BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

**ARTICLE DETAILS**

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>The Simplified Mortality Score for the Intensive Care Unit (SMS-ICU): Protocol for the development and validation of a bedside clinical prediction rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Granholm, Anders; Perner, Anders; Krag, Mette; Buhl Hjortrup, Peter; Haase, Nicolai; Holst, Lars; Marker, Søren; Collet, Marie; Jensen, Aksel; Møller, Morten</td>
</tr>
</tbody>
</table>

**VERSION 1 - REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Antonio Palazón-Bru</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Miguel Hernández University, Spain</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>07-Dec-2016</td>
</tr>
</tbody>
</table>

**GENERAL COMMENTS**

Comments to the authors:
1) A new risk score for the analyzed research question was recently published with simple variables. The validation was in accordance with the Prof. Gary S Collins's recommendations. They should be mentioned in the text. Their references are:
2) Please, use "participants" or "sample", instead of "population", because you are not using all the potential patients which you would use the new risk score. In other words, you are using a sample of the study population.
3) Could the interventions in the three trials affect the main outcome variable (90-day mortality)?
4) "Hospital length of stay prior to inclusion/ICU admission". Could this variable take the null value?
5) Why will not you use means and standard deviations so as to describe the quantitative variables?
6) I suppose the algorithm to select the final variables will be based on the likelihood ratio test, but this has not been mentioned.
7) I would need a more detailed explanation in order to understand how you will transform the multivariate model into a risk score. I do not know if you will use this reference: Sullivan LM, Massaro JM,

8) If you read the letter written by Prof. Gary S Collins (see comment #1), you will see that it is better to use smooth calibration instead of the calibration framework (intercept and slope).

9) Please, revise the order of your references. For example, I have seen that DeLong et al. is the #33 in the list and #34 in the text.

10) My main concern is the origin of your patients; would you analyze only the control group of all the trials?

11) I would like to see more information about the development of the mobile application. For example, which operating system (Android, IOS…) will you use?

12) I do not know the statistical software which you will use to analyze the data sets.

13) Please, could you add a figure to explain the flow chart of the protocol?

No more comments. In general, I think you have a great idea with a fantastic methodology to be used in clinical practice. Congratulations!

---

**REVIEWER**

Romain Pirracchio  
Department of Anesthesia and Intensive Care Medicine  
Hôpital Européen Georges Pompidou  
Université Paris Descartes  
Paris, France

**REVIEW RETURNED**

19-Dec-2016

---

**GENERAL COMMENTS**

I enjoyed reading the protocol for the development of a Simplified Mortality Score for the Intensive Care Unit (SMS-ICU). Overall, the research question is sound and the protocol designed to develop this new score makes sense. However, I still have few concerns:

- **primary outcome:** 90-day mortality. The authors claim they are willing to build up a score that would be relevant at the bedside. Is 90-day realistically the main clinician preoccupation when he admits a patient in the ICU? Because ICU mortality is far from being homogeneous over time, I would pick up a shorter term mortality as primary outcome measure.

- **table 1:** in the footnote, they authors state the values can either be the one obtained during the 24 hours prior to ICU admission or during the first 24 hours in the ICU. I believe these 2 pieces are far from being exchangeable. In many situation, they carry very different information. In addition, data form the day before ICU admission are very often missing in practice, which makes again the score poorly usable in practice.

- **logistic regression:** why did the authors choose to use a fully parametric model, when there is so many evidence in the literature showing that machine learning based algorithm perform way better in this context?

- **external validation:** I am confused when the authors state they are planning to use multiple imputation in the external validation cohort. External validation is supposed to reflect real life condition. In real life, if a value is missing, we are not using multiple imputation to get it. So missing values should not be imputed here.
2. A new risk score for the analyzed research question was recently published with simple variables. The validation was in accordance with the Prof. Gary S Collins’s recommendations. They should be mentioned in the text. Their references are:


Response:
We thank the reviewer for providing this information, including references, which we have read with great interest.
We have mentioned and cited the score in the revised manuscript (page 4).

3. Please, use “participants” or “sample”, instead of “population”, because you are not using all the potential patients which you would use the new risk score. In other words, you are using a sample of the study population.

Response:
We thank the reviewer for this comment. We agree and have replaced the word “population” in the revised manuscript.

