PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Systematic Review and Meta-Analysis of the Diagnostic Value of Four Biomarkers in Detecting Neonatal Sepsis in Low- and Middle-Income Countries
AUTHORS	Rees, Chris A Lim, Jamie Westbrook, Adrianna L El Helou, Rachelle Schmid, Alexis Rubin-Smith, Julia Shreeve, Kyra Rotman, Chloe Govindapillai, Sindu Dorney, Kate Niescierenko, Michelle

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Joan L Robinson Institution and Country: Faculty of Medicine & Dentistry, University of Alberta Canada Competing interests: None
REVIEW RETURNED	10-Aug-2022

GENERAL COMMENTS	Comments to the Author I recently reviewed the same manuscript for another journal. As far as I can tell, the authors implemented only two of my minor suggestions. I would have expected them to at least have paid attention to suggestions to make their meaning more clear. The remaining suggestions are: The authors performed a systematic review of CRP, WBC count, ESR and PCT in neonatal sepsis and in pediatric pneumonia, doing a sub-group analysis of LMIC versus high income countries. The manuscript is remarkably concise given the amount of data that are presented. It would have been a tremendous amount of work to do this study. Major points 1. Most pneumonia in young children is viral. Is it not more useful to look at biomarkers for bacterial versus viral pneumonia rather than pneumonia versus "no pneumonia" since one main goal is to decrease inappropriate antibiotic use? A study of the latter comparison was recently published in JPIDS (Gunaratnam et al. .Systematic Review and Meta-Analysis of Diagnostic Biomarkers for Pediatric Pneumonia J Pediatric Infect Dis Soc. . 2021 Oct 27;10(9):891-900. doi: 10.1093/jpids/piab043.). Given this major concern and the fact that the combination of sepsis in neonates and pneumonia in all age groups is not very logical, I would recommend looking only at neonatal sepsis in this manuscript. I will not comment further on the pneumonia sections of the manuscript. 2. The search only included articles up to December 2020. That is 18 months ago now. Many journals will not accept systematic reviews if the search is more than 12 months old.

	3. A prior report compiled results of studies that directly compared CRP to PCT for diagnosis of neonatal sepsis (J Perinatol. 2019 Jul;39(7):893-903. doi: 10.1038/s41372-019-0363-4.). That is probably a better way to look at the utility of CRP and PCT than are the indirect comparisons in the current study 4. I am far from being an expert about the Youden index. However, using the formula in Youdens' original paper (Index for rating diagnostic tests (wiley.com)), one should not be able to obtain a negative result. The Wikipedia page states: "While it is technically possible to obtain a value of less than zero from this equation, e.g. Classification yields only False Positives and False Negatives, a value of less than zero just indicates that the positive and negative labels have been switched. After correcting the labels the result will then be in the 0 through 1 range." 5. It looks like it is possible to determine the Youden optimal cut-point rather than just choosing the cut-point with the highest Youden index with the latter being restricted by the cut-points that were studied. 6. Biomarker performance may vary dramatically for early onset sepsis (where biomarkers may still reflect maternal values) versus late onset sepsis (where biomarkers no longer reflect maternal values). I suspect that this is why the 2020 JAMA study that the authors quote excluded early onset cases. Antibiotics given to the neonate prior to measurement of biomarkers may also influence results. A prior study excluded pre-treated patients (Crit Care 2018 Nov 21;22(1):316. doi: 10.1186/s13054-018-2236-1). It seems like the authors should at least mention these limitations. 7. The practical implications of the findings are not totally clear. Unfortunately, the biomarkers that were studied are not sufficiently sensitive to use them to rule out neonatal sepsis. This is not a novel finding. Minor points 1. I did not understand "A priori, up to two reference standards were included for each disease so that diagnostic accuracy could be meta-analyzed within each standard and for a composite reference standard of the two." Is there a better way to explain that? 2. It is up to the editor but I would think that SI units should be used throughout the manuscript as they are the units used in almost every country other than the United States. 3. In the tables, I did not understand "All country study income groups". I assumed that this was a compilation of data from all countries. However, in Table 1 looking at a CRP cut-off of 0.5, there were 804 cases in all countries and 664 in LMIC but no indication of where the other 140 cases are accounted for. 4. It looks like in the tables, the authors report studies that required a blood culture and those that required only a clinical diagnosis of sepsis and then combined the data in separate rows. For example, for CRP cut-point of 1.0 in all countries, there were 7948 with a positive blood culture and 323 with a clinical diagnosis so one row reports the expected total of 8271. However, for a CRP cut-point of 0.6, the group with clinical sepsis are not reported in a separate row. Why?

