BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers’ comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Use of C-reactive protein to tailor antibiotic use: a systematic review and meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Petel, Dara; Winters, Nicholas; Gore, Genevieve C.; Papenburg, Jesse; Beltempo, Marc; Lacroix, Jacques; Fontela, Patricia</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Heiman Wertheim Radboudumc, Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>REVIEW RETURNED</td>
<td>23-Mar-2018</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>The authors performed a systematic review &amp; meta-analysis on the performance of CRP testing on the initiation and duration of antibiotic therapy. The study was performed in a robust manner and the results can aid in shaping strategies to reduce antibiotic use worldwide. There are some minor comments: -the study by Do NT (ref 30) does include children. Why are these not presented? -the study populations are quite diverse: NICU versus primary care and high income setting versus low middle income. Good to reflect on this in the discussion. -I recommend to discuss the use CRP versus other biomarkers like PCT in the discussion section.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Chien-Chang Lee Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td>REVIEW RETURNED</td>
<td>02-Apr-2018</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>Reviewer’s comment: The authors conducted a systemic review and meta-analysis of studies comparing the effects of c-reactive protein use and routine care on various outcomes, including the duration of antibiotic use, antibiotic initiation, mortality, infection relapse, and hospitalization. I have some concerns and questions regarding this study. Major concerns: 1. The research question(s) were not clear. There have been</td>
</tr>
</tbody>
</table>
several meta-analyses of CRP on various outcomes (diagnostic accuracy, guide of antibiotic use, mortality, etc) of different infectious diseases (pneumonia, sepsis, etc) in adults, infants, and other population. The authors focused on two major outcomes (duration of antibiotic use and antibiotic initiation). The values of CRP may vary in different infectious diseases. The authors did not consider the disease entities during analyses. This reduced the clinical values of the results.

2. The novelty of this study was not impressive. According to the two major outcomes (duration of antibiotic use and antibiotic initiation) in this study, there have been meta-analyses of the antibiotics initiation in adult and neonate population separately. Indeed, there was no meta-analysis in the antibiotic duration. However, there were merely two RCTs and two cohort studies in the results. Overall, the results were lack of novelty.

3. It seems that the authors only included studies about “CRP-guided antibiotic use” and excluded studies regarding “CRP to improve the diagnosis of bacterial infection”. Typically, the use of CRP in the diagnosis of bacterial infection implied the further initiation of antibiotics. The authors should include the studies of diagnostic accuracy in the analyses of initiation of antibiotics.

4. Why the authors excluded case-control study and cross-sectional study in their analyses? Many meta-analyses of CRP related studies included case-control studies and cross-sectional studies because the RCTs were limited in this topic.

Minor concerns:
1. In the abstract, line10….antibiotic duration by -1.45 days (95%CI -1.55, -0.71) in two RCTs. Is it correct? How could it be possible that a point estimate of -1.45 had a 95%CI ranging -1.55 to -0.71?

2. The authors should check their manuscript carefully. Some typos or mis-spellings were found. For example, in the abstract line 3….for randomized “control” trials….it should be “controlled”.

3. The statement that CRP-guided treatment could reduce antibiotic duration in adults came from single cohort study (Gao et al.) This cannot be used as the result of a meta-analysis.

4. In the supplementary table 1, the study by Jaswal et al. had no information on antibiotics duration. Why the authors included this study? How could the authors analyze the result in Figure 6 including this study by Jaswal et al.?

5. In the CRP-guided decision, the cut-off level of CRP would determine the decision. Different cut-off levels of CRP may result in different outcomes. Is it reasonable to conduct a meta-analysis including studies using different cut-off levels of CRP? Please explain why or why not.

REVIEWER
Sitaram Vangala
Principal Statistician, Department of Medicine Statistics Core, University of California Los Angeles, United States of America

REVIEW RETURNED 18-May-2018

GENERAL COMMENTS
In this manuscript, the authors describe a systematic review and meta-analysis of studies investigating the use of CRP-based algorithms to guide decisions regarding appropriate initiation and duration of antibiotic courses. The methodology and results are clearly described, and sufficient information is provided to facilitate replication. The review and analysis were undertaken with great care and attention to possible sources of bias.
Major points:

1. The use of a CRP-based algorithm would necessarily have some impact on the endpoints of initiation and duration of antibiotic treatment. It is useful to estimate the extent to which these differ between patients for whom such an algorithm was used and patients for whom it was not, but of what interest is the null hypothesis of no difference? Most of the relevant language in the manuscript concerns estimation, but multiple references to statistical significance are made in the results section on p. 10 (lines 14-18 and 41-45), which seem to be based on whether the 95% confidence interval excludes 0. I would recommend that analysis of these outcomes be consistently formulated in terms of estimation given the implausibility of no impact.

