PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	PROGNOSTIC FACTORS FOR STREPTOCOCCAL TOXIC SHOCK SYNDROME: SYSTEMATIC REVIEW AND META-ANALYSIS
AUTHORS	Bartoszko, Jessica; Elias, Zeyad; Rudziak, Paulina; Lo, Carson KL; Thabane, Lehana; Mertz, Dominik; Loeb, Mark

VERSION 1 – REVIEW

REVIEWER	Chis Ster, Irina St. George's University of London
REVIEW RETURNED	26-Apr-2022

GENERAL COMMENTS	I think the authors address a question of interest. Nevertheless, I found the conclusions too optimistic given the great deal of uncertainty of these studies. The meta-analyses are based on many studies with 0 events - such studies do not provide any indication of either the direction or magnitude of the relative effect (OR). Also, many studies are subjected to a very small number if events and therefore require great caution with their analyses. I refer to Peto method for meta-analysis and also to the paper at this link (https://ebmh.bmj.com/content/21/2/72). Classical random effects analyses applied by the authors would produce biased results. I am not entirely sure that aggregated type data analysis produce estimates which can be interpreted as modifiable factors aiming at lowering mortality. The authors should engage a professional statistician who should shed a light on the analyses (authorship included).

REVIEWER	Taylor, Kathryn University of Oxford, Nuffield Dept of Primary Care Health Sciences
REVIEW RETURNED	28-May-2022

GENERAL COMMENTS	This manuscript is mostly well written and the methods are sound but some of the descriptions are unclear and the presentation of the results could be improved. I arrange my comments chronologically. Abstract 1. Reporting that pairs of reviewers screened records without stating that they did so independently suggests that perhaps they did not screen independently so I would suggest adding the word “independently”. 2, The CI abbreviation for confidence interval is standard in BMJ Open and does not require to be defined.

	3. General point for abstract and main text: I appreciate that all the studies had high risk of bias and low certainty of evidence and the need to be transparent but I not think it is helpful to combine the two assessments repeatedly in statements like “Low certainty evidence suggests the odds of mortality may be…..”. It jars and makes me less interested in reading on. It would to separate the assessments to say something like: We found a statistically significant association between factor X and mortality (quoting OR with CI, n=) but the certainty of evidence was low. 4. Strengths and limitations statement – The authors write: “These strengths directly address limitations of a narrative synthesis of STSS prognosis restricted to the critical care setting”. It would be better to say “..of an existing narrative synthesis..” or otherwise it begs the question of why a narrative synthesis is being mentioned. Introduction 5. The GAP abbreviation would be clearer if “group” had a capital G. Methods 6. The authors reported that to resolve disagreements in extracting data, where necessary, there was adjudication “by a third party” and for disagreements in applying QUIPs, there was adjudication “by a senior co-investigator” – was there a difference in the adjudicator or if not, it would better to describe them in the same way. 7. The authors report modifying QUIPS and GRADE. It would be helpful to the reader to be explicit about what modification was made and why i.e. “We modified the QUIPS tool by…….. because….” 8. The pooling of data from two studies is only valid if the two studies are similar. I would suggest that two studies reporting the same prognostic factor and outcome is not sufficient – How similar were the patient populations of the two studies? 9. The description of the I-squared categories is confusing with the overlapping ranges of 50% to 90% and 75% to 100% and with their description as “substantial” and “considerable” respectively when these terms are generally used synonymously. 10. A reference should be given for the imputation of a mean by the median and using the IQR to estimate the SD. For the latter the equation should be given. As this is an estimate 11. Age is a continuous variable so what were the categories? 12. I don’t see the value of including Table 2 in the main text given that all studies have high overall risk of bias. BMJ Open allows 5 figures and tables but the authors have only included 3 tables and 1 figure. I would suggest moving a couple of forest plots into the main text e.g. IVIG in all STSS patients and IVIG in STSS patients treated with clindamycin. In fact these could be merged to a single forest plot with subgroups “with clinidamycin” Results 13. Figure 1 is confusing as the removal of duplicates is not shown and therefore the numbers do not add up. Also the changes in the total number of records due to the study’s update to the search is not interesting and would be better removed as it adds clutter (this is not a systematic review that is updating a previously published systematic review). Table 1 14. Using indentation for geographical regions, study designs etc would made the table easier to read. 15. Probable and confirmed STSS patients do not need “N (%)” as the title of the table already says numbers (percentages) can be assumed unless otherwise stated.

