuracy with which
codes identify cases. For example, if the aim is to screen a population to identify as many IE patients as possible for recruitment into a clinical trial where each patient will then be evaluated individually to see if they meet the study inclusion and exclusion criteria, then you want codes that cast the widest possible net i.e. you want high sensitivity for detecting cases but may be happy to sacrifice specificity. If, on the other hand, you were working with a big data set and the aim is to just focus on true (Duke +) IE cases as far as possible. Then you might wish to prioritise specificity over sensitivity, even if that means you lose some cases, in the knowledge that you would non-the-less still have a sufficiently big sample size to detect a clinically significant change. This was the case with the Dayer study that you cite. Furthermore, as you have pointed out, because different healthcare systems collect coding data in different ways and for different purposes, the codes that provide the greatest specificity (or sensitivity) in one country/healthcare system may be different in another setting. We have found significant differences in the coding practices, and therefore most appropriate codes to use, in the US and UK. Hence, the different codes used in our US (J Am Coll Cardiol 2018;72:2443-54) and UK studies (your reference 3). Incidentally, when you cite all the 'big data' studies that have looked at the impact of the AHA/ESC guidelines on IE incidence (line 83), why do you fail to cite the largest and most recent of these studies (J Am Coll Cardiol 2018;72:2443-54), even though you include it in your list of such studies in Table S1?

It is naïve, therefore, to accuse other researchers of being naïve. It is also wrong to condemn them of "assuming that a diagnostic code represents a clinical case" of not being careful about their choice of codes and of being unaware and naïve about the pitfalls of clinical coding data in clinical data studies. That is a preconceived, prejudiced and naïve criticism of fellow researchers, that is unsupported by evidence, and not the sort of thing that should be part of a scientific paper. Most of the studies cited have been very careful and thoughtful about which codes they have chosen for their studies and justified their choices in their paper. Indeed, in most cases your data has validated, rather than discredited, their choices. Yet the authors of this paper have bizarrely chosen to spin their data to disparage other studies rather than highlighting the true importance of their own findings.

Because the Dayer paper was a UK study and their ICD-10 code validation study is also a UK study, the authors have framed their whole paper as a critique of the coding choices made in the Dayer paper. The Dayer paper used only I33.0 primary codes in order to maximise the likelihood that identified cases were true Duke + cases of IE. Furthermore, a strategy was used to include only supersells. This was done in order to ensure that cases transferred from one hospital to another or who were re-admitted for the same episode of IE were counted only once and elective admissions were excluded. In support of this, this study found that a primary I33.0 code, excluding elective and readmissions, had the highest specificity (0.97) and PPV (0.88) of any of the ICD-10 codes, or code combinations studied. In other words, this study completely validates the choice of codes made in the Dayer paper - where the aim was to maximise selection of Duke + IE cases, even if that meant losing some cases i.e. sacrificing sensitivity. However, the authors have chosen not to mention this anywhere in the papers text and to hide it away in Supplementary Material File 11 Table S6. Instead when commenting about the Dayer paper, they chose to focus only on the sensitivity value saying: "We also considered using the most reliable I33.0 code in the primary position only, similar to previous publications using UK data3,30, although this had poor sensitivity in Leeds (0.43) and underestimated incidence of clinical cases (Figure 3)." Even though our study design was not aimed at achieving high sensitivity. Rather it was aimed at achieving high specificity/PPV, and according to your data, it achieved the highest specificity/PPV possible. This is an example of selective use of your data and spinning it to discredit another study, when in fact your data strongly validates the approach that was taken.

In the following paragraph, you say "Estimating incidence using the steps outlined above (removing
codes with low predictive power, short stays and readmissions) substantially improved agreement between estimated and true incidence of endocarditis, although it similarly tended to overestimate incidence increases and suggest stronger statistical evidence to support them (Figure 3)." The data you provide, along with Figure 3 just don't support this. The graph 2nd from the left in Figure 3, most accurately reflects the description in parenthesis above and the protocol used in the Dayer paper. If this graph is directly superimposed on the 'Clinician-assessed Incidence' graph (far right, figure 3), the two trend lines are almost precisely parallel - therefore the trends are identical, not overestimated, using this approach. i.e. this coding strategy replicated that actual trend with very high precision. And, far from overestimating IE incidence, it underestimates it (the trend line is parallel to but lower than the Clinical-assessed trend line). And this means there is reduced statistical power to show a change - it does not suggest stronger statistical evidence than should be the case - as implied by the wording in the papert. Again this data supports the strategy we used in the Dayer paper, where we deliberately chose a strategy where we could be confident that we were doing the best we could to identify real IE cases (hence high specificity and PPV), even if that was at the cost of missing some IE cases (lower sensitivity) and reduced statistical power, because we wanted to minimise cases that might not be real IE cases and we knew we had a big enough sample size and the statistical power to detect a significant change, even with the potential loss of some less well defined IE cases.