REVIEWER	Reviewer name: Javad Heshmati Institution and Country: United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	13-Aug-2022

GENERAL COMMENTS	Dear Editor This is a good manuscript reviewing "A Systematic Review and Meta-Analysis of the Diagnostic Value of Four Biomarkers in Detecting Neonatal Sepsis and Pneumonia Among Children". The subject of the manuscript is fully consistent with the aims and scope of the journal « BMJ Paediatrics Open». The research methodology

	is fully consistent with the aims declared by the authors. Their conclusions are also consistent with the set goals, however, some issues need to be reconsidered: - Please explain all abbreviations in the abstract and manuscript. Abstracts 1) Abstract should be informative, background did not explain the question of this review and the answer which authors search for it 2) Abstract should be informative, did they have any language or publication preference? 3) Keywords: are these keywords are Mesh terms? Word that serves as a keyword, as to the meaning of that condition must be a Mesh term Introduction The Introduction needs adjustments in order to answer these questions: - What are the uncertainties and conflicts that underlie the hypotheticals? - How important is the evidence of studies for the healthy individuals and patients? - What is the focused clinical question your review will address? Methods - List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. -I would suggest that authors include a section where all eligibility criteria as well as the PICO statement used are presented, as recommended by Preferred Reporting Items for Systematic Reviews and MetaAnalysis (PRISMA) guidelines. - Authors should include a paragraph where all the outcomes (primary and secondary) of this meta-analysis are clearly summarized. This way the reader can easily track down each outcome of interest.

REVIEWER	Reviewer name: Flom, Peter Institution and Country: Peter Flom Consulting, United States Competing interests: None
REVIEW RETURNED	17-Aug-20223

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. The general approach is fine, but I have some issues to resolve before I can recommend publication. p 7 bottom - I am not sure I follow why using Reitsma's method means you don't have to show I^2, The value of that statistic (or other measures of heterogeneity) is not simply one of calculation, but of whether it makes sense to combine the studies at all. Adjusting the CI for heterogeneity is fine, IF the studies should be combined. p. 8 top - How were the CI around the AUC calculated? There doesn't seem to be a consensus on the best way to do this. One suggestion is to bootstrap and use random forests (see

	https://www.r-bloggers.com/2019/08/how-to-get-an-auc-confidence-interval/) Others use a normal approximation (with large N) but may use complex formulas for the standard error (e.g. Hanley, J.A. and NcNeil, B.J. 1982. 'The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.' Radiology, Vol 148, 29-36.) I think there are other methods as well. I also think a good "future paper" would be to try to model risk of disease based on continuous values of the biomarkers, using logistic regression. I understand why the authors used cutoffs, but d doing so increases both type 1 and type 2 errors. Peter Flom

VERSION 1 – AUTHOR RESPONSE

November 17, 2022

Dear Dr. Escobedo and Prof. Choonara,

Thank you for the opportunity to revise and resubmit our article, “A Systematic Review and Meta-Analysis of the Diagnostic Value of Four Biomarkers in Detecting Neonatal Sepsis and Pneumonia Among Children” (bmjpo-2022-001627). In response to the recommendation from the editor and reviewer, we removed the reporting of our results on pneumonia, so we changed the title of our revised submission to “A Systematic Review and Meta-Analysis of the Diagnostic Value of Four Biomarkers in Detecting Neonatal Sepsis in Low- and Middle-Income Countries”. We have indicated where the corresponding changes were made in the revised manuscript. We are happy to address further questions that may arise.

