Peer Review File

Article Information: https://dx.doi.org/10.21037/atm-23-432

Round 1

1) First of all, my major concerns regarding this study is the unclear focuses, the prognostic role of sTM for mortality in ARDS, the prognosis prediction accuracy of sTM for mortality in ARDS, or both. It seems that the authors focused on both so the methodology of this meta-analysis should be the meta-analysis of cohort studies and predictive accuracy studies. For the latter one focus, the authors should use PROBAST to assess the risk of bias of included studies but the authors did not such procedure.

   Reply: We absolutely agree with the reviewer's opinion. Our research focused on the prognostic role and predictive value of sTM for hospital mortality in ARDS. The predictive role and predictive accuracy are highly connected, numerous meta-analyses with same research design have been published [1-3]. PROBAST (Prediction model Risk Of Bias Assessment Tool) was developed for assessing the risk of bias (ROB) and applicability of diagnostic and prognostic prediction model studies. The developers of this tool informed "Studies that use multivariable modeling techniques to identify predictors (such as risk or prognostic factors) associated with an outcome but do not attempt to develop, validate, or update a model for making individualized predictions are not covered by PROBAST" [4]. The nine studies included in our meta-analysis intended to investigate predictive accuracy of sTM for mortality of ARDS, but not to develop or validate a prediction model incorporating sTM. Therefore, we did not use PROBAST to assess the ROB.

2) Second, the title is not accurate and not clear, which did not indicate the he prognostic role of sTM for mortality in ARDS, the prognosis prediction accuracy of sTM for mortality in ARDS, or both, and the prognosis outcome to be predicted.

   Reply : We have revised the title to explicitly recapitulate the main ideas of our research.

   Changes in the text: See Page 1, line 3-4.

3) Third, the abstract needs to be revised. The background did not describe the clinical controversy on the prognostic role and the prognosis prediction accuracy of sTM and whether a meta-analysis is suitable to address the controversy. The methods did not describe the inclusion criteria, data extraction and the meta-analysis for analyzing the prognostic role.

   Reply: Due to words limitation, we cannot extensively elucidate research background and methods in the abstract. These important points were addressed in the Introduction and Method section. In fact, we have to furtherly simplify the abstract as editor requested.

4) Fourth, the introduction of the main text needs to provide detailed examples on the clinical controversy on the prognostic role and the prognosis prediction accuracy of sTM, analyze the potential reasons, and explain why a meta-analysis is suitable to address this issue.

Reply 4: We have revised the introduction section.

Changes in the text: See Page 3, line 92.

5) Fifth, in the methodology of the main text, the authors need to clarify the criteria for the inclusion of prediction accuracy studies, which is different from prognostic role studies. The inclusion criteria need to be strictly defined according to the PICOS principles, in particular the clinical research design of studies to be included and whether unadjusted or adjusted prognostic role data of sTM were extracted. The three level criteria of quality in NOS are not convincing, the authors need to revise them. In fact, only studies scored 9 were of high quality or low risk of bias. The authors need to describe the risk of bias assessment for the prediction accuracy studies. In statistics, procedures for analyzing the heterogeneity, test of sources of heterogeneity, and test of publication for the prediction accuracy studies should also be described.

Reply: We rewrote and clearly demonstrated the inclusion and exclusion criteria according to PICOS principle. NOS is a mainstream tool adopted to assess bias risk for observable study in the meta-analysis. NOS was used in the studies published in journals with high impact factor and scored 7-9 generally considered low risk of bias [5-7]. The bivariate method for the prediction accuracy is based on random effects and the distribution of sensitivity and specificity of individual studies were demonstrated in the Figure 5, so we did not further analyze the heterogeneity. Test of publication for the prediction accuracy studies was shown in the revised manuscript.

Changes in the text:

See Page 4-5, line 121-141.

See Page 6, line 171(Method section) and Page 8, line 237 (Results section), Added Figure S3, (the numbers of original Figure S3 and Figure S4 were changed to Figure S4 and Figure S5 without revision)

6) Finally, please consider to cite the below related papers: Zheng Z, Chang Z, Chen Y, Li J, Huang T, Huang Y, Fan Z, Gao J. Total bilirubin is associated with all-cause mortality in patients with acute respiratory distress syndrome: a retrospective study. Ann Transl Med 2022;10(21):1160. doi: 10.21037/atm-22-1737. Sun W, Luo Z, Cao Z, Wang J, Zhang L, Ma Y. A combination of the APACHE II score, neutrophil/lymphocyte ratio, and expired tidal volume could predict non-invasive ventilation failure in pneumonia-induced mild to moderate acute respiratory distress syndrome patients. Ann Transl Med 2022;10(7):407. doi: 10.21037/atm-22-536.