	Forest plots 16. It would be helpful to label the axes “worse STSS prognosis” and “better STSS prognosis”. 17. ne, Ne, nc and Nc – why are some not reported in the forest plots? 18. Ne, ne, Nc and nc would be a more natural order. 19. For consistency mean and SD should be shown in the forest plots for continuous outcomes. Discussion 20. “statistical evidence” is too vague. The authors should refer to statistical significance. 21. The authors refer to the lack for cohort studies and RCTs in passing but, given the high risk of bias of studies and low certainty evidence, in conclusion, the authors should be calling for high quality cohort studies and RCTs to investigate prognostic factors for STSS. 22. The pooled result for age at or over 65 years features in the conclusion along with the result for other pooled results (for NSAIDs, clindamycin and IVIG in clindamycin) but the result for age is only based on 2 studies whilst the results for the other factors are based on 4 to 9 studies. The result based on 2 studies should be given less weight for the reasons I give above. 23. The authors report: “These findings support the use of IVIG as an adjunctive treatment in clindamycin-treated STSS patients.” All the studies have high risk of bias and the certainty of evidence is low and therefore I do not see the justification for such an upbeat conclusion.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Irina Chis Ster, St. George's University of London
Comments to the Author:

1. *I think the authors address a question of interest.*

Nevertheless, I found the conclusions too optimistic given the great deal of uncertainty of these studies. The meta-analyses are based on many studies with 0 events - such studies do not provide any indication of either the direction or magnitude of the relative effect (OR).

Response: Thank you for your thoughtful review of our manuscript. To Dr. Ster’s comment that the conclusions are too optimistic, we have updated the abstract and main text to better reflect the low certainty of evidence. Further, we have removed “modified” from the abstract conclusion. The changes are as follows in the revised manuscript:

Abstract (Page 2): “Treatment with clindamycin and within clindamycin-treated patients, IVIG, was each significantly associated with mortality, but the certainty was evidence was low. Future research should focus on morbidity post-infection in STSS survivors.”

Discussion (Page 18, Para 2): “Although these findings support the use of IVIG as an adjunctive treatment in clindamycin-treated STSS patients, the certainty of evidence was low due to serious risk of bias and imprecision.”

2. *Also, many studies are subjected to a very small number if events and therefore require great caution with their analyses. I refer to Peto method for meta-analysis and also to the paper at this link (https://ebmh.bmj.com/content/21/2/72).*

Response: We agree with Dr. Ster that the small number of events is a limitation to the data. To mitigate the risk of misinterpreting our findings from the inclusion of small studies with few events, we present absolute effect estimates in addition to relative effect estimates in the manuscript. We have elaborated on this in the Discussion section as follows (Page 17, Para 1):

"Small numbers of events further contributed to the imprecision around summary estimates and limited the interpretation of our findings. With few participants and events, minor changes in the data can cause major changes in the results. In such instances, results can be exaggerated by the presentation of relative effect estimates only. To minimize the risk of misinterpreting results from the inclusion of small studies in our meta-analyses, we calculated an absolute effect estimate for each relative effect estimate (table 2)."

3. *Classical random effects analyses applied by the authors would produce biased results.*
I am not entirely sure that aggregated type data analysis produce estimates which can be interpreted as modifiable factors aiming at lowering mortality.
The authors should engage a professional statistician who should shed a light on the analyses (authorship included).

Response: It is important to note that not all of the meta-analyses in our review were limited by few events, only a select few. One example is the NSAIDs-mortality meta-analysis, in which two of the four included studies in each comparison group had zero events. Dr. Ster proposes the Peto method for meta-analysis is better suited for such situations compared to the DerSimonian and Laird (DL) meta-analysis method we used. Thus, based on guidance from Dr. Lehana Thabane, a professional biostatistician and co-author on this paper, we performed a post-hoc sensitivity analysis with the NSAIDs-mortality data comparing results from the Peto method to the original DL method. We present our sensitivity analysis findings below.

Original DL method: OR 4.14 (95% CI 1.13 to 15.14)
Peto method: OR 5.42 (95% CI 1.45 to 20.27)

The direction of both point estimates is the same and both confidence intervals do not cross one. Because results are consistent, we believe it is appropriate to continuing reporting results from our original meta-analysis method in the manuscript.

We have updated the Methods and Results sections of the manuscript text to reflect our sensitivity analysis findings as follows.