We thank you very much for reviewing our revised submission. Please do not hesitate to contact me with any questions.

Gratefully,

Chris A. Rees, MD, MPH (on behalf of the co-authors)

Assistant Professor of Pediatrics and Emergency Medicine

Emory University School of Medicine

Email: chris.rees@emory.edu

Phone: 1-801-664-5280

Editor’s Comments

You MUST update your search

Author response: We appreciate the editor’s comment. As requested, we conducted an updated search. We have included this in the Methods under Data Sources as follows, “We searched the Medline, EMBASE, DARE, CINAHL, and Babelmesh databases on February 12, 2021 and conducted an

updated search on August 29, 2022. We extracted articles that were included in each of these databases from their inception through August 29, 2022." We have also updated all results to include the additional articles we included.

You MUST exclude pneumonia

Author response: We appreciate this comment and have excluded all analyses and data related to pneumonia as requested by the editor and reviewer 1.

See the comments of reviewer 1 that MUST be answered. Also the comments of reviewers 2 & 3

Author response: We appreciate the opportunity to respond to the thoughtful comments of the reviewers. We have responded to the reviewers comments as described below. Please let us know if there are additional questions.

You excluded 110 papers due to language. This is a major limitation and should be mentioned

Author response: We appreciate the editor's suggestion. We have added the following to the Limitations paragraph as a result, "Lastly, though we attempted to review articles in as many languages as our team was capable to, several articles were excluded from our analysis because they were not published in English, Spanish, French, German, Dutch, and Arabic. This may have introduced some selection bias, potentially excluding more articles reporting research conducted in LMICs where these languages are not spoken."

If you need more time let us know

Author response: We contacted the Editorial Office and requested an extension so we would have adequate time to update our search. Thank you for graciously offering us more time to perform these major revisions. Our manuscript has been greatly strengthened from responding to your comments as well as to those of the reviewers. We sincerely appreciate your insight.

Reviewer #1

Dr. Joan Robinson, Faculty of Medicine & Dentistry, University of Alberta

Comments to the Author

I recently reviewed the same manuscript for another journal. As far as I can tell, the authors implemented only two of my minor suggestions. I would have expected them to at least have paid attention to suggestions to make their meaning more clear. The remaining suggestions are:

Author response: We thank the reviewer for their thoughtful comments. We are glad to have the opportunity to respond to each of them with our submission to BMJ Paediatrics Open.

The authors performed a systematic review of CRP, WBC count, ESR and PCT in neonatal sepsis and in pediatric pneumonia, doing a sub-group analysis of LMIC versus high income countries. The manuscript is remarkably concise given the amount of data that are presented. It would have been a tremendous amount of work to do this study.

Author response: We thank the reviewer for acknowledging the importance of this work and its implication for improving the care of neonates in low- and middle-income countries.

Major points

1. Most pneumonia in young children is viral. Is it not more useful to look at biomarkers for bacterial versus viral pneumonia rather than pneumonia versus “no pneumonia” since one main goal is to decrease inappropriate antibiotic use? A study of the latter comparison was recently published in JPIDS (Gunaratnam et al. .Systematic Review and Meta-Analysis of Diagnostic Biomarkers for Pediatric Pneumonia J Pediatric Infect Dis Soc. . 2021 Oct 27;10(9):891-900. doi: 10.1093/jpids/piab043.). Given this major concern and the fact that the combination of sepsis in neonates and pneumonia in all age groups is not very logical, I would recommend looking only at neonatal sepsis in this manuscript. I will not comment further on the pneumonia sections of the manuscript.

