PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Risk of Recurrent Spontaneous Preterm Birth: A Systematic Review and Meta-Analysis
AUTHORS	Phillips, Courtney; Velji, Zain; Hanly, Ciara; Metcalfe, Amy

VERSION 1 - REVIEW

REVIEWER	Shayna Conner, M.D., M.S.C.I. Washington University School of Medicine. USA.
REVIEW RETURNED	06-Dec-2016

GENERAL COMMENTS	This is a systematic review and meta-analysis investigating the risk of recurrent spontaneous preterm birth after a prior spontaneous preterm birth. The manuscript is well written and the effort is shown. However there are many critical flaws in the construction and execution of this study which make it difficult to interpret and impossible to apply clinically. Additionally, many standard inclusions for a high-quality meta-analysis were either not performed, or not included. I have the following questions and comments. 1. Please justify the use of only 3 databases. 2. Spelling: PPROM in abstract line 37, topIC page 4 line 27. Please correct. 3. Reference for Page 5 line 28 “The clinical pathways that lead to sPTB include both preterm labor and preterm premature rupture of membranes” This seems like an opinion and not based on established data. If no proper reference for this assumption, please remove. Clinically they are on a spectrum and often occur in same clinical setting. 4. Please specify why the date restrictions on the searches page 6 line 13-15 5. Are all preterm births in direct succession in all studies, please clarify? 6. Was study protocol registered? If not, why not? 7. Why only include studies with N>=20? Considered case series? 8. Why only 24 studies in primary outcome? Looking at appendix B, all but 2 had outcome of PTB<37 weeks. This is very important as transparency is vital. 9. Separate out the risk estimates found from RCTs vs. cohort. 10. Please report in more detail in results section on study quality and why these specific measures were chosen in the methods section, especially as ¼ of the criteria were N/A. What was considered high vs. low quality, etc. 11. Small sample size for secondary analysis and therefore difficult to make any interpretation, yet seems to be the focus of the paper. 12. Was there any additional analysis investigating the heterogeneity, why not? 13. In this paper devoted to recurrence risk of sPTB, there is no

	mention whatsoever of the likely heterogenous use of therapies targeted to reduce the risk of sPTB. Some women in the study likely received 17OHP, cerclage, pessary, vaginal progesterone, etc, whereas some did not. Especially noted in the setting of RCTs. This fact likely significantly influences the results of the primary outcome. Further analysis investigating the recurrence risk for sPTB after these therapies and without these therapies are likely of more clinical benefit. 14. The funnel plot for publication bias should be included, at least in an appendix. Were any statistical analyses used to investigate publication bias? 15. How were studies chosen in the cases of duplicated cohorts (n=17)?

REVIEWER	Gabriele Saccone, MD University of Naples Federico II
REVIEW RETURNED	10-Jan-2017

GENERAL COMMENTS	Well written SR with MA 1) Please add more details in Table 1, see PICOS 2) Please split Table 1 in two Table (i.e. SPTB and IPTB) 3) SPTB and IPTB are two different pathway and should be analyzed separately

REVIEWER	Oudijk, Martijn A AMC, Amsterdam, The Netherlands
REVIEW RETURNED	01-Feb-2017

GENERAL COMMENTS	The authors are to be complimented for their extensive work. Recurrence risk of sPTB is an important topic, and clinically useful data. These data can be used in clinical practice, to inform patients, and to decide upon preventive measures. The article would greatly improve if the authors are able to provide data on subgroups. In the current article only 1 figure is given, that is, the recurrence risk of sPTB < 37 weeks, whereas the risk is also affected by the gestational age at the time of sPTB in the index pregnancy. The earlier the sPTB, the greater the recurrence risk. Could you please provide these data in a new version of the paper? Adding the subgroups will add so much more than just data on < 37 weeks. And will also provide clinicians with useful data on whether or not to administer preventive measures. In addition these data may be used to calculate sample sizes of RCT's on the subject.

VERSION 1 – AUTHOR RESPONSE

REVIEWERS' COMMENTS TO AUTHOR:
Reviewer: 1
Reviewer Name: Shayna Conner, M.D., M.S.C.I.
Institution and Country: Washington University School of Medicine. USA.
Please state any competing interests or state ‘None declared’: None declared.

Please leave your comments for the authors below
This is a systematic review and meta-analysis investigating the risk of recurrent spontaneous preterm

birth after a prior spontaneous preterm birth. The manuscript is well written and the effort is shown. However there are many critical flaws in the construction and execution of this study which make it difficult to interpret and impossible to apply clinically. Additionally, many standard inclusions for a high-quality meta-analysis were either not performed, or not included. I have the following questions and comments.