2. The authors repeatedly conflate non-significance with no differences (e.g., in the results component of the abstract p.3 lines 22-28, the results section p.11 lines 1-5, and the discussion p.14 lines 32-37). If the hypothesis of interest is that an outcome does not differ between groups, or is no worse in the CRP-based algorithm group, an equivalence or non-inferiority test should be used, rather than a difference test. The problem with the approach taken is that insufficient power may well be the explanation for any non-significance. Since confidence intervals are consistently reported, the appropriate test can be performed simply by specifying some equivalence/non-inferiority margins in the methods section against which these intervals should be compared. It would also be useful to investigate what sorts of margins these meta-analyses are powered to detect.

3. These inappropriate tests threaten the validity of some of the safety conclusions drawn by the authors. The abstract and results sections report no differences in hospitalization, mortality or relapse rates. This is the basis on which the authors conclude that CRP-based algorithms are safe. However, in the single RCT conducted in adults, the confidence interval for the mortality risk difference is (-14%, 17%), which would surely breach any reasonable equivalence/non-inferiority margin. If the authors would state upfront how they would define equivalence or non-inferiority for each endpoint, then clinical audiences would be in a better position to evaluate whether their safety conclusions are actually supported by the meta-analysis.

Minor points:

1. Though the authors exercise great care in stratifying different study designs and patient populations, this may result in low power for many of these analyses. It would be illuminating to see what these results look like when pooled in the context of a regression model, where the stratification factors are included as covariates instead of defining entirely separate meta-analyses. This may result in tighter confidence intervals, allowing clearer conclusions to be drawn about safety, alongside the more conservative results already reported in the paper.

2. It would be interesting to see how these treatment effects vary by the CRP cutoff used. It is possible that some cutoffs may be too aggressive, and so safety may be easier to demonstrate in studies looking at less aggressive cutoffs.
3. How sensitive are the results to the inclusion/exclusion of Little et. al. (2013) and Llor et. al. (2012), each of which has a sample size far larger than any other study? These are likely dominating the relevant adult meta-analyses.

VERSITON 1 – AUTHOR RESPONSE

Reviewer 1 (Dr. Heiman Wertheim)

1. **Minor comments:**
   1.1. The study by Do NT (ref 30) does include children. Why are these not presented?

We appreciate the comment made by Dr. Wertheim. In our study protocol registered with PROSPERO (ID#: CRD42016038622), we had defined studies including a pediatric population as the following: “a study population will be considered to be pediatric if >85% of the study subjects are younger than 18 years of age or if the study was carried out in a pediatric hospital.” Consequently, the study of Do et al. was initially excluded because 51% of the study subjects were ≤ 15 years of age. However, we agree with Dr. Wertheim that the total number of subjects ≤ 15 years of age was substantial (1059 patients). Thus, we revised our manuscript to include Do et al. study and we report that the results of a meta-analysis using a random effects model including the RCTs of Do et al. and Rebnord et al. was inconclusive (RD -3% of antibiotic treatment initiation with a 95%CI -14%, 8%):

“…Finally, two RCTs including children indicated no difference between CRP and control groups (RD -3%; 95% CI -14, 8).”

1.2. The study populations are quite diverse: NICU versus primary care and high income setting versus low middle income. Good to reflect on this in the discussion.

We have added the following sentence to the discussion:

“…The above findings were consistent regardless of the varied designs of included studies in this review and also diversity of the study populations, which come from both high and low income countries…”

1.3. I recommend to discuss the use CRP versus other biomarkers like PCT in the discussion section.

We follow Dr. Wertheim’s suggestion and added the following paragraph to our discussion:

“…There is scarce literature comparing the performance of CRP to other biomarkers to guide antibiotic use. The RCT of Oliveira et al. compared the use of CRP and procalcitonin algorithms to determine antibiotic treatment duration in 94 critically ill adult patients with severe sepsis or septic shock. No difference in the median duration of treatment was observed between the procalcitonin (7 days; IQR 6, 8.5) and CRP groups (6 days; IQR 5, 7). Importantly, the study treatment algorithm imposed an upper limit of 7 days of antibiotic treatment for patients who showed signs of clinical resolution of sepsis, independently of CRP and procalcitonin levels, which may have contributed for this lack of difference.”
Reviewer 2 (Dr. Chien-Chang Lee)