Similarly, your study uses the broadest list of possible ICD-10 codes I have ever seen for identifying IE cases (I33.0, I33.9, 1, I38.0, I39.0, I39.8, I01.1, I09.1, I42.3, B37.6, T82.6), much broader than in any of the studies you cite and criticise. Of the ICD-10 code studies you cite, most use I33.0 primary codes only and one uses I33.0 and I33.9 primary codes only. Of the ICD-9 code studies you cite, most use the ICD-9 equivalents of I33.0, I33.9 or I39.0 and none used the ICD-9 equivalents of I01.1, I09.1, I42.3 or T82.6. Throughout the paper, however, when you talk about 'codes' and provide data that you compare with previous studies, your data is based on using all of the codes listed above. Yet none of the other studies used such an extensive list of codes. Far from it in most cases. So, when you talk about 'codes' and provide data you are not comparing like with like. Indeed, according to Supplementary data File 11, Table 6, this produces a very low specificity (0.47) and PPV (0.44) but improved sensitivity (0.74) and NPV (0.76). Although poor specificity and improved sensitivity may be acceptable if you are screening for IE cases for a clinical trial and will then check that each individual identified really is Duke + before recruiting them into the trial, it is not at all acceptable if you are looking in a 'big data' study at the impact of guideline change on the incidence of IE, as all the cited studies were. In this situation, you want to ensure, as far as possible, that you only identify individuals with true (Duke+) IE. Therefore, you want to prioritise specificity and PPV over sensitivity. The ridiculously large panel of ICD-10 codes chosen for this study was appallingly bad at identifying IE cases - with a specificity worse than tossing a coin (0.47) and bore no relationship to the much more restricted code panels chosen by the studies cited in the paper. It is completely misleading therefore to suggest that the cited studies suffered from the sort of poor specificity that the panel of codes chosen for this study did.

Despite this, the first thing you say in the results is "Less than half of admissions with an endocarditis code recorded in electronic health records represented a new clinical case of endocarditis". You go on to say "In Leeds 738/1681 (44%) endocarditis-coded admissions between 2006-2016 represented Duke definite/possible cases (Figure). In Oxford, 307/552 (56%) reviewed admissions between 2010-2016 represented Duke definite/possible cases." This data cannot be found in any of the main paper figures or after carefully working through all 19 additional supplementary data files. While I don't doubt it, these figures are the result of using a bizarrely broad set of ICD-10 codes for IE and therefore bear no relationship to the number of Duke definite/possible cases that would have been identified by the codes used in any of the studies cited and criticised in the paper. Yet, these figures are used in the abstract and elsewhere as if they are a true representation of the likely inability of any 'big data' studies to identify
Duke + cases. This is wholly misleading and denigrating of other studies. Immediately after citing this data in the abstract you say: "Estimating endocarditis incidence using admissions with any diagnostic code overestimated incidence trends two-fold." And conclude with "Studies conducted using health records that assume a diagnostic code represents a clinical case, without examining the predictive ability of the diagnostic codes used, can give inaccurate estimations of incidence and trends." This is clearly an attempt to impugn most of the 'big data' studies cited in the paper and imply that the codes they have used result in a massive overestimation of IE incidence. When this is clearly not the case. Indeed, as you later point out in the results section, use of I33.0 (as used in several studies) will "underestimate incidence of clinical cases." But again, you write this in a pejorative way about the study. So, on the one hand you criticise studies for overestimating IE incidence and on the other you criticise them for underestimating incidence. Yet, as discussed above, having a wide net for capturing cases may be appropriate for some studies and being as precise as possible in identifying Duke+ cases, even at the loss of some cases, may be appropriate in other situations, particularly if you have a big enough sample size to accommodate that loss. Criticising these studies as being naïve is, therefore naïve in itself.