1. Please justify the use of only 3 databases.

The decision to only search 3 databases was related to both the anticipated yield and resources available for this project. An analysis of the effectiveness of different databases for systematic reviews related to maternal morbidity/mortality concluded that Medline and Embase had the highest yield. We have expanded on that in this study by also searching clinicialtrials.gov and hand searching the bibliographies of included studies. We do acknowledge articles in journals that are not indexed by Medline or Embase may have been missed and have added a statement to that effect to the limitations.

2. Spelling: PPROM in abstract line 37, topIC page 4 line 27. Please correct.

Thank you for identifying this typo. We have corrected this.

3. Reference for Page 5 line 28 “The clinical pathways that lead to sPTB include both preterm labor and preterm premature rupture of membranes” This seems like an opinion and not based on established data. If no proper reference for this assumption, please remove. Clinically they are on a spectrum and often occur in same clinical setting.

We have re-phrased this sentence

4. Please specify why the date restrictions on the searches page 6 line 13-15

We searched the entirety of these databases. The dates provided indicate the first year that studies were indexed in a particular database.

5. Are all preterm births in direct succession in all studies, please clarify?

Yes, we only included successive preterm births. This is stated at the end of the second paragraph on page 7.

6. Was study protocol registered? If not, why not?

Unfortunately, the study protocol was not registered. This was an oversight on our part. We did not feel it was appropriately to retroactively register the protocol once we had already begun work on this project.

7. Why only include studies with N>=20? Considered case series?

We only included studies with a N≥20 to exclude case studies. This may have resulted in some small case series also being excluded. We have added a statement to the discussion to acknowledge this as a limitation.

8. Why only 24 studies in primary outcome? Looking at appendix B, all but 2 had outcome of PTB<37 weeks. This is very important as transparency is vital.

Due to heterogeneity in the definition of preterm birth in the index pregnancy, the main analysis was restricted to studies that defined preterm birth as occurring prior to 37 weeks of gestation in the index and the subsequent pregnancy. Too few studies used alternative cut-points to define preterm birth (i.e. <34 weeks, <32 weeks) to create pooled estimates. We have clarified this in the results section.

9. Separate out the risk estimates found from RCTs vs. cohort.

We have created a stratified analysis based on study design and updated the methods, results, and discussion sections accordingly.

10. Please report in more detail in results section on study quality and why these specific measures were chosen in the methods section, especially as ¼ of the criteria were N/A. What was considered high vs. low quality, etc.

As we were examining the recurrence risk of spontaneous preterm birth, by definition, there was no 'unexposed' cohort in observational studies as everyone had to have an index preterm birth. This meant that 2 of the 8 items in the Newcastle Ottawa Scale (specifically selection of the non-exposed cohort and comparability of the exposed and non-exposed cohorts) were not applicable. We have further expanded upon this in the result section and conducted a stratified analysis based on quality score.

11. Small sample size for secondary analysis and therefore difficult to make any interpretation, yet seems to be the focus of the paper.

We acknowledge the small sample size for the secondary analysis regarding subtypes of preterm birth. While we agree that these results should be interpreted with caution, to our knowledge this is the first time this phenomenon has been examined in a meta-analysis and makes a novel contribution to the literature. As such, we have devoted a large proportion of our discussion to this. We hope that the findings from this study will motivate others to differentiate between spontaneous preterm birth due to preterm labour vs. PPROM in other original investigations so in the future these effect estimates can be updated with more precision.

12. Was there any additional analysis investigating the heterogeneity, why not?

Heterogeneity was assessed visually by examining the clinical pathway leading to preterm birth (addressed statistically by a stratified analysis) and by the gestational age at the index and recurrent preterm birth (addressed statistically by restriction of the meta-analysis to studies with common definitions). Additionally, we examined the I2 and Cochrane Q statistics to examine statistical heterogeneity in our pooled estimates. Random effect models were used throughout due to the underlying clinical heterogeneity. We felt this was sufficient to quantify and address the heterogeneity of the data used in this study, but would be happy to conduct some additional analysis if you have something specific that you'd like us to assess.

13. In this paper devoted to recurrence risk of sPTB, there is no mention whatsoever of the likely heterogenous use of therapies targeted to reduce the risk of sPTB. Some women in the study likely received 17OHP, cerclage, pessary, vaginal progesterone, etc, whereas some did not. Especially noted in the setting of RCTs. This fact likely significantly influences the results of the primary outcome. Further analysis investigating the recurrence risk for sPTB after these therapies and without these therapies are likely of more clinical benefit.