Author response: As requested, in this revised submission we have excluded pneumonia. The differentiation of viral from bacterial pneumonia indeed is a challenge. We acknowledged our exclusion of articles related to pneumonia in the Methods under Inclusion and Exclusion Criteria as follows, “There were >20 studies that reported the test characteristics of the included biomarkers for pneumonia. However, these were not included in our manuscript because those studies did not differentiate viral from bacterial disease.”

2. The search only included articles up to December 2020. That is 18 months ago now. Many journals will not accept systematic reviews if the search is more than 12 months old.

Author response: As requested, we conducted an updated search. We have included this in the Methods under Data Sources as follows, “We searched the Medline, EMBASE, DARE, CINAHL, and Babelmesh databases on February 12, 2021 and conducted an updated search on August 29, 2022. We extracted articles that were included in each of these databases from their inception through August 29, 2022.” We have also updated all results to include the additional articles we included.

3. A prior report compiled results of studies that directly compared CRP to PCT for diagnosis of neonatal sepsis (J Perinatol. 2019 Jul;39(7):893-903. doi: 10.1038/s41372-019-0363-4.). That is probably a better way to look at the utility of CRP and PCT than are the indirect comparisons in the current study

Author response: Thank you for bringing this recent study to our attention. Our study differs from the one mentioned as our study uses meta-analysis techniques, which required at least 3 distinct studies that used the same threshold for the same biomarker. Unfortunately, there was no instance where at least 3 studies used the same thresholds for the same 2 different biomarkers making us unable to make direct comparisons in this analysis. The cited study averaged the sensitivity and specificity, respectively, for CRP and PCT, without regard to differing cut off values and sample sizes which influence the precision of individual sensitivities and specificities. Furthermore, the relationship between sensitivity and specificity is unaccounted for when calculating simple averages. Our study uses proper meta-analysis techniques which allow us to account for the relationship between sensitivity and specificity and provide proper weight to each included study.

In response to the reviewer's comment, we have acknowledged our inability to compare biomarkers in the Limitations paragraph as follows, "Most of the included studies did not assess all four biomarkers of interest, making unclear their comparative test characteristics in the same populations."

4. I am far from being an expert about the Youden index. However, using the formula in Youdens' original paper (Index for rating diagnostic tests (wiley.com)), one should not be able to obtain a negative result. The Wikipedia page states: "While it is technically possible to obtain a value of less than zero from this equation, e.g. Classification yields only False Positives and False Negatives, a value of less than zero just indicates that the positive and negative labels have been switched. After correcting the labels the result will then be in the 0 through 1 range."

Author response: We appreciate the reviewer's comment. We checked our code and once we updated our search and incorporated several additional articles, there were no longer negative values for the Youden's index.

5. It looks like it is possible to determine the Youden optimal cut-point rather than just choosing the cut-point with the highest Youden index with the latter being restricted by the cut-points that were studied.

Author response: We thank the reviewer for their comment. Youden's index was originally created to objectively determine which test was more diagnostically "accurate" when confronted with similar but different sensitivities and specificities. Since then, there have been several methods created to determine an optimal cut off or threshold of a continuous measure, such as a biomarker, based on Youden's index. In the context of our study design, there is a multivariate method created by Dr. Susanne Steinhauser et al. (Steinhauser, et al. BMC Med Res Methodol. 2016.) which aimed to use various cut points and their accompanying sensitivity and specificity to extrapolate an optimal threshold based on Youden's index and its respective diagnostic accuracy. We had originally chosen this method for analysis but found two issues that ultimately caused us to evaluate the biomarkers by cut points rather than estimate an optimal cut point: 1) the model was unable to converge due to too much heterogeneity that comes from evaluating many cut points at once or 2) the model estimated the optimal cut point to be out of the possible range for that biomarker and therefore unusable to practicing clinicians (example table included below). Thus, we did not include this in our manuscript.

