Reviewer’s report

Title: Development of a clinical prediction rule for sepsis in primary care: protocol for the TeSDIT study

Version: 0 Date: 18 Mar 2020

Reviewer: Joy Allen

Reviewer's report:

Loots et. al present a protocol for a community based study to collect data to develop a clinical prediction rule for identification of sepsis in the Netherlands. This is commendable and in line with recent efforts for transparency and improved quality over diagnostic test evaluation research. Sepsis recognition is a worldwide unmet need and is particularly complex within primary care. It is noted that the study is in progress and close to completion. Therefore, the below comments regarding potential improvements to the protocol aim to improve the quality and detail presented, in order to aid transferability to other investigators in this area as well as inform the subsequent data analysis.

General comments:
- CPR is commonly used acronym in healthcare for cardiopulmonary resuscitation. Suggest that the authors consider a differ acronym to avoid confusion.
- While the title of the study includes the words diagnostic and prognostic study. The protocol describes a study which is developing models therefore it is not strictly diagnostic nor prognostic (even though the intention for these models will be to diagnose sepsis and predict admission to hospital). Further external validation would be needed to measure diagnostic and prognostic accuracy. Consider revising.
- The TRIPOD (and potentially the STARD) checklist should be followed and all relevant changes made throughout the protocol.

Major comments:
- It is difficult to follow the study processes from patient screening to participant follow up. A study flowchart would help the reader. A lot of the information presented in the discussion should be moved to the methods section for.
- The protocol does not adequately describe the process of data collection and follow up of the patients not referred to hospital. If these patients were subsequently referred to hospital, would this data be captured? This is important as they are false negatives in the current care pathway and so, a potential improvement over current care would be to ensure these patients were referred sooner.
- Could the authors make it clear their a priori rules for the expert panel diagnosis? It is unclear what data they will be presented with when, and in which patients in order to form their diagnosis. For example, in line 343 it states that the expert panel will be instructed to use the SOFA score. The supplementary material suggested that the SOFA score will be inputted for those patients where it is not available i.e. those not admitted to hospital, is this the case. Imputation based on this data risks bias within the development of the model. In general, more clarity is needed on the analysis plan in the protocol. An appended statistical analysis plan could add strength to this protocol and aid other investigators in this complex area.
Within the data cleaning section (lines 251 - 255): the explanation of removal of outliers 'more than three standard deviations from the mean' is worrying as it could lead to selection bias. By nature of sepsis, these patients may have extremely high or extremely low clinical test results. A more appropriate plan would be to look for statistical outliers and explore whether these are data recording errors. If so, they could be removed. If not, they should be included as they may represent the most important data of all. Failure to include these data points could result in selection bias. Please see reference: Rousseeuw, P.J. and Hubert, M. (2011), Robust statistics for outlier detection. WIREs Data Mining Knowl Discov, 1: 73-79. doi:10.1002/widm.2

Do you have a health economic evaluation plan? If so, please add to the supplementary material as it will be useful for other researchers.

Minor comments:
- Presentation of the inclusion and exclusion criteria could be improved and separated on individual lines to help the reader. Further clarity could be added to also help the reader from other countries understand the exact patient group which are eligible.
- The protocol is inconsistent in the description of the QoL instrument (EQ-5D or EQ-5D-5L) which will be used. Please choose appropriate tool and make consistent.
- How the data regarding confidence in diagnosis was going to be analysis was not clear. Please clarify analysis plan.
- Line 89: propose changing diagnosis to recognise…
- Line 155: Change "in" to "into".
- Lines 186-188: there is an apparent time delay in taking blood samples from those patients who are referred to hospital and now. Will this be accounted for in the analysis?
- Lines 213-214: Asking patients to complete an EQ-5D-5L questionnaire from recall for the worst day they remember from their recent sepsis episode is going to produce results of questionable accuracy. A symptom of sepsis is acute confusion, and this, coupled with the duration of time from their sepsis episode to the end of follow-up when they are asked to complete the questionnaire (30 days), would lead to poor accuracy of recall in this largely elderly population. This should be discussed as a limitation.
- Line 226: EMV not defined.
- Lines 232 - 234: It is important to note the evidence advocating 10 events per variable in binary logistic regression analysis is weak. Further explanation should be provided why you are using this rule. The authors may find this paper helpful: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710621/pdf/10.1177_0962280218784726.pdf
- Line 248: primary outcome should read "sepsis within 72 hours"
- Table 1 - add references to the sources in order to future proof this publication.
- Line 289 - Using AUROC to compare between different models should be used with caution, please see reference: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-13
- Line 331 - try to implement as good as possible: reword to 'try to implement as much/well as possible'
- Line 360 - change retroactive to retrospective.
- Are the investigators planning to make the data available for other investigators? It would add strength to make their intentions clear in like 367.
Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

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.