The same criticism is repeated but inverted in the results when you state: "Raw admissions data can give massively inflated estimates of incidence" and "Estimating cases of IE using all admission with an endocarditis code overestimated the apparent incidence in Leeds during 2006-2016 by over twofold compared to clinical cases in the Leeds service database (Figure 3) (sensitivity/PPV 0.74/0.44). Why use 'an' when it should be 'any' (and better still the codes should be listed) and why just show sensitivity and PPV? Again, in the discussion etc the implication is there that this is the sort of sensitivity/PPV etc of the data being collected in the 'big data' studies despite the fact that none of them used this coding protocol, or anything like it. Indeed, the graph 2nd to the right in Figure 3 uses a coding protocol not dissimilar to some of the ICD-9 studies cited in the paper and the resulting plot is very similar indeed to that of the clinically determined cases (graph, far right, Figure 3). i.e. this study strongly validates the methods used in these studies rather than undermining them. So, the critical and pejorative narrative concerning these studies, that runs throughout the paper seems entirely unreasonable naïve, and biased.

You then go on to talk about the use of the I33.0 primary code citing our papers. However, you only talk about this having poor sensitivity (0.43) and underestimating the incidence of clinical cases. You say nothing about the fact that is the most effective strategy for identifying Duke + cases with a specificity of 0.97 and PPV of 0.88. Again, your data validates the coding protocol we chose for those specific studies. Yet the tone of the paper is entirely negative and disparaging about these studies.

The data on identifying the causal organism for IE cases using supplementary ICD-10 codes, essentially found that coding was not sufficiently accurate or reliable and changed over time due to changes in coding practice. Thus, you concluded it was not reliable enough to depend on for quantifying Streptococcal IE trends. This is exactly the same conclusion we drew in the Dayer et al paper, and, as we elaborated, was the reason we did not include causal organism data in our study. However, this paper only seeks things to criticise in 'big data' studies not things to agree on or support.

It is also hard to see why Figures 1, 3, 4 and 5 all show time course data when nearly all of the reported data and sensitivity/specificity/PPV/NPV values are based on total numbers at each site for the period covered by the study. Presumably, the purpose of collecting time course data was to see if the 2008 NICE guidelines resulted in a change in the incidence of IE? It is interesting therefore that, the data on clinically assessed incidence (Figure 3) shows IE incidence falling before 2008 and rising in the period after. Furthermore, this pattern is reflected in the IE incidence data with all 3 ICD-10 coding protocols shown in Figure 3. This pattern is also seen in the Figure 1 data for Leeds, both in the ICD-10 and clinical cases data. The Oxford ICD-10 data covers a larger time frame and the pre-2008 data shows a
flat or shallow downward trajectory but from 2008 onwards shows a significant upward trajectory. Unfortunately, the Oxford clinical data for this period is not available for comparison. Bizarrely, therefore, given that the whole paper seems to be devoted to the issue of what effect guideline change had on IE incidence, the authors provide no analysis designed to identify if there was a change in IE incidence trends comparing the periods before and after 2008 and no comment on this. Why did they not perform an interrupted time series analysis, change point analysis, poison model analyses comparing projected post-change incidence data modelled from pre-change data with actual post change data, autoregressive integrated moving average (ARIMA) analysis or another scientifically and statistically accepted analysis to compare the IE incidence before and after the guideline change? Without this, the trend data shown in Figures 1 and 3 is pointless other than to show a general upward trend in IE incidence and how accurately the incidence trajectory of clinical case is tracked by I33.0 - which does not require so many figures.

Again, the time course causal organism data tells us nothing since, the sample size for each organism type was too small to draw any conclusions - particularly when analysed at an individual year level rather than for the entire period of study. Furthermore, the % of cases for which an organism code was recorded was only around 57% and 65% in Leeds and even less in Oxford i.e. between 18% and 60%. As mentioned in the paper, changes in coding practice changed substantially during the study. These findings very closely mirror what we reported in the Dayer et al paper i.e. "Secondary or supplementary coding was unreliable and relevant codes were recorded in only 30-49% of cases (and we cannot be certain that these represented a random subset of the entire population). Additionally, the rate of improvement in secondary or supplementary coding was uneven" i.e. your data further validates our findings. However, you fail to mention this, or two other well recognized problems related to the use of ICD-9/10 codes for identifying the causal organism for IE cases. First, because the codes are secondary or supplemental codes that are assigned at discharge, you cannot know if they refer to the cause of the IE or to another intercurrent infection such as a chest infection, bed sore, urinary infection, post-op wound infection etc. Particularly, as these are very ill and debilitated inpatients, many with co-morbidities and a high proportion of which undergo surgical intervention during their hospital stay. i.e. the presence of a streptococcal or staphylococcal code does not mean that those organisms were necessarily the cause of the IE. Particularly when the codes are for "sepsis" or "diseases classified elsewhere" and none are for IE. Furthermore, since you have framed the whole paper around the issue of antibiotic prophylaxis following invasive dental procedures (this is what the 2008 NICE guidelines recommended should stop), the relevant organisms are what are defined by most as oral viridans group streptococci (OVGS) i.e. those organisms that you defined in Supplemental data, Additional file 17, Table S9, and not other Streptococci. Yet the ICD-9/10 codes you have used are not specific for OVGS but cover a broad spectrum of other streptococcal species. Indeed, most studies that have attempted to use ICD-9/10 codes to look at OVGS related IE, have designed protocols to try and get nearer to identifying OVGS species by subtracting those codes that identify non-OVGS streptococcal species from all streptococcal codes. However, this is not a satisfactory approach either since it means identifying OVGS species negatively rather than positively. Overall, therefore, the organism-specific data is underpowered and too unreliable to draw any conclusions - and this needs to be stated clearly, or the data should not be used.