CRP Cut Point (mg/L), ≥	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index	Area Under Receiver Operating Characteristic Curve*	Reference Standard(s)
All Included Studies					
Failure to Converge					Blood culture and clinical sepsis
Failure to Converge					Blood culture
Failure to Converge					Clinical sepsis
High Income Countries Only					
Failure to Converge					Blood culture and clinical sepsis

Failure to Converge	Blood culture			
Failure to Converge	Clinical sepsis			
Low Income Countries Only				
7234.237	0.69 (0.52, 0.81)	0.80 (0.69, 0.88)	0.49	0.80 (0.35, 0.94)
	Blood culture and clinical sepsis			
8199.016	0.69 (0.52, 0.82)	0.81 (0.69, 0.89)	0.50	0.80 (0.34, 0.95)
	Blood culture			
Failure to Converge	Clinical sepsis			

6. Biomarker performance may vary dramatically for early onset sepsis (where biomarkers may still reflect maternal values) versus late onset sepsis (where biomarkers no longer reflect maternal values). I suspect that this is why the 2020 JAMA study that the authors quote excluded early onset cases. Antibiotics given to the neonate prior to measurement of biomarkers may also influence results. A prior study excluded pre-treated patients (Crit Care 2018 Nov 21;22(1):316. doi: 10.1186/s13054-018-2236-1). It seems like the authors should at least mention these limitations.

Author response: We appreciate the reviewer's comment. We have added the following to the limitations in response: "It is possible that some studies included neonates that had been pre-treated with antibiotics, which could affect the level of biomarkers. Most included studies did not differentiate early from late onset neonatal sepsis. Biomarkers in early onset sepsis may reflect maternal values."

7. The practical implications of the findings are not totally clear. Unfortunately, the biomarkers that were studied are not sufficiently sensitive to use them to rule out neonatal sepsis. This is not a novel finding.

Author response: We thank the reviewer for their critical feedback. From a practical standpoint, given the diagnostic limitations, many of our colleagues in West and East Africa have traditionally relied on things like CRP or ESR to determine if antibiotics should be administered (or stopped) to neonates. That was one of the observations that led us to conduct this study. We have added language to this effect in the fourth paragraph of the Discussion as follows, "In practice in many resource-limited settings, elevations in biomarkers such as CRP and ESR are used to make decisions around the initiation of antibiotics for neonates. However, our study suggests that the sole reliance on a single biomarker to make such a decision may not have sufficient discriminatory value. The development and validation of clinical prediction models including historical findings, other risk factors, as well as biomarkers for neonatal sepsis in LMICs may enhance the diagnostic capabilities in such settings."

Minor points

1. I did not understand "A priori, up to two reference standards were included for each disease so that diagnostic accuracy could be meta-analyzed within each standard and for a composite reference standard of the two." Is there a better way to explain that?

Author response: We thank the reviewer for pointing to the need for clearer language. We have changed this sentence in the first paragraph under Statistical Analyses to read, “We reported the aggregate performance of each biomarker cut point with up to two reference standards in the same studies (e.g., blood culture or clinical sepsis) for neonatal sepsis and alone in cases in which ≥3 studies reported the same cut point.”

2. It is up to the editor but I would think that SI units should be used throughout the manuscript as they are the units used in almost every country other than the United States.

Author response: Per the reviewer’s suggestion, we have changed to SI units throughout the manuscript as follows:

CRP in mg/L

Procalcitonin in ng/mL

ESR in mm/hour

WBC in cells per cubic millimeter of blood

3. In the tables, I did not understand “All country study income groups”. I assumed that this was a compilation of data from all countries. However, in Table 1 looking at a CRP cut-off of 0.5, there were 804 cases in all countries and 664 in LMIC but no indication of where the other 140 cases are accounted for.

Author response: We have clarified the language in the tables to state “All Included Studies” instead of “All Study Country Income Groups”. Regarding the question about Table 1, the other 140 neonates (i.e., 804-664=140) came from 1-2 studies. In order to meaningfully report aggregated numbers in the tables, we required that ≥3 studies reported the same cut point. We have clarified this in the Methods in the second paragraph under Statistical Analyses as follows, “Many of the studies that met our inclusion criteria used different cut points for their respective biomarker. We looked at each cut point used by ≥3 studies individually, using a bivariate model created by Reitsma et al. through the reitsma function in the R package Mada.”