Methods (Page 9, Para 1): "The analysis plan included performing subgroup analyses of STSS patients treated with clindamycin vs no clindamycin, STSS patients with necrotizing fasciitis vs without necrotizing fasciitis, age (0 to 17 years old vs 18 to 64 years old vs 65 years old or older), sex (male vs female) and risk of bias (high vs moderate vs low) when at least two studies were present for each subgroup. Because select meta-analyses were limited by small numbers of events, we performed a post-hoc sensitivity analysis using the Peto method for meta-analysis, which is recommended for meta-analysis of rare events [32], and compared the results to those from the DerSimonian and Laird method we applied in this review."

Results (Page 16, Para 1) : "We also found no statistical evidence that the association between sex and mortality differed between studies with patients <18 years and patients 18-64 years

(p=0.666). Because results were consistent across Peto, and DerSimonian and Laird methods, our post-hoc sensitivity analysis showed that our meta-analyses based on few events were robust."

Reviewer: 2
Dr. Kathryn Taylor, University of Oxford
Comments to the Author:

4. *This manuscript is mostly well written and the methods are sound but some of the descriptions are unclear and the presentation of the results could be improved. I arrange my comments chronologically.*

Response: Thank you for your thoughtful review of our manuscript.

Abstract

5. *Reporting that pairs of reviewers screened records without stating that they did so independently suggests that perhaps they did not screen independently so I would suggest adding the word "independently".*

Response: We agree with Dr. Taylor and have made modified the abstract text as follows (page 2).

"We searched MEDLINE and EMBASE from inception to 6 August 2021, along with CINAHL from inception to 16 September 2021 and citations of included studies. Pairs of reviewers independently screened potentially eligible studies of patients with GAS-induced STSS that quantified the association between at least one prognostic factor and outcome of interest."

6. *The CI abbreviation for confidence interval is standard in BMJ Open and does not require to be defined.*

Response: We have removed the definition of CI from the abstract text (page 2).

"One randomized trial and 39 observational studies were eligible (n=1,914 patients). Low certainty evidence suggests the odds of mortality may be significantly reduced by clindamycin treatment (n=144; odds ratio [OR] 0.14, 95% CI 0.06 to 0.37) and within clindamycin-treated STSS patients…"

7. *General point for abstract and main text: I appreciate that all the studies had high risk of bias and low certainty of evidence and the need to be transparent but I not think it is helpful to combine the two assessments repeatedly in statements like "Low certainty evidence suggests the odds of mortality may be…..". It jars and makes me less interested in reading on. It would to separate the assessments to say something like: We found a statistically significant association between factor X and mortality (quoting OR with CI, n=) but the certainty of evidence was low.*

Response: Where relevant, we have modified the abstract and main text as per Dr. Taylor's comment.

Abstract (Page 2): "One randomized trial and 39 observational studies were eligible (n=1,914 patients). We found a statistically significant association between clindamycin treatment and

mortality (n=144; odds ratio [OR] 0.14, 95% CI 0.06 to 0.37), but the certainty of evidence was low. Within clindamycin-treated STSS patients, we found a statistically significant association between intravenous immunoglobulin (IVIG) treatment and mortality (n=188; OR 0.34, 95% CI 0.15 to 0.75), but the certainty of evidence was also low. The odds of mortality may increase in patients ≥65 years when compared to patients 18-64 years (n=396; OR 2.37, 95% CI 1.47 to 3.84), but the certainty of evidence was low."

Main text (Page 11): Eleven prognostic factors from 31 studies including 1339 patients were eligible for analysis (table 3, supplementary data). We found a statistically significant association between clindamycin treatment and mortality (n=144; odds ratio [OR] 0.14, 95% CI 0.06 to 0.37), but the certainty of evidence was low. Within clindamycin-treated STSS patients, we found a statistically significant association between intravenous immunoglobulin (IVIG) treatment and mortality (n=188; OR 0.34, 95% CI 0.15 to 0.75), but the certainty of evidence was also low. We are uncertain whether IVIG treatment reduces the odds of mortality in all STSS patients regardless of concurrent clindamycin treatment (n=365, OR 0.37, 95% CI 0.17 to 0.80; very low certainty of evidence). The odds of mortality may increase in patients ≥65 years when compared to patients 18-64 years (n=396; OR 2.37, 95% CI 1.47 to 3.84), but the certainty of evidence was low. We are less certain whether the same is true for patients ≥65 years compared to patients <18 years (n=136, OR 10.66, 95% CI 1.28 to 88.57; very low certainty of evidence).

8. *Strengths and limitations statement – The authors write: "These strengths directly address limitations of a narrative synthesis of STSS prognosis restricted to the critical care setting". It would be better to say "..of an existing narrative synthesis.." or otherwise it begs the question of why a narrative synthesis is being mentioned.*

Response: We have modified the text as per Dr. Taylor's comment (page 3).