1. **Major concerns:**
   1.1. The research question(s) were not clear. There have been several meta-analyses of CRP on various outcomes (diagnostic accuracy, guide of antibiotic use, mortality, etc) of different infectious diseases (pneumonia, sepsis, etc) in adults, infants, and other population. The authors focused on two major outcomes (duration of antibiotic use and antibiotic initiation). The values of CRP may vary in different infectious diseases. The authors did not consider the disease entities during analyses. This reduced the clinical values of the results.

   We appreciate the comment of Dr. Lee, but we disagree with it. We believe that our research questions were very clear. As stated in our manuscript, our main objective was to determine the effect of using a CRP-based algorithm on antibiotic consumption in patients with a suspected bacterial infection. We later specified that antibiotic consumption was measured as “antibiotic treatment initiation” and “antibiotic treatment duration”. Moreover, we also aimed to determine the safety of using a CRP-based strategy to guide antibiotic use. For this aim, we chose “infection relapse”, “mortality” and hospitalization, which are appropriate safety outcomes for the research question. Although other systematic reviews and meta-analyses about the use of CRP to guide antibiotic initiation in adult patients with respiratory tract infections have been published, there have been no such studies evaluating the use of CRP to guide antibiotic duration and none including neonates and children.

   It is true that we have not stratified studies by infection type. However, this was naturally achieved when we stratified the included studies by their patient population, as all neonatal studies included newborns with sepsis and all adult and pediatric studies evaluating antibiotic initiation included patients with respiratory tract infections. Therefore, the clinical validity of our results was maintained.

   1.2. The novelty of this study was not impressive. According to the two major outcomes (duration of antibiotic use and antibiotic initiation) in this study, there have been meta-analyses of the antibiotics initiation in adult and neonate population separately. Indeed, there was no meta-analysis in the antibiotic duration. However, there were merely two RCTs and two cohort studies in the results. Overall, the results were lack of novelty.

   We hear Dr. Lee’s concern about the novelty of our study. We would like to explain that its novelty resides in the fact that, so far, no systematic review/meta-analysis evaluating the use of CRP to guide antibiotic duration and none including neonatal and pediatric patients has been published.

   The fact that we include few studies on the use of CRP to guide antibiotic treatment duration in neonates reflects the quality of the literature being published and its suitability to answer the proposed research question. We do not believe that this can be used as a surrogate for “novelty”. In addition, even if not “novel”, our research question is very pertinent in the neonatal setting, where many neonatal intensive care units have been implementing CRP-based algorithms to guide antibiotic treatment duration/discontinuation. It is therefore imperative to evaluate the available evidence with the methodological rigor that we propose in this study.

   1.3. It seems that the authors only included studies about “CRP-guided antibiotic use” and excluded studies regarding “CRP to improve the diagnosis of bacterial infection”. Typically, the use of CRP in the diagnosis of bacterial infection implied the further initiation of antibiotics. The authors should include the studies of diagnostic accuracy in the analyses of initiation of antibiotics.

   We thank Dr. Lee for his comment. As previously explained, in our study, we would like to evaluate the effect of CRP-based algorithms on antibiotic use in different patient populations/settings. To do so, we needed to include studies that had implemented such algorithms as their intervention. It is true that there are several studies evaluating the performance of CRP to diagnose bacterial infections. However, these are diagnostic studies whose outcome is not “antibiotic use” but the
sensitivity, specificity, positive and negative predictive values of different CRP cutoffs to recognize bacterial infections. Thus, such studies are not suitable to be used to answer our research questions and should not be included in this systematic review/meta-analysis.

1.4. Why the authors excluded case-control study and cross-sectional study in their analyses? Many meta-analyses of CRP related studies included case-control studies and cross-sectional studies because the RCTs were limited in this topic. We have excluded cross-sectional and case-control studies because these designs are not suitable to study an outcome as “antibiotic treatment duration”. Such continuous outcome implies the need for follow up, which is not something that happens in cross-sectional studies. In addition, continuous outcomes are not suitable for case-control studies, which need a categorical outcome to classify groups as cases and controls.