Reply 6: The relevance of aforementioned articles with our research is weak. We will consider to cite the articles published on the ATM in future research.

References:

1. Aloisio E, Braga F, Puricelli C, et al. Prognostic role of Krebs von den Lungen-6 (KL-6) measurement in idiopathic pulmonary fibrosis: a systematic review and meta-analysis. Clin Chem Lab Med. 2021 Apr 9;59(8):1400-1408. doi: 10.1515/cclm-2021-0199. PMID: 33831978.

2. Medrinal C, Combret Y, Hilfiker R, et al. ICU outcomes can be predicted by noninvasive muscle evaluation: a meta-analysis. Eur Respir J. 2020 Oct 1;56(4):1902482. doi: 10.1183/13993003.02482-2019. PMID: 32366493.

3. Shaikh N, Lee MC, Stokes LR,et al. Reassessment of the Role of Race in Calculating the Risk for Urinary Tract Infection: A Systematic Review and Meta-analysis. JAMA Pediatr. 2022 Jun 1;176(6):569-575. doi: 10.1001/jamapediatrics.2022.0700. Erratum in: JAMA Pediatr. 2022 Jun 21;:null. PMID: 35435935; PMCID: PMC9016605.

4. Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-58. doi: 10.7326/M18-1376. PMID: 30596875.

5. Zacher Kjeldsen MM, Bricca A, Liu X,et al. Family History of Psychiatric Disorders as a Risk Factor for Maternal Postpartum Depression: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2022 Oct 1;79(10):1004-1013. doi: 10.1001/jamapsychiatry.2022.2400. PMID: 35976654; PMCID: PMC9386615.

6. Wolf J, Hubbard S, Brauer M,et al. Effectiveness of interventions to improve drinking water, sanitation, and handwashing with soap on risk of diarrhoeal disease in children in low-income and middle-income settings: a systematic review and meta-analysis. Lancet. 2022 Jul 2;400(10345):48-59. doi: 10.1016/S0140-6736(22)00937-0.

7.Vai B, Mazza MG, Delli Colli C,et al. Mental disorders and risk of COVID-19-related mortality, hospitalisation, and intensive care unit admission: a systematic review and meta-analysis. Lancet Psychiatry. 2021 Sep;8(9):797-812. doi: 10.1016/S2215-0366(21)00232-7. Epub 2021 Jul 17. PMID: 34274033; PMCID: PMC8285121.

Round 2

1) First, essential information of this meta-analysis should be briefly provided in the abstract, regardless of the restriction on the length of abstract including just describing that there is controversy regarding the prognostic role and mortality prediction accuracy of STM in the background, the clinical research design of studies to be included in the methods, and number of studies with low risk of bias in the results. The authors need to delete other unnecessary sentences in the abstract to meet the requirement on length.

Reply: We have revised the manuscript according to reviewer's suggestion.

Changes in the text: See Page 1-2, line 28-61.

2) Second, detailed examples on the controversy regarding the prognostic role and mortality prediction accuracy of STM in the introduction of the main text is still lacking, which is necessary for a meta-analysis, since meta-analysis is often used to address controversy. Without convincing evidence for the controversy, the meta-analysis should not be performed.

Reply: We highlighted the controversy in the revised manuscript.

Changes in the text: See Page 4, line 103-109.

3) Third, in the methodology of the main text, it is clear that NOS is only suitable for assessing the risk of bias of case-control and cohort studies for studies of prognostic factors, but for the predictive accuracy (the so called "predictive value"), if the authors believed PROBAST is not appropriate, please consider QUADAS-2. Anyway, risk of bias assessment predictive value studies should be performed, which cannot be mixed with the prognostic role focus of this study. The authors described that "cross-sectional studies" are eligible, but NOS cannot assess the risk of bias of such studies. Please clarify whether the extracted OR values are adjusted or unadjusted OR values.

Reply: The risk of bias of studies on predictive value of sTM for ARDS mortality was evaluated by QUAPAS (Quality Assessment of Prognostic Accuracy Studies) tool, which is a modified tool derived from QUADAS-2 and PROBAST [1], as following Figure 1. QUAPAS was used to evaluate bias risk for studies on predictive value of sTM in the revised manuscript. Presently, no ‘‘gold standard’’ tool exists for assessing methodological quality in cross-sectional studies. The NOS is inappropriate to assess study quality for cross-sectional studies, some researchers adopted modified NOS for cross-sectional studies [2-3], as following Figure 2. So we used modified NOS to reevaluate the cross-sectional study by the Benatti MN (2020) and revised the manuscript. The extracted OR values were adjusted OR values, which was highlighted in the revised manuscript.