"These strengths directly address limitations of an existing narrative synthesis of STSS prognosis restricted to the critical care setting."

*Introduction *

9. *The GAP abbreviation would be clearer if "group" had a capital G.*

Response: We agree and have modified the text as per Dr. Taylor's comment (page 4, para 1).

"Streptococcal toxic shock syndrome (STSS) is an acute and severe life-threatening complication of predominantly invasive Group A *Streptococcus* (GAS) infections."

Methods

10. The authors reported that to resolve disagreements in extracting data, where necessary, there was adjudication "by a third party" and for disagreements in applying QUIPs, there was adjudication "by a senior co-investigator" – was there a difference in the adjudicator or if not, it would better to describe them in the same way.

Response: We agree and have updated the methods text to describe adjudication in the same way – by a senior co-investigator (page 7, para 1).

"We used the proportions to calculate crude ORs when no adjusted ORs were provided. Reviewers resolved discrepancies by discussion and, when necessary, with adjudication by a senior co-investigator."

11. The authors report modifying QUIPS and GRADE. It would be helpful to the reader to be explicit about what modification was made and why i.e. "We modified the QUIPS tool by…….. because…."

Response: For QUIPS, we specified that a "modified version" was used because not all prompts were relevant in assessing risk of bias in each study. Further, regardless of prompt responses, we rated all case-series at high risk of bias overall. This is not explicit guidance provided by the QUIPS tool. For GRADE, we specified rules to facilitate rating the certainty of evidence for each domain (detailed in the supplement). These are not explicit rules provided by GRADE. Thus, these are not modified versions, but additional considerations made when applying each tool/framework. To clarify, we have removed the reporting of "modified version" and updated the text as follows (page 7, para 2).

"Following training and calibration exercises, reviewers, independently and in duplicate, used the Quality in Prognosis Studies (QUIPS) tool to rate each prognostic factor and outcome combination at low, moderate or high risk of bias. Based on prespecified sets of questions, we assessed risk of bias across the following domains: participation, attrition, prognostic factor measurement, outcome measurement, confounding, and statistical analysis and reporting [25]. For studies addressing more than one prognostic factor and outcome combination, we reported the highest risk of bias rating among the prognostic factor and outcome combinations within a study for each domain. In addition to assessing risk of bias at the domain-level as outlined in the QUIPS tool, we applied the following rules to assess risk of bias overall at the study-level. We rated overall study risk of bias as low if the study was prospective and five or more domains were assessed as low risk of bias, and high if two or more domains were assessed as high risk of bias. All other studies were rated as moderate risk of bias overall. Due to high risk of selection bias and residual confounding, we rated all case series as high risk of bias overall… This approach considers whether confidence intervals include the null effect. The supplementary file presents the detailed guidance we developed to facilitate the certainty of the evidence assessment in this review."

12. The pooling of data from two studies is only valid if the two studies are similar. I would suggest that two studies reporting the same prognostic factor and outcome is not sufficient – How similar were the patient populations of the two studies?

Response: We meta-analyzed data from studies that enrolled patients with GAS-induced STSS. The patient populations were similar in this regard. We have clarified the criterion we used for meta-analysis as per Dr. Taylor's comment (page 8, para 2).

"When at least two included studies reported on the same prognostic factor and outcome in patients with GAS-induced STSS, we conducted…"

13. The description of the I-squared categories is confusing with the overlapping ranges of 50% to 90% and 75% to 100% and with their description as "substantial" and "considerable" respectively when these terms are generally used synonymously.

Response: Our methods, to consider "substantial" and "considerable" heterogeneity separately, are consistent with the Cochrane Handbook (https://training.cochrane.org/handbook/current/chapter-10). We have clarified our method to address an I-squared value that is in the range of overlapping I-squared values as follows (page 8, para 2).

"We interpreted an I2 statistic value of 0%-40%, 30%-60%, 50%-90%, or 75%-100% as not likely important, moderate, substantial, or considerable heterogeneity, respectively [29]. If an I2 statistic value was within a range of overlapping values (e.g. 80%), we would interpret heterogeneity as more important (e.g. considerable instead of substantial) if we observed inconsistent magnitudes and

directions of summary estimates upon visual inspection of the forest plots, and the chi-square test was significant [29]."