These points have been made previously in several other studies including the Dayer et al paper: "there are no pathogen-specific ICD-10 codes that identify oral viridans group streptococci. Furthermore, we could not always be certain that the organism coded was the organism that caused the infective endocarditis and not an organism that caused some other intercurrent infection—e.g. a chest or wound infection. Finally, because of the small amount of data for each type of organism, the study was underpowered to detect a significant change. In view of these limitations, it was impossible to draw any
conclusions from the organism-specific data." Your study again, therefore, validates the Dayer et al study, and similar studies, and yet you misconstrue your data to suggest otherwise.

Even the clinical causal organism data had a significant number of cases (up to 46% of cases) where the culture result was either negative or the culture was not taken, with the % differing widely and inconsistently over time and between hospitals. This again makes it difficult to interpret or compare this data. It is further complicated by the fact that the Oxford data recorded negative cultures and when a sample was not taken but the Leeds data did not record when samples were not taken, potentially inflating the % of cases where a positive result appeared to have been achieved.

Further, confusion is caused by the fact that in Leeds, ICD-10 codes were only used to identify Staph, Strep, other organisms and none, while in Oxford they were used to identify Staph, Strep, Enterococci, other organisms and none. While the clinically recorded data was different again recording Staph, Strep, Enterococci, other organisms and none in Leeds and Staph, Strep, Enterococci, polymicrobial, other organisms, negative and not taken in Oxford i.e. there are no two graphs that use the same parameters, making comparisons difficult or impossible. Figure 5, like the other time trend studies, adds nothing useful to the results and is not mentioned in the discussion and should be dropped, at least in its present form. Furthermore, there is no mention that the type of IE cases seen at major centers like Leeds and Oxford are likely to be skewed towards Staphylococcal IE more than other general hospitals.

Indeed, there is an issue throughout about the different methodology used in Leeds and Oxford. They really seem like different studies and the data from the two sites do not sit comfortably together in a single paper. In particular, the fact that the 'gold standard' against which the ICD-10 codes were tested in Leeds was clinician diagnosed cases (which seems appropriate) but was cases identified through electronic database screening using ICD-10 codes in Oxford (rather than clinician diagnosed cases), seems incongruous, particularly when the whole slant of the paper is to say how poor these codes are at identifying cases.

You have mentioned none of these issues in your paper and have just assumed that all streptococcal codes code for streptococcal IE cases - when this is clearly not the case. Furthermore, despite framing your paper around antibiotic prophylaxis (used to prevent OVGS IE) you have just looked at streptococcal IE which includes many species of Streptococci not relevant to OVGS IE. Many of the 'big data' studies you have criticised for their use of ICD-9/10 codes did not make these errors and/or highlighted these issues, so the conclusion you draw and the criticisms you level against them are again not based on a like for like comparison and are therefore unjustified. Similarly, the 'Clinical and policy implications' you draw in the conclusions of the paper are an over/miss-interpretation of the data. Indeed, much of the data would support different conclusions.

This paper is so focussed on being negative and pejorative about the value of properly conducted 'big data' studies and projects such a strong and dogmatic bias in its use of data to try and undermine such studies, that it renders itself unpublishable in its current form. This is a tragedy because the study is essentially well done and contains some really important, interesting and positive data hidden within it that is highly publishable if done in a scientific manner and without some axe to grind.

