PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Screening tools for predicting mortality of adults with suspected sepsis: an international sepsis cohort validation study
AUTHORS	Blair, Paul; Mehta, Rittal; Oppong, Chris; Tin, Som; Ko, Emily; Tsalik, Ephraim; Chenoweth, Josh; Rozo, Michelle; Adams, Nehkonti; Beckett, Charmagne; Woods, Christopher; Striegel, Deborah; Salvador, Mark G; Brandsma, Joost; McKean, Lauren; Mahle, Rachael; Hulsey, William R; Krishnan, Subramaniam; Prouty, Michael; Letizia, Andrew; Fox, Anne; Faix, Dennis; Lawler, James V; Duplessis, Chris; Gregory, Michael; Vantha, Te; Owusu-Ofori, Alex Kwame; Ansong, Daniel; Oduro, George; Schully, Kevin; Clark, Danielle

VERSION 1 – REVIEW

REVIEWER	Leichtle, Alexander University of Bern
REVIEW RETURNED	09-Oct-2022

GENERAL COMMENTS	Review of the manuscript "The performance of screening tools for predicting mortality across multi-site international sepsis cohorts" by Blair et al. In their manuscript, the authors compare the performance of 5 commonly used sepsis screening tools in different countries. The study included n=567 participants, with a pooled mortality of 16.4% over 28 days. They found regional differences of the prediction accuracy and conclude that scores should be validated in the respective target populations. Key points: Sepsis is an important health condition, and its accurate mortality prediction could improve patient care. The authors included also data from non western countries and thereby contribute to diversity in medicine, which is a merit of the study. The study is relatively small, which might hamper the validity of the outcomes as described by the authors in the limitations. The study has been waived by IRBs as a precondition of ethical research. The inclusion criteria are partly "soft" (deferral of enrollment in case of "too sick"). Specific comments: 1. Why were continuous predictors dichotomized? What were "clinically relevant" cutoffs? 2. What is the rationale of using sepsis scores to predict mortality? Wouldn't specific mortality scores more adequate?

	3. The statement, that models should be validated in the respective cohort is important and could be emphasized more, maybe even in the heading. 4. The figure on page 26 seems distorted. 5. What is the potential cause for sodium having sich a high HR (page 34)? Is this iatrogenic?

REVIEWER	Li, Wei Southeast University, Epidemiology and Health Statistics
REVIEW RETURNED	13-Oct-2022

GENERAL COMMENTS	1. The methods of this study are unclear, and the statistical methods are confusing 2. Statistical analysis of baseline demographic characteristics stratified by different sites needs to be carried out in Table 1. 3. Statistical analysis of Pathogens detected results between different sites needs to be carried out to identify differences 4.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Alexander Leichtle, University of Bern
Comments to the Author:
Review of the manuscript "The performance of screening tools for predicting mortality across multi-site international sepsis cohorts" by Blair et al.
In their manuscript, the authors compare the performance of 5 commonly used sepsis screening tools in different countries. The study included n=567 participants, with a pooled mortality of 16.4% over 28 days. They found regional differences of the prediction accuracy and conclude that scores should be validated in the respective target populations.

Key points:
Sepsis is an important health condition, and its accurate mortality prediction could improve patient care.
The authors included also data from non western countries and thereby contribute to diversity in medicine, which is a merit of the study.
The study is relatively small, which might hamper the validity of the outcomes as described by the authors in the limitations.
The study has been waived by IRBs as a precondition of ethical research.
The inclusion criteria are partly "soft" (deferral of enrollment in case of "too sick").

Specific comments:
1. Why were continuous predictors dichotomized? What were "clinically relevant" cutoffs?
RESPONSE: These were dichotomized to improve the clinical utility or inference based on physiologically abnormal cut points. While continuous parameters can have more statistical power, they are less translatable to clinical care. Where possible, these cut points were at individual components of existing risk scores. We have added information in the methods about this. The cut points are included in Figure S3 rather than listed in the text for ease of reading.