We agree with this comment, but the underlying literature either did not report on this or was too heterogeneous to examine. We have added a statement to this effect in the discussion. We have also

updated the tables in Appendix B to provide more detail on inclusion/exclusion criteria of the original studies which includes some data on the use of therapeutics.

14. The funnel plot for publication bias should be included, at least in an appendix. Were any statistical analyses used to investigate publication bias?

Unfortunately a funnel plot was not created for this study as to our knowledge the funnel plot commands in Stata have not been developed to work with the metaprop commands that are used in the meta-analysis of proportions.

15. How were studies chosen in the cases of duplicated cohorts (n=17)?

In the case of duplicate data, the study with the largest sample size was chosen. We have updated the methods section to reflect this.

Reviewer: 2
Reviewer Name: Gabriele Saccone, MD
Institution and Country: University of Naples Federico II
Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below
Well written SR with MA
1) Please add more details in Table 1, see PICOS

We have updated Table B1 and B2 to provide more information on the included studies.

2) Please split Table 1 in two Table (i.e. SPTB and IPTB)

We apologize for the confusion – Table B1 only reports the recurrence rate of spontaneous preterm birth. We have added table B2, which reports the occurrence rate of iatrogenic preterm birth following a spontaneous preterm birth in the prior pregnancy.

3) SPTB and IPTB are two different pathway and should be analyzed separately

We agree and confirm that sPTB and iPTB have been analyzed separately.

Reviewer: 3
Reviewer Name: Oudijk
Institution and Country: AMC, Amsterdam, The Netherlands
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The authors are to be complimented for their extensive work. Recurrence risk of sPTB is an important topic, and clinically useful data. These data can be used in clinical practice, to inform patients, and to decide upon preventive measures.
The article would greatly improve if the authors are able to provide data on subgroups. In the current article only 1 figure is given, that is, the recurrence risk of sPTB < 37 weeks, whereas the risk is also affected by the gestational age at the time of sPTB in the index pregnancy. The earlier the sPTB, the greater the recurrence risk. Could you please provide these data in a new version of the paper? Adding the subgroups will add so much more than just data on < 37 weeks. And will also provide clinicians with useful data on whether or not to administer preventive measures. In addition these data

may be used to calculate sample sizes of RCT's on the subject.

We agree with this point, unfortunately there were not enough studies that have reported on sPTB prior to 37 weeks to create meaningful pooled estimates. We have provided the raw data for this in Table B1.

VERSION 2 – REVIEW

REVIEWER	Martijn Oudijk AMC, Amsterdam The Netherlands
REVIEW RETURNED	08-Apr-2017

GENERAL COMMENTS	I'm satisfied with the reply of the authors on all subjects, and the addition of the appendix.

REVIEWER	Javier Zamora Hospital Ramon y Cajal. Spain
REVIEW RETURNED	14-May-2017

GENERAL COMMENTS	The manuscript reports on an interesting question that authors addressed after a thorough and comprehensive systematic review. Although the topic has been previously studied, authors focus on estimating the risk of recurrent sPTB and iPTB following sPTB in singleton pregnancies but stratified by pathophysiological pathway and stratified the analysis for preterm labour (PTL) or preterm premature rupture of membranes (PPROM) subgroups. Unfortunately, literature does not provide too many studies to inform this subgroup analysis which limits the conclusions of the manuscript in this point. Minor revisions Methods Search was updated on July 2016. At the time the paper would be published, the search is going to be slightly (near one year) obsolete. It is not clear if title and abstracts were scrutinised by only one reviewer. It seems that two authors performed the searches but it is not clear if the same reviewers made the screening of records. The study aims to estimate the risk of recurrent sPTB following sPTB in singleton pregnancies. To achieve this aim, authors collected evidence from both observational studies (cohorts) and RCT. It is fine to use a risk of bias tool for observational studies like Newcastle-Ottawa-score but the use of Jadad scale to assess the quality of randomised controlled trials (RCT) is not adequate. The scale is rather obsolete and does not focus on the critical points that can bias the results. Focusing on randomization methods and blinding is not critical at this time given the observational nature of the study aim. Moreover, it is not clear whether the authors included both arms of RCT included in the review or only the control arms that used standard care. I guess that many of the trials included would have been designed to assess if different interventions have an impact on