The importance of this study is that it provides exactly the sort of data needed to support and facilitate 'big data' studies. Essentially, this is a well conducted study comparing the sensitivity, specificity, PPV and NPV of different ICD-10 codes, and combinations of codes for identifying Duke definite and Duke possible IE cases. The really important data in that respect is in Supplementary data File 11, Table S6. Although, it really should show this data for all of the individual codes (not just I33.0 and all) and
sensible combinations of codes, particularly those combinations used commonly in big data studies. Including all the codes: I33.0, I33.9, 1, I38.0, I39.0, I39.8, I01.1, I09.1, I42.3, B37.6, T82.6 as the main basis for all your analyses is just not sensible or reasonable and certainly not what those researchers who performed the studies you have cited, and keep deriding as naïve, would have considered as a sensible starting place.

It would be helpful to have the sensitivity/specificity/PPV/NPV data for the codes and combinations of codes first (without the effect of other factors such as excluding readmissions etc). You could then provide data on how things like excluding re-admission, excluding stays less than 3 days and excluding elective admissions affect the sensitivity/specificity/PPV/NPV. All of this would be extremely valuable and positive data for those planning future 'big data' studies in order to help them ensure they choose the very best code protocols for their specific study. This data would be much more useful to show than the time course studies that, as outlined above, contribute nothing in their current form.

Figure 2 is also of some value but would be greatly improved if it showed sensible individual codes and coding combinations rather than just comparing I38 alone, and your long list of codes that no one else would consider appropriate for use in a study. This long list of codes was not used in any of the studies cited and seems only to have been chosen to generate the most extreme data possible.

The paper is currently framed entirely around the potential impact of guideline change on IE incidence - which is a time course related issue. It might be more appropriate, however, to re-focus it around identifying the best coding strategies for 'big data' studies - which is not a time course related issue. If the study is going to continue to include a component related to the impact of guideline change, then the time course data needs to be re-analysed to investigate any change in IE incidence trends, comparing before and after the NICE guideline change.

With regard to the causal organism data, the use of ICD-10 codes, as currently defined, is not at all helpful in identifying the causal organism of IE cases. You study provides important additional data to support this and this should therefore be the focus of the discussion on causal organism coding. At present, causal organism data defined by ICD-9/10 codes are too unreliable to be used to draw conclusions in 'big data' studies - certainly for primary study objectives, and this should be the focus of the discussion of this data. Clearly, having ICD codes that specifically identify OVGS as part of any review of ICD codes would be helpful along with better policies on secondary/supplemental coding practices and any discussion should also highlight these issues.

The most striking thing about the causal organism data, particularly the clinically acquired data, is that the % of IE case caused by streptococci remains high and similar to or only a little less than the % caused by staphylococci. This is an interesting and important finding since this is quite different to the finding in some other countries, particularly the USA, where streptococcal IE has been falling steadily in recent years while Staphylococcal cases have risen dramatically. The rise in the USA, at least in part, has been due to the increase in injection drug use, particularly related to the 'opioid crisis'. These differing data help to highlight the diverging pattern of IE disease in the US and UK. Furthermore, they are important because all the data on the replacement of Streptococcal IE by Staphylococcal IE emanating from the US has led to assumptions that this is a widespread phenomenon common to developed countries. And has led to, assumptions that Streptococcal IE is a disease of the past that perhaps no longer merits attention. Your data, clear demonstrates that Streptococci are still an important cause of IE and one that still merits attention.

I would strongly support publication of the important and useful data in this study. Clearly, however,
the paper needs considerable modification to re-focus its narrative on that data rather than on a crusade to criticise other studies, particularly when that data substantially validates, rather than undermines, those studies.

The title of the paper is: "Caveat emptor: the cautionary tale of endocarditis and the potential pitfalls of clinical coding data: an electronic health records study." I don't know about 'let the buyer beware'. As it stands right now, it's the reader who needs to beware the negative and misleading narrative and use of data within this paper. This certainly is a cautionary tale, but it is a cautionary tale about how not to write a paper despite having excellent data and it's an example of the pitfalls of focusing on trying to promulgate a biased and preconceived message rather than talking about the positive data identified in the study. Perhaps a more factual title would be appropriate.

Finally, I should add how difficult was to review this paper when the most important data was spread across 19 additional and separate files of Supplementary Material each of which had to be down loaded separately and was not titled - meaning you had to then keep referring back to the 'List of Additional Figures' in the manuscript. It would be much more helpful for all of this to be properly presented in a single file of Supplementary material with a list of each at the beginning and a title/figure legends attached to each figure, table etc.

**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.

Yes

**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 recommend additional statistical review

**Quality of written English**
Please indicate the quality of language in the manuscript:
**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 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/). 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