2. What is the rationale of using sepsis scores to predict mortality? Wouldn't specific mortality scores more adequate?

RESPONSE: These scores are readily available, frequently used in clinical care, and supported by multiple single-site studies. Certain risk scores including the UVA score and qSOFA were created to predict mortality for risk stratification. The APACHE score however includes parameters that were not available in this dataset.
We have corrected the background to describe this: “The qSOFA and other sepsis screening tools (i.e., Modified Early Warning Score [MEWS], National Early Warning Score [NEWS], and Universal Vital Assessment [UVA]) are often used clinically to identify those at risk of sepsis, but these tools have been studied for their ability to prognosticate mortality or poor composite outcomes among hospitalized adults9-12”

3. The statement, that models should be validated in the respective cohort is important and could be emphasized more, maybe even in the heading.
RESPONSE: We agree and have changed the headline to “validation” of screening tools to fit the study objectives more appropriately and also changed the introduction wording to emphasize this.

4. The figure on page 26 seems distorted.
RESPONSE: Our page numbers appear to be different. Please provide the figure number that appears distorted, and we will be happy to make any figure file fixes.

5. What is the potential cause for sodium having such a high HR (page 34)? Is this iatrogenic?
RESPONSE: We added the hypernatremia HR in the results as this is worth highlighting. High sodium (hypernatremia) can occur in sepsis due to intravascular fluid loss due to breakdown of vascular cell junctions, insensible fluid losses/ dehydration from the disease process. There could be an iatrogenic contribution with diuresis in patients or with inadequate fluid resuscitation. Ultimately, there is not data available to determine precisely the cause of hypernatremia, but this is known finding that can occur during critical illness.

Reviewer: 2
Dr. Wei Li, Southeast University
Comments to the Author:
1. The methods of this study are unclear, and the statistical methods are confusing
RESPONSE: To improve the flow and clarity we have broken up the methods section into subheadings and edited throughout to improve readability.

2. Statistical analysis of baseline demographic characteristics stratified by different sites needs to be carried out in Table 1.
RESPONSE: We have added a column to include the statistical test p-value comparing the characteristics across sites.

3. Statistical analysis of Pathogens detected results between different sites needs to be carried out to identify differences
RESPONSE: We have added a statistical test to describe that confirmed infections differed across sites. Infection classes were different among sites ($p<0.001$, chi-squared test).

VERSION 2 – REVIEW

REVIEWER	Leichtle, Alexander University of Bern
REVIEW RETURNED	03-Jan-2023

GENERAL COMMENTS	Re.review of the manuscript "Validation of screening tools for predicting mortality across multi-site international sepsis cohorts" by Blair et al. As this is a revised manuscript, I comment on the authors' responses directly: 1. Resolved 2. Resolved 3. Resolved 4. By counting I would say it's Fig. 2 5. I'd like to see this answer paragraph in the text. The authors resolved most of the comments, and there are only very minor additions to be made that do not require an additional review round.

REVIEWER	Li, Wei Southeast University, Epidemiology and Health Statistics
REVIEW RETURNED	14-Dec-2022

GENERAL COMMENTS	There are significant differences in baseline demographic characteristics among different sites(Table 1 and Table S1), and they found regional differences of the prediction accuracy of survival (Figure 1). How to interpret the impact of baseline demographic characteristics on prediction accuracy?

VERSION 2 – AUTHOR RESPONSE

"As this is a revised manuscript, I comment on the authors' responses directly:
1. Resolved
2. Resolved
3. Resolved
4. By counting I would say it's Fig. 2"
RESPONSE: We have uploaded a new Figure 2 in a new file format.
"5. I'd like to see this answer paragraph in the text."
RESPONSE: We have added a section about hypernatremia in the discussion including the information from this paragraph.

VERSION 3 – REVIEW

REVIEWER	Li, Wei Southeast University, Epidemiology and Health Statistics
REVIEW RETURNED	31-Jan-2023

GENERAL COMMENTS	There are many variables in Table 1 and Table S1, why the model only adjusted for age and gender in table 3.