4. Could the interventions in the three trials affect the main outcome variable (90-day mortality)?

Response:
Thank you for this valid point. Please find our considerations regarding the three trials below:

1) Scandinavian Starch for Severe Sepsis/Septic Shock (6S) trial: In the 6S trial, an increased mortality was seen in the group allocated to hydroxyethyl starch (HES) compared with Ringer’s acetate. This group comprises a total of 398 patients or approximately 9 % of the full development sample. We believe that including the HES group in the analyses and thus increasing the sample size is the optimal approach; however, following the comment, we have decided to conduct a sensitivity analysis where analyses are repeated in the development sample after exclusion of patients allocated to HES in 6S. This has been added to the revised manuscript (page 12).

2) Transfusion Requirements In Septic Shock (TRISS) trial: In the TRISS trial, 90-day mortality rates were similar among both groups, and as such we consider the inclusion of all patients from the TRISS trial appropriate. Additionally, mortality rates did not differ at 1 year or at longest follow-up, which was 1 year after randomisation of the last patient, median time of follow-up of 21 months (Rygård SL et al., Long-term outcomes in patients with septic shock transfused at a lower versus a higher haemoglobin threshold: the TRISS randomised, multicentre clinical trial. Intensive Care Med. 2016 Nov;42(11):1685-1694).

3) Conservative versus Liberal Approach to fluid therapy of Septic Shock in Intensive Care (CLASSIC) trial: In the CLASSIC trial, no statistically significant difference was found between the groups regarding 90-day mortality. Additionally, fluid therapy widely varies in clinical practice between departments (Hjortrup PB et al., Associations of Hospital and Patient Characteristics with Fluid
Resuscitation Volumes in Patients with Severe Sepsis: Post Hoc Analyses of Data from a Multicentre Randomised Clinical Trial. PLoS ONE. 2016 May 19;11(5):e0155767, and consequently we believe that it is appropriate to include the full CLASSIC sample in the analyses.

Regarding the Stress Ulcer Prophylaxis in the Intensive Care Unit (SUP-ICU) trial (external validation sample):

We do not currently know whether the use of stress ulcer prophylaxis (SUP) increases or decreases mortality in critically ill patients. In the revised manuscript, we have added plans for the conduction of two sensitivity analysis - one in each arm of the trial - after the external validation has been conducted in the full sample (page 14).

5. “Hospital length of stay prior to inclusion/ICU admission”. Could this variable take the null value?

Response:
Yes, this variable can be null as it will be recorded as full days. The variable definition has been slightly changed in the revised manuscript.

Generally, in the study databases, inclusion dates and times are registered, but only hospital admission dates (and not times) are registered. The variable will be calculated as ICU admission date minus hospital admission dates, and consequently be zero for patients admitted to the ICU on the same day as they were admitted to the hospital.

This is a minor limitation in the source data, however, we do not expect it to have any influence of importance on the resulting score. In the Simplified Acute Physiology Score (SAPS) 3 (Moreno et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005 Oct;31(10):1345-1355) a “hospital length of stay prior to ICU admission” variable is included, and in the SAPS 3, points are only given when the variable is > 14 and > 28 days, respectively. Additionally, using hours instead of whole days would complicate a score that is developed with the intention of being simple and easy to use.

6. Why will not you use means and standard deviations so as to describe the quantitative variables?

Response:
This is primarily a matter of preference. In the original trials/studies median (IQR) were used, as this in our opinion is the most conservative estimate. Consequently, we prefer using/reporting median (IQR).

Additionally, this approach may be more appropriate if some variables turn out to be skewed with a need of transformation (page 10) before they are included in the logistic regression analysis. This will likely apply to the “hospital length of stay” variable and possibly to other variables in the dataset.

7. I suppose the algorithm to select the final variables will be based on the likelihood ratio test, but this has not been mentioned.

Response:
This is true indeed. The selection of variables will be performed using a backward stepwise elimination approach (as described in the protocol). In each step, the variable with the largest P-value reported by the likelihood ratio test will be removed. This has been further clarified in the revised manuscript (page 10).

We have deliberately decided not to pre-specify a significance threshold for variables that stay in the model, as we aim to create a simple score that can be calculated easily in daily clinical practice.
8. I would need a more detailed explanation in order to understand how you will transform the multivariate model into a risk score. I do not know if you will use this reference: Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med. 2004;23:1631-60.