4. It looks like in the tables, the authors report studies that required a blood culture and those that required only a clinical diagnosis of sepsis and then combined the data in separate rows. For example, for CRP cut-point of 1.0 in all countries, there were 7948 with a positive blood culture and 323 with a clinical diagnosis so one row reports the expected total of 8271. However, for a CRP cut-point of 0.6, the group with clinical sepsis are not reported in a separate row. Why?

Author response: We appreciate the reviewer’s comment. This is due to the same issue raised in minor point 3 above. Please see our response above for clarification and how we changed the manuscript accordingly.

Reviewer #2

Javad Heshmati

Dear Editor

This is a good manuscript reviewing “A Systematic Review and Meta-Analysis of the Diagnostic Value of Four Biomarkers in Detecting Neonatal Sepsis and Pneumonia Among Children”. The subject of the manuscript is fully consistent with the aims and scope of the journal « BMJ Paediatrics Open». The research methodology is fully consistent with the aims declared by the authors. Their conclusions are also consistent with the set goals, however, some issues need to be reconsidered:

Author response: We thank the reviewer for the time they spent reviewing our work and for their thoughtful input. We believe that the suggested changes have strengthened our manuscript.

- Please explain all abbreviations in the abstract and manuscript.

Author response: We appreciate the reviewer’s comment. In response, we have ensured that all abbreviations are defined the first time they appear in the abstract and the full text.

Abstracts

1) Abstract should be informative, background did not explain the question of this review and the answer which authors search for it

Author response: We thank the reviewer for their suggestion. In response, we have made the Background clearer with the following text, “Biomarkers may enhance diagnostic capability for common pediatric infections, especially in low- and middle-income countries (LMICs) where standard diagnostic modalities are frequently unavailable, but disease burden is high. A comprehensive understanding of the diagnostic capability of commonly available biomarkers for neonatal sepsis in LMICs is lacking. Our objective was to systematically review evidence on biomarkers to understand their diagnostic performance for neonatal sepsis in LMICs.”

2)Abstract should be informative, did they have any language or publication preference?

Author response: We have added additional language to the Abstract as recommended by the reviewer. The Methods of the Abstract now read, “We conducted a systematic review and meta-analysis of studies published in English, Spanish, French, German, Dutch, and Arabic reporting the diagnostic performance of C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), white blood cell (WBC) count, and procalcitonin (PCT) for neonatal sepsis.”

3) Keywords: are these keywords are Mesh terms? Word that serves as a keyword, as to the meaning of that condition must be a Mesh term

Author response: We thank the reviewer for this comment. We have changed the Key Words to the following, “neonate; sepsis; biomarkers; procalcitonin; c reactive protein”. We have also verified that these show up as MeSH terms in the National Library of Medicine here: https://www.ncbi.nlm.nih.gov/mesh/

Introduction

The Introduction needs adjustments in order to answer these questions:

- What are the uncertainties and conflicts that underlie the hypotheticals?

Author response: We have added the following to the first paragraph in response to the reviewer's comment, "Neonatal sepsis is a common cause of neonatal morbidity and mortality in LMICs. Nonetheless, there is no unified criteria for the diagnosis of neonatal sepsis, which makes clinicians in resource-limited settings with scarce access to blood cultures rely on a clinical diagnosis."

- How important is the evidence of studies for the healthy individuals and patients?

Author response: We have added the following to the final paragraph of the Introduction in response to the reviewer's comment, "A comprehensive understanding of the diagnostic capability of commonly available biomarkers for neonatal sepsis in LMICs is lacking but may allow for more accurate diagnoses among neonates and more judicious antibiotic use."

- What is the focused clinical question your review will address?