14. A reference should be given for the imputation of a mean by the median and using the IQR to estimate the SD. For the latter the equation should be given. As this is an estimate

Response: In the main text, we have included two references containing the equations we used for the imputations listed above (page 8, para 2). The equation to calculate SD from IQR was SD= (P75-P25)/1.35.

"For meta-analyses of continuous outcomes, we imputed means and standard deviations for studies reporting medians and (interquartile) ranges, respectively [30, 31]."

15. Age is a continuous variable so what were the categories?

Response: We have clarified in the main text as follows (page 8, para 3).

"We planned to perform a regression analysis for each study for which age was reported at the patient level to generate a study and age category (0 to 17 years old vs 18 to 64 years old vs 65 years old or older) specific OR that could be used in meta-analysis when a study had at least 10 observations for continuous outcomes and 10 events for dichotomous outcomes; however, no study met the sample size or event number requirements."

16. I don't see the value of including Table 2 in the main text given that all studies have high overall risk of bias. BMJ Open allows 5 figures and tables but the authors have only included 3 tables and 1 figure. I would suggest moving a couple of forest plots into the main text e.g. IVIG in all STSS patients and IVIG in STSS patients treated with clindamycin. In fact these could be merged to a single forest plot with subgroups "with clinidamycin"

Response: We have removed Table 2 from the main text and added it to the supplementary file (page 25 to 26). Further, we have added the two forest plots Dr. Taylor suggested into the main text as Figures 2A and 2B, respectively.

*Results *

17. Figure 1 is confusing as the removal of duplicates is not shown and therefore the numbers do not add up. Also the changes in the total number of records due to the study's update to the search is not interesting and would be better removed as it adds clutter (this is not a systematic review that is updating a previously published systematic review).

Response: We have added a box explicitly reporting the number of duplicates removed (n=7,729). We have also updated the total number of records as per Dr. Taylor's comment.

*Table 1 *

18. Using indentation for geographical regions, study designs etc would made the table easier to read.

Response: We have used indentation in table 1 as per Dr. Taylor's comment (please see below).

19. Probable and confirmed STSS patients do not need "N (%)" as the title of the table already says numbers (percentages) can be assumed unless otherwise stated.

Response: We have removed "N (%)" from table 1 as per Dr. Taylor's comment (please see below).

Characteristics	**(40 studies, 1914 patients)**
Range of publication year	1989 to 2021
Median (IQR) No. of patients	11 (5 to 29)
Geographical region:	
North America	19 (48)
Europe	14 (35)
Central/South America	0 (0)
Asia	3 (7)
Other	4 (10)
Study design:	
Randomized trial	1 (2)
Cohort	19 (48)
Case-control	2 (5)

Case-series	18 (45)
Case definition:	
Probable STSS patients	115 (6)
Confirmed STSS patients	223 (12)
Prognostic factor type:	
Demographic	28 (70)
Medical history	5 (13)
Early disease	11 (28)
Treatment	16 (40)

Forest plots

20. It would be helpful to label the axes "worse STSS prognosis" and "better STSS prognosis".

Response: In the supplementary file, we have provided the following explanation under the "Forest plots" header.

"For mortality, ICU admission and need for mechanical ventilation outcomes, an odds ratio greater than 1 corresponds with a worse STSS prognosis and an odds ratio less than 1 corresponds with a better STSS prognosis.

For clinical cure or improvement outcome, an odds ratio greater than 1 corresponds with a better STSS prognosis and an odds ratio less than 1 corresponds with a worse STSS prognosis.

For duration of hospitalization and ICU stay outcomes, a mean difference greater than 1 corresponds with a worse STSS prognosis and a mean difference less than 1 corresponds with a better STSS prognosis."

We have added these labels to the two forest plots we moved to the main manuscript (i.e. IVIG-mortality in all STSS patients and IVIG-mortality in STSS patients treated with clindamycin). Please see figures 2A and 2B, respectively.

21. ne, Ne, nc and Nc – why are some not reported in the forest plots?

Response: Proportions are blank for study rows for which we meta-analyzed adjusted odds ratios instead of crude proportions. We have clarified this in the Figure 2 heading in the main text of the manuscript.

"…Please note proportions are blank for study rows where we meta-analyzed adjusted odds ratios instead of crude proportions."

22. Ne, ne, Nc and nc would be a more natural order.

Response: We have opted to continue reporting numerator first and denominator second because we believe this is a more natural order and is how fractions/proportions will be understood by readers.