	the risk of SPTB as an outcome. In such a case, patients randomised to intervention arms could not be valid to assess the risk of sPTB. This surely deserves discussion. Statistical methods section does not describe in detail the model authors have used for the analysis. It is merely described as a random effects model without further details about neither the weights used or standard errors calculations or any transformation of data before the analysis. Some more details will be appreciated in order to permit reproducibility. Results The number of records retrieved independently from the two main searching platforms (EMBASE and MedLine) was huge and it is not clear if the number included in the top boxes in figure 1 represent unique records or contain duplicates. I suppose that after deduplicating records these numbers could be much lower. As discussed above, the analysis stratified by study quality in RCT is not adequate. I suggest using the same instrument to assess the risk of bias for both cohort and RCT (standard care control arm) restricting to the domains regarding population representativeness, outcome assessment and follow-up.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3
Reviewer Name: Martijn Oudijk
Institution and Country: AMC, Amsterdam, The Netherlands Please state any competing interests or state ‘None declared’: None declared.

I'm satisfied with the reply of the authors on all subjects, and the addition of the appendix.

Thank you

Reviewer: 4
Reviewer Name: Javier Zamora
Institution and Country: Hospital Ramon y Cajal. Spain Please state any competing interests or state ‘None declared’: None declared

The manuscript reports on an interesting question that authors addressed after a thorough and comprehensive systematic review. Although the topic has been previously studied, authors focus on estimating the risk of recurrent sPTB and iPTB following sPTB in singleton pregnancies but stratified by pathophysiological pathway and stratified the analysis for preterm labour (PTL) or preterm premature rupture of membranes (PPROM) subgroups. Unfortunately, literature does not provide too many studies to inform this subgroup analysis which limits the conclusions of the manuscript in this point.

Thank you for your comments. We agree with this limitation in the existing literature and hope that this review will motivate authors of primary studies to consider reporting on the pathophysiological pathway leading to preterm birth.

Minor revisions

Methods

Search was updated on July 2016. At the time the paper would be published, the search is going to be slightly (near one year) obsolete.

Thank you for your comment. We have updated the search. This resulted in the addition of one small study, but the overall pooled estimate of the recurrence rate of spontaneous preterm birth remained identical.

It is not clear if title and abstracts were scrutinised by only one reviewer. It seems that two authors performed the searches but it is not clear if the same reviewers made the screening of records.

This was completed by 2 reviewers. We have updated the methods section to make this clearer.

The study aims to estimate the risk of recurrent sPTB following sPTB in singleton pregnancies. To achieve this aim, authors collected evidence from both observational studies (cohorts) and RCT. It is fine to use a risk of bias tool for observational studies like Newcastle-Ottawa-score but the use of Jadad scale to assess the quality of randomised controlled trials (RCT) is not adequate. The scale is rather obsolete and does not focus on the critical points that can bias the results. Focusing on randomization methods and blinding is not critical at this time given the observational nature of the study aim. Moreover, it is not clear whether the authors included both arms of RCT included in the review or only the control arms that used standard care. I guess that many of the trials included would have been designed to assess if different interventions have an impact on the risk of SPTB as an outcome. In such a case, patients randomised to intervention arms could not be valid to assess the risk of sPTB. This surely deserves discussion.

In the context of randomized controlled trials, we did include both the intervention and control arms in the point estimates. This is also reflective of the inclusion criteria for observational studies. Ultimately many women included in both the observational and randomized controlled trials would have received some form of treatment to prevent the recurrence of spontaneous preterm birth. Unfortunately, to date, none of these proposed treatments have been shown to be particularly effective in a real-world setting, thus we think it is appropriate to include all women from all eligible studies in this review. We have added a section to the discussion to reflect this. As you suggested below, we have also re-done the quality assessment for the RCTs using the Newcastle Ottawa Scale.

Statistical methods section does not describe in detail the model authors have used for the analysis. It is merely described as a random effects model without further details about neither the weights used or standard errors calculations or any transformation of data before the analysis. Some more details will be appreciated in order to permit reproducibility.

We have added more information on the statistical methods used in the method section and provided a reference to a manuscript that provides a more detailed account should readers be interested in a greater level of detail.

Results
The number of records retrieved independently from the two main searching platforms (EMBASE and MedLine) was huge and it is not clear if the number included in the top boxes in figure 1 represent unique records or contain duplicates. I suppose that after deduplicating records these numbers could be much lower.

These large estimates do include duplicates.

As discussed above, the analysis stratified by study quality in RCT is not adequate. I suggest using the same instrument to assess the risk of bias for both cohort and RCT (standard care control arm)

restricting to the domains regarding population representativeness, outcome assessment and follow-up.

We have re-done the quality assessment for the RCTs using the Newcastle Ottawa Scale as suggested.