Changes in the text:

See Table 3 and Page 5-6, line 16-171 (Method section) and Page 8, line 247(Result section)

See Table 1 and Page 5, line 161-164.

See Page 6, line 179(Method section) and Page 8, line 237-238(Result section).

Domain	Participants	Index Test	Outcome	Flow and Timing	Analysis
Description	Describe methods for recruiting participants Describe participants (previous testing, presentation, intended use of index test, and setting)	Describe the index test (definition, context of use, method of measurement, and interpretation)	Describe the outcome (definition, method of measurement, and interpretation)	Describe any participants lost to follow-up or excluded from the analysis Describe the time horizon from the index test to the outcome	Describe the statistical methods
Signaling questions (yes, no, unclear)	S1.1: Was a consecutive or random sample of participants enrolled?† S1.2: Was a case–control design avoided?‡ S1.3: Did the study avoid inappropriate selection criteria?†	S2.1: Was the method used to perform the index test valid and reliable?§ S2.2: Was the method for performing the index test the same for all participants?§‖ S2.3: Were the index test results interpreted without knowledge of the outcome?† S2.4: If a threshold was used, was it prespecified?‡	S3.1: Was the method used to measure the outcome valid and reliable?† S3.2: Was the method for measuring the outcome the same for all participants?† S3.3: Was the outcome measured without knowledge of the index test results?†	S4.1: Did all participants receive the index test?‖ S4.2: Was treatment avoided after the index test was performed?¶ S4.3: Was the time horizon sufficient to capture the outcome?† S4.4: Was information on the outcome available for all participants?†	S5.1: Were all enrolled participants included in the analysis?† S5.2: If data were missing, were appropriate methods used?‖ S5.3: Were appropriate methods used to account for censoring?‖ S5.4: In case of competing events, were appropriate methods used to account for them?‖
Risk of bias (high, low, unclear)	Could the selection of participants have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could measurement of the outcome have introduced bias?	Could the study flow have introduced bias?	Could the analysis have introduced bias?

Figure 1. QUAPAS.

S1 Text
NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
(adapted for cross sectional studies)

Selection: (Maximum 5 stars)

1) Representativeness of the sample:
   a) Truly representative of the average in the target population. * (all subjects or random sampling)
   b) Somewhat representative of the average in the target population. * (non-random sampling)
   c) Selected group of users.
   d) No description of the sampling strategy.
2) Sample size:
   a) Justified and satisfactory. *
   b) Not justified.
3) Non-respondents:
   a) Comparability between respondents and non-respondents characteristics is established, and the response rate is satisfactory. *
   b) The response rate is unsatisfactory, or the comparability between respondents and non-respondents is unsatisfactory.
   c) No description of the response rate or the characteristics of the responders and the non-responders.
4) Ascertainment of the exposure (risk factor):
   a) Validated measurement tool. **
   b) Non-validated measurement tool, but the tool is available or described.*
   c) No description of the measurement tool.

Comparability: (Maximum 2 stars)

1) The subjects in different outcome groups are comparable, based on the study design or analysis. Confounding factors are controlled.
   a) The study controls for the most important factor (select one). *
   b) The study control for any additional factor. *

Outcome: (Maximum 3 stars)

1) Assessment of the outcome:
   a) Independent blind assessment. **
   b) Record linkage. **
   c) Self report. *
   d) No description.
2) Statistical test:
   a) The statistical test used to analyze the data is clearly described and appropriate, and the measurement of the association is presented, including confidence intervals and the probability level (p value). *
   b) The statistical test is not appropriate, not described or incomplete.

Figure 2 Modified NOS.

References:

1. Lee J, Mulder F, Leeflang M, et al. QUAPAS: An Adaptation of the QUADAS-2 Tool to Assess Prognostic Accuracy Studies. Ann Intern Med. 2022 Jul;175(7):1010-1018. doi: 10.7326/M22-0276. Epub 2022 Jun 14. PMID: 35696685.

2. Modesti PA, Reboldi G, Cappuccio FP, et al. Panethnic Differences in Blood Pressure in Europe: A Systematic Review and Meta-Analysis. PLoS One. 2016 Jan 25;11(1):e0147601. doi: 10.1371/journal.pone.0147601. PMID: 26808317; PMCID: PMC4725677.

3. Baccolini V, Isonne C, Salerno C, et al. The association between adherence to cancer screening programs and health literacy: A systematic review and meta-analysis. Prev Med. 2022 Feb;155:106927. doi: 10.1016/j.ypmed.2021.106927. Epub 2021 Dec 23. PMID: 34954244.