23. For consistency mean and SD should be shown in the forest plots for continuous outcomes.

Response: Three studies reported on duration of hospital stay for the prognostic factor-outcome combination IVIG-mortality. Of these, one reported on the mean difference already and one had the mean imputed. Two studies reported on duration of ICU stay. Of these, one reported on the mean

difference already. Because not all studies reported mean (SD) across comparison groups (IVIG vs no IVIG), we have opted to not report mean and SD in the forest plots for continuous outcomes.

Discussion

24. "statistical evidence" is too vague. The authors should refer to statistical significance.

Response: We have updated the main text as per Dr. Taylor's comment (page 16, para 2).

"Prognostic factors for which there was a statistically significant association with mortality in STSS patients were…"

25. The authors refer to the lack for cohort studies and RCTs in passing but, given the high risk of bias of studies and low certainty evidence, in conclusion, the authors should be calling for high quality cohort studies and RCTs to investigate prognostic factors for STSS.

Response: We have updated the main text of the Discussion section as per Dr. Taylor's comment (page 17, para 1).

"Creation of an international registry of STSS patients may improve the credibility of prognostic evidence for STSS, and facilitate the conduct of high-quality cohort studies. Although we…"

26. The pooled result for age at or over 65 years features in the conclusion along with the result for other pooled results (for NSAIDs, clindamycin and IVIG in clindamycin) but the result for age is only based on 2 studies whilst the results for the other factors are based on 4 to 9 studies. The result based on 2 studies should be given less weight for the reasons I give above.

Response: In an effort to give less weight to the result for age, we have re-ordered the conclusion and now present the result for age following results from meta-analyses based on more studies (page 18, para 2).

"After analyzing 30 different prognostic factor and outcome combinations, we found that clindamycin treatment was significantly associated with an improved STSS prognosis. Further, we found that IVIG treatment may reduce the odds of mortality in STSS patients who receive clindamycin treatment, but we are uncertain if this is true for all STSS patients, regardless of clindamycin treatment. These findings support the use of IVIG as an adjunctive treatment in clindamycin-treated STSS patients. Age equal to or older than 65 years and treatment with NSAIDs was significantly associated with a worse STSS prognosis. Results from very low to low certainty evidence failed to show a significant association between any other factors of interest and STSS prognosis."

27. The authors report: "These findings support the use of IVIG as an adjunctive treatment in clindamycin-treated STSS patients." All the studies have high risk of bias and the certainty of evidence is low and therefore I do not see the justification for such an upbeat conclusion.

Response: We have clarified that although the findings support the use of IVIG, the certainty of evidence was low due to serious risk of bias and imprecision as follows (page 18, para 2).

"Further, we found that IVIG treatment may reduce the odds of mortality in STSS patients who receive clindamycin treatment, but we are uncertain if this is true for all STSS patients, regardless of clindamycin treatment. Although these findings support the use of IVIG as an adjunctive treatment in clindamycin-treated STSS patients, the certainty of evidence was low due to serious risk of bias and

imprecision. Age equal to or older than 65 years and treatment with NSAIDs was significantly associated with a worse STSS prognosis."

Thank you again for your consideration of this manuscript. We hope that we have satisfactorily addressed the reviewer comments and look forward to your response.

Sincerely,
Jessica Bartoszko
On behalf of the authors

VERSION 2 – REVIEW

REVIEWER	Chis Ster, Irina St. George's University of London
REVIEW RETURNED	31-Jul-2022

GENERAL COMMENTS	The authors resonably addressed the issues that I have raised in the first hand review. My minor suggestion would be that of consulting a good/close to native English speaker. I am not one of them either but "certainty is low" for example does not read well to me. Perhaps better "the results exhibit high levels of uncertaincy" sounds better? After all - this is part of the findings for such paper and clearly signals the need for (more) research in the area.

REVIEWER	Taylor, Kathryn University of Oxford, Nuffield Dept of Primary Care Health Sciences
REVIEW RETURNED	12-Jul-2022

GENERAL COMMENTS	The authors have improved the manuscript and I am happy with most of the changes that they have made in response to my comments. The PRISMA diagram has improved in that the numbers now add up but the first box should just refer to the total number of records initially (currently 33123). As I said before, this is not an update of a published study and the PRISMA template does not suggest that numbers from each search and updates are given. The current PRISMA diagram also raises the question of why CINAHL was not searched initially and whether it was correct that no extra records arose from MEDLINE and EMBASE between 6 Aug and 16 September. I also note that the search is well over 6 months old and should be updated. I appreciate the scale of the task in updating a systematic review of prognostic factors but it is worthwhile presenting an up-to-date review. The authors were correct about the terms to interpret the I-squared measure. Personally, I don't find the ranking of "considerable" and "substantial" heterogeneity helpful. The Cochrane Handbook do state they are part of a rough guide and should be considered with other factors including the direction of effects and uncertainty when number of studies is small. The authors reference the Cochrane Handbook and I think it would be helpful to state that when the numbers of studies are small (as in this review) the uncertainty in the value of I-squared is "substantial". There are only few studies in many of the meta-analyses of this review. I am also concerned that the authors have now added sensitivity analysis using Peto's method but they do not report the results