Response:

We thank the reviewer for this comment!
We will indeed perform the transformations according to the Framingham Heart Study approach, which in several regards is similar to the reference already in the manuscript (Møller et al. The Peptic Ulcer Perforation (PULP) score: a predictor of mortality following peptic ulcer perforation. A cohort study. Acta Anaesthesiol Scand. 2012 May;56(5):655-662).
This has been further elaborated in the revised manuscript (page 11).

9. If you read the letter written by Prof. Gary S Collins (see comment #1), you will see that it is better to use smooth calibration instead of the calibration framework (intercept and slope).

Response:

We thank the reviewer for this comment.
The choice of calibration method can be discussed and the most appropriate measure also depends on the sample size and the event rate. Intercept and slope will still be calculated/reported, as they are easy-to-understand summary measures. However, in addition we will provide calibration curves (which was already planned but not described clearly in the protocol) and the Hosmer-Lemeshow goodness-of-fit C-statistic to further assess the calibration. This is in line with recommendations in the literature, including the paper that Professor Gary S. Collins refers to in his letter mentioned in comment #1 (Van Calster B et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016 Jun;74:167-176 and Labarère et al. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med. 2014 Apr;40(4):513-527).
This has been highlighted in the revised manuscript (page 11).

10. Please, revise the order of your references. For example, I have seen that DeLong et al. is the #33 in the list and #34 in the text.

Response:

We have updated/revised the references in the revised manuscript.

11. My main concern is the origin of your patients; would you analyze only the control group of all the trials?

Response:

Please see our response to comment #4.

12. I would like to see more information about the development of the mobile application. For example, which operating system (Android, IOS…) will you use?

Response:

We thank the reviewer for the interest!
This has not been decided upon yet. Following successful development and validation of the score, we will go more into details on this matter. Most likely, we will develop a web-based app, eventually augmented by application for the different operating systems.
13. I do not know the statistical software which you will use to analyze the data sets.

Response:
We plan to conduct all analyses using SAS version 9.4, SAS Institute Inc., Cary, NC, USA (or newer versions if updated). This has been added to the revised manuscript (pages 12 and 14).

14. Please, could you add a figure to explain the flow chart of the protocol?

Response:
We thank reviewer #1 for this suggestion, and a figure has been added to the revised manuscript (page 4).

15. No more comments. In general, I think you have a great idea with a fantastic methodology to be used in clinical practice. Congratulations!

Response:
Thank you very much.

Reviewer #2 - Romain Pirracchio:
16. I enjoyed reading the protocol for the development of a Simplified Mortality Score for the Intensive Care Unit (SMS-ICU). Overall, the research question is sound and the protocol designed to develop this new score makes sense.

Response:
Thank you very much.

17. Primary outcome: 90-day mortality. The authors claim they are willing to build up a score that would be relevant at the bedside. Is 90-day realistically the main clinician preoccupation when he admits a patient in the ICU? Because ICU mortality is far from being homogeneous over time, I would pick up a shorter term mortality as primary outcome measure

Response:
The European Societies for Anaesthesiology and Intensive Care Medicine (ESA and ESICM) recommend use of longer fixed-term mortality outcomes, i.e. 90-day mortality or longer (Jammer I et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions: a statement from the ESA-ESICM joint taskforce on perioperative outcome measures. Eur J Anaesthesiol. 2015 Feb;32(2):88-105), as patients may survive for the first couple of weeks or longer or for their entire ICU or hospital stay and subsequently die. Additionally, we have recently shown no difference between the use of in-hospital and 90-day mortality when using the SAPS II score (Granholm A et al. Predictive Performance of the Simplified Acute Physiology Score (SAPS) II and the Initial Sequential Organ Failure Assessment (SOFA) Score in Acutely Ill Intensive Care Patients: Post-Hoc Analyses of the SUP-ICU Inception Cohort Study. PLoS ONE. 2016 Dec 22;11(12):e0168948). Intensivists may be more concerned about in-ICU mortality than mortality at longer times of follow-up, however we believe that for most patients survival to ICU or hospital discharge is of little value if followed by death within a short time-frame.

For scores that aim to predict in-ICU or in-hospital mortality, it has been discussed that the scores may “predict the obvious” (Suistomaa M et al. Customized prediction models based on APACHE II and SAPS II scores in patients with prolonged length of stay in the ICU. Intensive Care Med. 2002 Apr;28(4):479-485), as patients that die within a short time frames often are so severely ill upon ICU admission that clinicians are often able to determine their risks without the use of scoring systems at
We believe that using 90-day mortality potentially adds to our score, as this may enable clinicians using the score to detect patients at high risk of dying where it is not as obvious and thus increase treatment efforts leading to better patient outcomes.