VERSION 3 – REVIEW

REVIEWER	Javier Zamora Hospital Ramón y Cajal, Spain
REVIEW RETURNED	04-Jun-2017

GENERAL COMMENTS	I'm satisfied with author's responses to the concerns raised in my previous review although a couple of minor points still needs to be addressed. 1. I asked for clarification on whether screening of TITLES AND ABSTRACTS were made by one or two reviewers. Authors have insisted in that FULL-TEXT reviewed articles were assessed in duplicate (in fact they wrote an additional word in methods: i.e. "jointly"). This was already clear in the first version of the manuscript. Maybe my question was not framed clear enough, and thus it is still is unaddressed. Please clarify whether TITLES AND ABSTRACTS were scrutinised by only one reviewer or alternatively two independent reviewers selected titles and abstract for further assessment as full-text articles. According to MECIR R40 criteria, it is mandatory to report how many people were involved from search results to included studies. This involves details on search (ZV y CH), selection of titles and abstracts (unknown) and full-text assessment (ZV and CH). 2. I suggested removing the reference to Jadad score. It is rather obsolete and has nothing to do with assessing the risk of bias of the RCTs included in this review. If a trial is blind, it is fine for a randomised comparison of two arms to assess the efficacy of an intervention. The same occurs with randomisation. If a trial is nicely randomised (and more important if the allocation is concealed), we know this is critical for a fair comparison to assess the effect of an intervention. However, for the aim of this review, both domains (i.e. blinding and randomization) are unimportant. Biases come from different sources. The review is assessing the risk of SB in both arms of a trial (the whole cohort of participants). Thus, the risk domains regard to the representativeness of the cohort (both arms), the ascertainment of the exposure (i.e. previous SB), demonstration that outcome of interest was not present at the start of the study, assessment of outcome, length and adequacy of follow-up. So, six domains are necessary to evaluate the risk of bias, independently of whether the study was observational or experimental. Finally, authors wrongly copied and pasted the rows from table C2 to table C3 and so, scores in C3 are wrong. To summarise, my suggestion is to remove any reference to Jadad score and to use the same tool (adapted NOS) for all studies as all of them are providing observational data to the review (prevalence of SB).

VERSION 3 – AUTHOR RESPONSE

Reviewer: 4
Reviewer Name: Javier Zamora
Institution and Country: Hospital Ramón y Cajal, Spain Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below I'm satisfied with author's responses to the concerns raised in my previous review although a couple of minor points still needs to be addressed.

Thank you. We have addressed your other points below.

1. I asked for clarification on whether screening of TITLES AND ABSTRACTS were made by one or two reviewers. Authors have insisted in that FULL-TEXT reviewed articles were assessed in duplicate (in fact they wrote an additional word in methods: i.e. "jointly"). This was already clear in the first version of the manuscript. Maybe my question was not framed clear enough, and thus it is still is unaddressed. Please clarify whether TITLES AND ABSTRACTS were scrutinised by only one reviewer or alternatively two independent reviewers selected titles and abstract for further assessment as full-text articles. According to MECIR R40 criteria, it is mandatory to report how many people were involved from search results to included studies. This involves details on search (ZV y CH), selection of titles and abstracts (unknown) and full-text assessment (ZV and CH).

Sorry about the confusion. The screening of titles and abstracts was also done by two reviewers. Articles identified by either reviewer at this stage as potentially relevant was considered for full text review. We have updated the methods section accordingly.

2. I suggested removing the reference to Jadad score. It is rather obsolete and has nothing to do with assessing the risk of bias of the RCTs included in this review. If a trial is blind, it is fine for a randomised comparison of two arms to assess the efficacy of an intervention. The same occurs with randomisation. If a trial is nicely randomised (and more important if the allocation is concealed), we know this is critical for a fair comparison to assess the effect of an intervention. However, for the aim of this review, both domains (i.e. blinding and randomization) are unimportant. Biases come from different sources. The review is assessing the risk of SB in both arms of a trial (the whole cohort of participants). Thus, the risk domains regard to the representativeness of the cohort (both arms), the ascertainment of the exposure (i.e. previous SB), demonstration that outcome of interest was not present at the start of the study, assessment of outcome, length and adequacy of follow-up. So, six domains are necessary to evaluate the risk of bias, independently of whether the study was observational or experimental.
Finally, authors wrongly copied and pasted the rows from table C2 to table C3 and so, scores in C3 are wrong.

To summarise, my suggestion is to remove any reference to Jadad score and to use the same tool (adapted NOS) for all studies as all of them are providing observational data to the review (prevalence of SB).

Thank you for the clarification. We have removed all references to the Jadad score, combined all of the quality assessments into a single table and fixed the errors in what was previously labeled Table C3