	(they only provide a comment about similarity of results) and they also do not justify the use of Peto's method in terms of the bias. I appreciate that the authors ref to bias in the GRADE assessments but I would dispute the statement "Because results were consistent across Peto, and DerSimonian and Laird methods, our post-hoc sensitivity analysis showed that our meta-analyses based on few events were robust" – this is describing possibly two biased methods with one being less biased than the other. Sensitivity analysis normally tests the robustness of conclusions but in the context of a review of few studies, low events, wide confidence intervals and low/very low certainty of evidence (as reflected in GRADE), I do not think the term "robust" can be used. I would suggest saying something like supporting the main results.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Irina Chis Ster, St. George's University of London

Comments to the Author:

2. *The authors resonably addressed the issues that I have raised in the first hand review.*

Response: Thank you for your thoughtful review of our manuscript.

3. *My minor suggestion would be that of consulting a good/close to native English speaker. I am not one of them either but "certainty is low" for example does not read well to me. Perhaps better "the results exhibit high levels of uncertaincy" sounds better? After all - this is part of the findings for such paper and clearly signals the need for (more) research in the area.*

Response: To Dr. Ster's comment about consulting a native English speaker, we confirm that the manuscript was written by English native speakers. Further, the terminology we used for reporting on the four possible GRADE certainty ratings (very low, low, moderate and high) is consistent with published GRADE guidance. Since our terminology is consistent with published GRADE guidance, and the limitations of the STSS evidence base are detailed in table 2 and the discussion section, we have opted to not update the wording around the GRADE ratings. However, we have clarified where the terminology comes from in the methods section.

Methods (page 7, paragraph 3): "Pairs of reviewers used the grading of recommendations, assessment, development, and evaluation (GRADE) approach to independently assess the certainty of prognostic evidence for each meta-analysed outcome. Criteria for rating the certainty for each prognostic factor and outcome as high, moderate, low, or very low, included considerations of risk of bias, inconsistency, indirectness, size and precision of the association and publication bias [26, 27].

Judgments of imprecision for this systematic review were made using a minimally contextualised approach. This approach considers whether confidence intervals include the null effect. Further, the terminology used to report GRADE ratings (e.g. low certainty evidence) is based on published GRADE guidance [28, 29]…"

Reviewer: 2

Dr. Kathryn Taylor, University of Oxford

Comments to the Author:

4. *The authors have improved the manuscript and I am happy with most of the changes that they have made in response to my comments.*

Response: Thank you for your thoughtful review of our manuscript.

5. *The PRISMA diagram has improved in that the numbers now add up but the first box should just refer to the total number of records initially (currently 33123). As I said before, this is not an update of a published study and the PRISMA template does not suggest that numbers from each search and updates are given.*

Response: We have updated figure 1 (the PRISMA diagram) as per Dr. Taylor's recommendations. It now reports the total number of records in the first box. Because each database yields different numbers of records, we feel it is important for readers to know the number of records coming from each database. Thus, in addition to the total number of records across databases, we have opted to keep the number of records from each database in the first box.

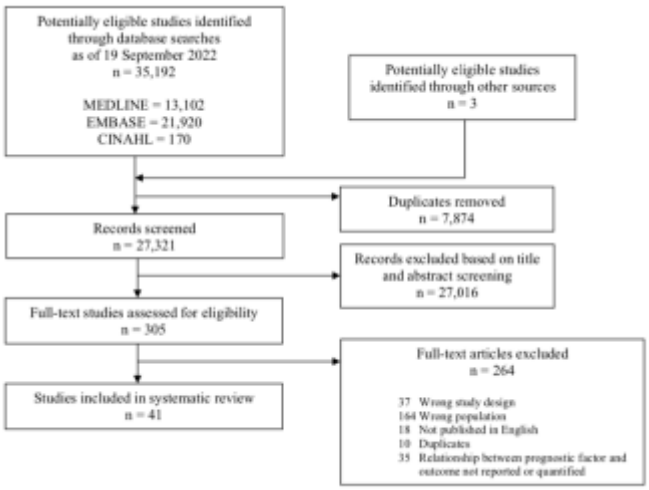
6. *The current PRISMA diagram also raises the question of why CINAHL was not searched initially and whether it was correct that no extra records arose from MEDLINE and EMBASE between 6 Aug and 16 September.*