18. Table 1: in the footnote, the authors state the values can either be the one obtained during the 24 hours prior to ICU admission or during the first 24 hours in the ICU. I believe these 2 pieces are far from being exchangeable. In many situation, they carry very different information. In addition, data form the day before ICU admission are very often missing in practice, which makes again the score poorly usable in practice.

Response:
We thank reviewer 2 for this comment.

To clarify: in the two cohort studies, patients were enrolled upon ICU admission and baseline data were collected during the first 24 hours in the ICU. In the RCTs, patients were enrolled after ICU admission and baseline data was collected for the past 24 hours, as patients were already in the ICU and the treatment assignments could potentially alter the physiologic and biochemical values recorded.

We agree that this is a limitation of the study, which will need to be acknowledged in manuscript to come. However, we believe the limitation is of minor importance, as the worst physiological variables are most likely to be recorded shortly after admission to an intensive care unit, followed by normalization of the values due to the treatments initiated in the ICU.

19. Logistic regression: why did the authors choose to use a fully parametric model, when there is so many evidence in the literature showing that machine learning based algorithm perform way better in this context?

Response:
We believe machine learning is an exciting field and that machine learning methods will be utilized increasingly in the future to the benefit of patients. Researchers from our department are currently initiating a research programme based on machine learning and big data analysis, which will hopefully reveal new predictors and patterns of physiological values that can be used to predict different ICU outcomes. However, as recently recommended we will use logistic regression (Labarère et al. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med. 2014 Apr;40(4):513-527).

20. - External validation: I am confused when the authors state they are planning to use multiple imputation in the external validation cohort. External validation is supposed to reflect real life condition. In real life, if a value is missing, we are not using multiple imputation to get it. So missing values should not be imputed here.

Response:
Multiple imputation will not be used bedside in the external validation sample, but will exclusively be used for missing values in the external validation database (to replace missing data), as recently recommended (Vesin A et al. Reporting and handling missing values in clinical studies in intensive care units. Intensive Care Med. 2013;39:1396-1404). This is in accordance with expert recommendations (Collins GS et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014 Mar;14:40). All candidate variables are possible to obtain within 24 hours after ICU admission.

This has been clarified in the revised manuscript (page 13).

Additional changes
21. Since the first submission, the project has received funding and thus the funding statement has been updated accordingly.

22. The study protocol that was initially prepared and co-submitted with the original version of this manuscript has been omitted from this revised manuscript. As this protocol article manuscript has been updated according to reviewer recommendations, this manuscript replaces the original protocol pdf, that is no longer up to date. The supplementary files have been renamed accordingly.

23. Details regarding the AID-ICU cohort study have been updated.

**VERSION 2 – REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Antonio Palazón-Bru</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Miguel Hernández University, Spain.</td>
</tr>
</tbody>
</table>

| REVIEW RETURNED           | 18-Jan-2017                             |

| GENERAL COMMENTS          | Great work! All my comments have been addressed correctly. |

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Romain Pirracchio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hopital Européen Georges Pompidou</td>
</tr>
<tr>
<td></td>
<td>Department of Anesthesia and Intensive Care Medicine</td>
</tr>
<tr>
<td></td>
<td>France</td>
</tr>
</tbody>
</table>

| REVIEW RETURNED           | 24-Jan-2017                             |

| GENERAL COMMENTS          | I would like to thank the authors for their answers. I have no more comment at that stage. |
Simplified Mortality Score for the Intensive Care Unit (SMS-ICU): protocol for the development and validation of a bedside clinical prediction rule

Anders Granholm, Anders Perner, Mette Krag, Peter Buhl Hjortrup, Nicolai Haase, Lars Broksø Holst, Søren Marker, Marie Oxenbøll Collet, Aksel Karl Georg Jensen and Morten Hylander Møller

BMJ Open 2017 7:
doi: 10.1136/bmjopen-2016-015339

Updated information and services can be found at:
http://bmjopen.bmj.com/content/7/3/e015339

These include:

References
This article cites 42 articles, 0 of which you can access for free at:
http://bmjopen.bmj.com/content/7/3/e015339#ref-list-1

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Intensive care (213)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/