Author response: We appreciate the reviewer's comment. We have included the following in the final paragraph of the Introduction in response, "Our objective was to systematically review existing evidence on the use of four biomarkers (CRP, ESR, WBC, and PCT) to understand their diagnostic performance against reference standards for neonatal sepsis, with a focus on studies conducted in LMICs."

Methods

- List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.

Author response: We appreciate the reviewer's recommendation. We have included the following in the second paragraph under Data Extraction and Risk of Bias Assessment, "We reviewed the full text of each article that was included after the initial phase of article title and abstract review. We extracted the following information from each included article: study location (e.g., outpatient, emergency department, inpatient, etc.), study design, study country, included patient ages, disease studied, biomarker(s) evaluated, reference standard, and study inclusion/exclusion criteria. Biomarkers were considered diagnostic if they were used to distinguish an infection in a child from healthy controls or children who had negative reference standard testing."

Regarding assumptions and simplifications made, we have included the following in the paragraph under Definitions, "We used the definitions used for neonatal sepsis reported in the included studies (i.e., either positive blood culture or clinical sepsis)."

-I would suggest that authors include a section where all eligibility criteria as well as the PICO statement used are presented, as recommended by Preferred Reporting Items for Systematic Reviews and MetaAnalysis (PRISMA) guidelines.

Author response: We thank the reviewer for pointing this out. Regarding the PICO statement, we have included the following in the final paragraph of the Introduction in response to the reviewer's comment, "Our objective was to systematically review existing evidence on the use of four biomarkers (CRP, ESR, WBC, and PCT) to understand their diagnostic performance against the reference standards of blood culture and clinical sepsis for neonatal sepsis, with a focus on studies conducted in LMICs."

Regarding the reviewer's comment about eligibility criteria, we have included the inclusion criteria in the following section pasted here for ease of reference.

"Inclusion and Exclusion Criteria

We included studies that met the following criteria 1) were peer-reviewed, original research articles published from the inception of each database through August 29, 2022, 2) evaluated the use of one of the four biomarkers of interest in the diagnosis of an infectious disease, 3) included participants aged 0 to 18 years old, and 4) included a control group that did not test positive with a reference standard as a comparison for the diagnostic performance of the biomarkers evaluated. Initially, our search was not restricted to specific diseases. However, post hoc, we decided to focus our analysis on neonatal sepsis as there were at least 20 studies that met our inclusion criteria, and it contributes to a large burden of childhood morbidity and mortality globally. There were >20 studies that reported the test characteristics of the included biomarkers for pneumonia. However, these were not included in our manuscript as most pneumonia in children is viral and the included studies did not differentiate viral from bacterial disease.

We excluded studies that met any of the following criteria: 1) articles that were not published in English, Spanish, French, German, Dutch, or Arabic, 2) abstracts without full text, 3) articles that only included highly medicalized populations, 4) articles reporting only mean or median values for biomarkers, 5) articles that did not evaluate children separately if adults >18 years were included, 6) articles that only assessed changes in biomarkers during treatment, and 7) case reports, editorials, study protocols, review articles, systematic reviews, and meta-analyses. We reviewed systematic reviews and meta-analyses for other articles reporting primary data our initial query did not capture. Any potential articles identified therein were included if they met inclusion criteria."

- Authors should include a paragraph where all the outcomes (primary and secondary) of this meta-analysis are clearly summarized. This way the reader can easily track down each outcome of interest.

Author response: We appreciate the reviewer's comment. We have included the following in the Definitions paragraph of the Methods in response, "We used the definitions used for our outcome of neonatal sepsis as reported in the included studies (i.e., either positive blood culture or clinical sepsis)."

Reviewer #3

Dr. Peter Flom, Peter Flom Consulting

I confine my remarks to statistical aspects of this paper. The general approach is fine, but I have some issues to resolve before I can recommend publication.