2. Minor concerns:
2.1. In the abstract, line10….antibiotic duration by -1.45 days (95%CI -1.55, -0.71) in two RCTs. Is it correct? How could it be possible that a point estimate of -1.45 had a 95%CI ranging -1.55 to -0.71?
We thank Dr. Lee for noting the typo. The correct result is -1.45 days (95%CI -2.61, -0.28) as reported in the Results section. We have corrected the mistake in the abstract:
“…. In neonates, CRP-based algorithms shortened antibiotic treatment duration by -1.45 days (95%CI -2.61, -0.28) in two RCTs, …”

2.2. The authors should check their manuscript carefully. Some typos or mis-spellings were found. For example, in the abstract line 3….for randomized “control” trials….it should be “controlled”.
Thank you for carefully checking our manuscript. We have judiciously reviewed our manuscript for typos and corrected the aforementioned one.

2.3. The statement that CRP-guided treatment could reduce antibiotic duration in adults came from single cohort study (Gao et al.) This cannot be used as the result of a meta-analysis.
Thank you for pointing our mistake. We have rewritten the sentence in the discussion section:
“….Similarly, CRP-based algorithms also reduced antibiotic initiation in adult outpatients…”

2.4. In the supplementary table 1, the study by Jaswal et al. had no information on antibiotics duration. Why the authors included this study? How could the authors analyze the result in Figure 6 including this study by Jaswal et al.?
Despite the fact that this study did not present data on our primary outcome, antibiotic treatment duration, it presented data on infection relapse, which was one of our secondary outcomes. This is why it was included in our systematic review/meta-analysis.

2.5. In the CRP-guided decision, the cut-off level of CRP would determine the decision. Different cut-off levels of CRP may result in different outcomes. Is it reasonable to conduct a
We agree with Dr. Lee that it is possible that the CRP cut-offs for different infectious diseases, or even patient populations, will differ. Nevertheless, this was not a problem in our study because the CRP cut-offs used in different studies were very similar once we stratified them by age group (which naturally stratified them by infection type too).

Reviewer 3 (Mr. Sitaram Vangala)

s1. Major points:

1.1. The use of a CRP-based algorithm would necessarily have some impact on the endpoints of initiation and duration of antibiotic treatment. It is useful to estimate the extent to which these differ between patients for whom such an algorithm was used and patients for whom it was not, but of what interest is the null hypothesis of no difference? Most of the relevant language in the manuscript concerns estimation, but multiple references to statistical significance are made in the results section on p. 10 (lines 14-18 and 41-45), which seem to be based on whether the 95% confidence interval excludes 0. I would recommend that analysis of these outcomes be consistently formulated in terms of estimation given the implausibility of no impact.

We thank Mr. Vangala for the comment. However, we disagree that the use of a CRP-based algorithm will inevitably decrease antibiotic use by making physicians withhold or stop such drugs. Studies evaluating the use of other biomarkers to guide antibiotic treatment (e.g., procalcitonin) showed that the potential impact of such algorithms is observed when the proposed cut-offs are achieved for study participants. However, one important fact is that, in many cases, physicians opt not to follow the algorithm when this says that antibiotic should be withheld or stopped for patients who they consider clinically ill (Bouadma L et al. Lancet 2010;375(9713):463-474). Therefore, our null hypothesis that CRP-based algorithms would not have an effect on antibiotic treatment initiation or duration is plausible and we planned our statistical analysis accordingly.

1.2. The authors repeatedly conflate non-significance with no differences (e.g., in the results component of the abstract p.3 lines 22-28, the results section p.11 lines 1-5, and the discussion p.14 lines 32-37). If the hypothesis of interest is that an outcome does not differ between groups, or is no worse in the CRP-based algorithm group, an equivalence or non-inferiority test should be used, rather than a difference test. The problem with the approach taken is that insufficient power may well be the explanation for any non-significance. Since confidence intervals are consistently reported, the appropriate test can be performed simply by specifying some equivalence/non-inferiority margins in the methods section against which these intervals should be compared. It would also be useful to investigate what sorts of margins these meta-analyses are powered to detect.

We will answer points 1.2 and 1.3 together as they refer to the same issue (see below).

1.3. These inappropriate tests threaten the validity of some of the safety conclusions drawn by the authors. The abstract and results sections report no differences in hospitalization, mortality or relapse rates. This is the basis on which the authors conclude that CRP-based algorithms are safe. However, in the single RCT conducted in adults, the confidence interval for the mortality risk difference is (-14%, 17%), which would surely breach any reasonable equivalence/non-inferiority margin. If the authors would state upfront how they would define equivalence or non-inferiority for each endpoint, then clinical audiences would be in a better position to evaluate whether their safety conclusions are actually supported by the metaanalysis.