Response: To clarify, we searched MEDLINE and EMBASE from the inception of each database up to 6 August 2021. In the previous search, records indexed in MEDLINE and EMBASE after 6 August

2021 were not included. Further, because of the limited availability of the methodologist that conducted the searches, the search in CINAHL was conducted at a later time point – after MEDLINE and EMBASE – resulting in a later search cut-off date (i.e. 16 September 2021). Because the difference in search cut-off dates between the databases was approximately one month, we did not anticipate the body of available evidence for STSS prognosis from MEDLINE and EMBASE to have changed substantially. In the updated search, we searched all databases up to 19 September 2022.

7. *I also note that the search is well over 6 months old and should be updated. I appreciate the scale of the task in updating a systematic review of prognostic factors but it is worthwhile presenting an up-to-date review.*

Response: We agree with Dr. Taylor that presenting an up-to-date review is worthwhile. We have updated the search and present an updated figure 1 below. In searching records up to 19 September 2022, we identified 1 new eligible study. We have updated the manuscript, tables, figures and supplement accordingly.



8. *The authors were correct about the terms to interpret the I-squared measure. Personally, I don't find the ranking of "considerable" and "substantial" heterogeneity helpful. The Cochrane Handbook do state they are part of a rough guide and should be considered with other factors including the direction of effects and uncertainty when number of studies is small. The authors reference the Cochrane Handbook and I think it would be helpful to state that when the numbers of studies are small (as in this review) the uncertainty in the value of I-squared is "substantial". There are only few studies in many of the meta-analyses of this review.*

Response: We would like to clarify that the I^2 statistic value was used to quantify heterogeneity across studies within a meta-analysis. "Considerable" and "substantial" are qualifiers for ranges of I^2 statistic values – they do not reflect the uncertainty in the value of the I^2 statistic or the uncertainty in the body of evidence. Given this, and the fact that the majority of meta-analyses in this review are based on

few studies as Dr. Taylor highlights, we maintain that we should consider heterogeneity as "considerable" instead of "substantial" (i.e. rate heterogeneity higher) when the I^2 statistic falls within the range of overlapping values. We have incorporated the additional criterion related to few studies that Dr. Taylor raised as follows:

Methods (page 8, paragraph 2): "We interpreted an I2 statistic value of 0%-40%, 30%-60%, 50%-90%, or 75%-100% as not likely important, moderate, substantial, or considerable heterogeneity, respectively [31]. If an I2 statistic value was within a range of overlapping values (e.g. 80%), we would interpret heterogeneity as more important (e.g. considerable instead of substantial) if the meta-analysis contained few studies, we observed inconsistent magnitudes and directions of summary estimates upon visual inspection of the forest plots, or the chi-square test was significant [31]."

9. *I am also concerned that the authors have now added sensitivity analysis using Peto's method but they do not report the results (they only provide a comment about similarity of results) and they also do not justify the use of Peto's method in terms of the bias. I appreciate that the authors ref to bias in the GRADE assessments but I would dispute the statement "Because results were consistent across Peto, and DerSimonian and Laird methods, our post-hoc sensitivity analysis showed that our meta-analyses based on few events were robust" – this is describing possibly two biased methods with one being less biased than the other. Sensitivity analysis normally tests the robustness of conclusions but in the context of a review of few studies, low events, wide confidence intervals and low/very low certainty of evidence (as reflected in GRADE), I do not think the term "robust" can be used. I would suggest saying something like supporting the main results.*

Response: We agree and have updated the wording of the post-hoc sensitivity analysis results as per Dr. Taylor's suggestion.

Results (page 16, paragraph 1): "Because results were consistent across Peto, and DerSimonian and Laird methods, our post-hoc sensitivity analysis applying the Peto method supported our main results."

Thank you again for your consideration of this manuscript. We hope that we have satisfactorily addressed the reviewer comments and look forward to your response.

Sincerely,

Jessica Bartoszko

On behalf of the authors

VERSION 3 – REVIEW

REVIEWER	Taylor, Kathryn University of Oxford, Nuffield Dept of Primary Care Health Sciences
REVIEW RETURNED	24-Oct-2022

GENERAL COMMENTS	I am happy with the changes made.