Author response: We thank the reviewer for the time they spent reviewing our work and for their thoughtful input. We believe that the suggested changes have strengthened our manuscript.

p 7 bottom - I am not sure I follow why using Reitsma's method means you don't have to show I^2, The value of that statistic (or other measures of heterogeneity) is not simply one of calculation, but of whether it makes sense to combine the studies at all. Adjusting the CI for heterogeneity is fine, IF the studies should be combined.

Author response: We appreciate the reviewer's comment. Given the lack of a consensus on how to best measure heterogeneity for diagnostic accuracy meta-analyses and that the bivariate method by Reitsma employs a linear mixed model which calculates confidence intervals to adjust for heterogeneity, we had originally opted to not display I2 values. However, in response to the reviewer's comment, we have added Holling sample size adjusted I2 values for each model so that readers are given the proper information needed to fully appreciate the results of this meta-analysis.

As already mentioned, there is no one right way to calculate heterogeneity for this type of study. The traditional Higgin's I2 has shown to be inflated when evaluating large studies and was not developed for bi- or multivariate analyses. We therefore chose to use Holling's sample size adjusted I2, which was developed for diagnostic accuracy meta-analyses and accounts for within-study sample sizes to mediate the inflation issues seen when evaluating large studies. These references detail the issues with common measures of heterogeneity and introduce the method we use in this current study:

Heinz Holling, Walailuck Böhning, Ehsan Masoudi, Dankmar Böhning & Patarawan Sangnawakij (2020) Evaluation of a new version of I2 with emphasis on diagnostic problems, Communications in Statistics - Simulation and Computation, 49:4, 942-972, DOI: 10.1080/03610918.2018.1489553

Sangnawakij P, Böhning D, Niwitpong SA, Adams S, Stanton M, Holling H. Meta-analysis without study-specific variance information: Heterogeneity case. Stat Methods Med Res. 2019 Jan;28(1):196-210. doi: 10.1177/0962280217718867. Epub 2017 Jul 6. PMID: 28681700.

We have updated the methods according to the reviewer's comment, "We also calculated Holling's sample size adjusted measure for heterogeneity (I2), which was developed for use in bivariate meta-analyses of diagnostic accuracy." Additionally, all tables now include I2 values as requested.

p. 8 top - How were the CI around the AUC calculated? There doesn't seem to be a consensus on the best way to do this. One suggestion is to bootstrap and use random forests (see https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.r-bloggers.com%2F2019%2F08%2Fhow-to-get-an-auc-confidence-interval%2F&data=05%7C01%7Cchris.rees%40emory.edu%7C624d5722e2d0418c145c08da81b092a7%7Ce004fb9cb0a4424fbcd0322606d5df38%7C0%7C0%7C637964892618588506%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=7hNdy4zeu1U5BrBkgQ5ZoSnEMh9MHWA%2BNp5f8l3ENL8%3D&reserved=0) Others use a normal approximation (with large N) but may use complex formulas for the standard error (e.g. Hanley, J.A. and NcNeil, B.J. 1982. 'The Meaning and Use of the Area under a Receiver Operating Characteristic

(ROC) Curve.' Radiology, Vol 148, 29-36.) I think there are other methods as well.

Author response: We thank the reviewer for this question. We used a bootstrapping procedure through dmetatools in R with 2,000 resamplings (according to Noma H, et al. Communications in Statistics: Case Studies, Data Analysis and Applications. 2021.). This method was developed for use with summary ROC and is complementary to Reitsma's bivariate method that we used for this study. We have updated the second paragraph of the Analyses in the Methods in response to the

reviewer's comment, "95% confidence intervals (CIs) for AUCs were calculated through bootstrapping with 2,000 resamplings via the AUC boot function in the dmetatools R package created by Noma, et al."

I also think a good "future paper" would be to try to model risk of disease based on continuous values of the biomarkers, using logistic regression. I understand why the authors used cutoffs, but d doing so increases both type 1 and type 2 errors.

Author response: We appreciate the reviewer's comment, and we agree that this is certainly a nice future study. As the reviewer pointed out, given our data source which does not include continuous data, we are unable to perform such an analysis in this manuscript.