We appreciate the reviewer’s thoughtful point. We were cautious to state that the use of CRP-based algorithms to guide antibiotic use appears to be safe because, when evaluating safety outcomes such as mortality (neonatal and adult outpatient studies), infection relapse (neonatal studies) and hospitalization (adult studies), all analyzed studies individually showed no difference between groups or a result favouring the CRP group (see below).
<table>
<thead>
<tr>
<th>Author / year</th>
<th>Study group</th>
<th>Study Design</th>
<th># of patients</th>
<th>CRP</th>
<th>Control</th>
<th>CRP</th>
<th>Control</th>
<th>CRP</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oliveira et al. 2013</td>
<td>Adults</td>
<td>RCT</td>
<td>45 49</td>
<td>21 (47%)</td>
<td>21 (43%)</td>
<td>1 (2%)</td>
<td>3 (6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gao et al. 2010</td>
<td>Adults</td>
<td>Cohort</td>
<td>18 28</td>
<td>NA</td>
<td>NA</td>
<td>0 (0%)</td>
<td>3 (11%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ehl et al. 1997</td>
<td>Premature</td>
<td>RCT</td>
<td>43 39</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coggins et al. 2013</td>
<td>Premature</td>
<td>Cohort</td>
<td>409 160</td>
<td>6 (1.5%)</td>
<td>14 (9%)</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbenjapon et al. 2015</td>
<td>Neonates</td>
<td>RCT</td>
<td>11 11</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (9%)</td>
<td>1 (9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaswal et al. 2003</td>
<td>Neonates</td>
<td>Cohort</td>
<td>14 14</td>
<td>NA</td>
<td>NA</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couto et al. 2007</td>
<td>Neonates</td>
<td>Cohort</td>
<td>138 85</td>
<td>22 (16%)</td>
<td>15 (18%)</td>
<td>2 (1.5%)</td>
<td>2 (2%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Initiation of antibiotic treatment**

<table>
<thead>
<tr>
<th>Author / year</th>
<th>Study group</th>
<th>Study Design</th>
<th># of patients</th>
<th>CRP</th>
<th>Control</th>
<th>CRP</th>
<th>Control</th>
<th>CRP</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diederichsen et al. 2000</td>
<td>Adults</td>
<td>RCT</td>
<td>414 398</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cals et al. 2010</td>
<td>Adults</td>
<td>RCT</td>
<td>129 129</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cals et al. 2013</td>
<td>Adults</td>
<td>RCT</td>
<td>203 176</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little et al. 2013</td>
<td>Adults</td>
<td>RCT</td>
<td>2224 2040</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do et al. 2016</td>
<td>Adults</td>
<td>RCT</td>
<td>507 501</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Llor et al. 2012</td>
<td>Adults</td>
<td>Cohort</td>
<td>545 4840</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rebndorf et al. 2016</td>
<td>Children</td>
<td>RCT</td>
<td>138 259</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franz et al. 2004</td>
<td>Neonates</td>
<td>RCT</td>
<td>656 635</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, we also based our conclusion on the fact that the 95%CI for the meta-analysis results for the aforementioned outcomes were not only non-significant, but also narrow enough to claim safety:

- **Mortality – neonatal studies (antibiotic duration):**
  - RCTs: RD 0% (95% CI -4, 4)  
  - Cohort studies: RD -5% (95% CI -10, 0)
- **Mortality – adult studies (antibiotic initiation):** zero mortality in all studies
- **Infection relapse – neonatal studies (antibiotic duration):**
  - RCTs: RD -4% (95% CI -12, 3)
  - Cohort studies: RD -1% (95% CI -4, 3)
- **Hospitalization – adult studies (antibiotic initiation):** RD 0% (95% CI -0.00, 0.01)

Considering the upper limit of the 95%CI, the use of CRP-based algorithms could potentially increase infection relapse and mortality in neonates by 3% and 4%, respectively. Following the reviewer’s
suggestion of analyzing the results against a non-inferiority margin, we can say that both results would be under a non-inferiority margin of 5%. Such results were obtained despite the fact that the RCTs included in the meta-analysis were small (which could lead to a wider 95%CI).

Thus, to clarify our conclusion regarding safety and also, as suggested by Mr. Vangala, to help the clinical audiences to evaluate safety using their own thresholds, we added the following sentences to the Discussion section:

“… Importantly, the use of CRP algorithms to guide antibiotic treatment does appear to be safe, as neonatal studies using CRP to determine duration of antibiotic treatment showed no difference in mortality or in infection relapse. Furthermore, adult studies that used CRP to guide antibiotic initiation showed no significant differences in mortality and hospitalization rates.”

“… Our meta-analysis showed that the use of CRP-based algorithms to determine antibiotic treatment duration seems to be safe for neonates as it did not impact mortality or infection relapse, using a non-inferiority margin of 5%. Furthermore, no deaths were observed and hospitalization rates were not statistically different (by a non-inferiority margin of 1%) in adult studies that used CRP to guide antibiotic initiation…”

2. Minor points:

2.1. Though the authors exercise great care in stratifying different study designs and patient populations, this may result in low power for many of these analyses. It would be illuminating to see what these results look like when pooled in the context of a regression model, where the stratification factors are included as covariates instead of defining entirely separate metaanalyses. This may result in tighter confidence intervals, allowing clearer conclusions to be drawn about safety, alongside the more conservative results already reported in the paper.

We thank Mr. Vangala for this interesting comment. We agree that the use of meta-regression would allow us to increase the number of observation points, which would lead to tighter 95%CI. However, we decided not to do it because the study populations were very heterogeneous regarding their risk of mortality (e.g., mortality risk for neonates with sepsis is much higher compared to mortality risk for adult outpatients with non-severe respiratory tract infections), as well as the CRP cut-off used (adults vs. neonates). In addition, pooling observational studies and RCTs could introduce more bias, something that we would like to avoid.

2.2. It would be interesting to see how these treatment effects vary by the CRP cutoff used. It is possible that some cutoffs may be too aggressive, and so safety may be easier to demonstrate in studies looking at less aggressive cutoffs.

Thank you for the very interesting point. As previously mentioned, all cut-offs used in the included studies once these were stratified by patient population are very similar. Therefore, we will not be able to answer this question.

2.3. How sensitive are the results to the inclusion/exclusion of Little et. al. (2013) and Llor et. al. (2012), each of which has a sample size far larger than any other study? These are likely dominating the relevant adult meta-analyses.

We have performed the suggested sensitivity analysis and re-run our meta-analysis including adult RCTs that used CRP algorithms to guide antibiotic treatment initiation after having removed the study of Little et al. (see forest plot below). The results we have obtained re the following:

- Including Little et al.: RD -0.07 (-0.10, -0.04), I² 37.9%
- Excluding Little et al.: RD -0.07 (-0.11, -0.02), I² 50.3%
Thus, the results were still significant but presented, as expected, higher statistical heterogeneity. Regarding the study of Lior et al.: this was the only cohort study for initiation of antibiotic in adult patients. Therefore, it was not included in any meta-analysis.
that the discussion makes various safety claims, I would additionally urge inclusion of the assumed non-inferiority margins on which these claims are based in the statistical methods section, rather than being introduced in the discussion.

3. The authors did not respond to my observation that the only RCT in adults looking at duration of antibiotic treatment actually observed a higher mortality rate in the CRP arm, with a confidence interval wide enough to breach any reasonable non-inferiority margin (-14%, 17%). The adult RCTs looking at initiation of antibiotic treatment did not see a higher mortality rate in the CRP arm, but only because there were no deaths in any of these studies (meaning they provide no information about the difference in mortality risks). These results suggest that there is a great degree of uncertainty about the effect of this intervention on mortality risk in adults, and so I remain skeptical about safety conclusions in adults that make any reference to mortality. At a minimum, I think a clear disclaimer on this point should be included in the discussion section, and ideally in the abstract as well.

4. The responses to my minor points were all satisfactory.

VERSION 2 – AUTHOR RESPONSE

Overall, it seems that sufficient information is included in the current draft for a discerning reader to understand what this valuable meta-analysis actually tells us about the safety and other effects of CRP-guided antibiotic treatment. I remain concerned, however, that the characterization of the conclusions about safety may leave less careful readers with the impression that we have some evidence of safety in places we really have none, particularly in adults with regards to mortality (see point 3 below). My specific responses to the authors’ comments are below:

1. The authors explain in their response that a null hypothesis of no difference in initiation or duration of antibiotic therapy is worth evaluating, given that physicians do not consistently follow biomarker-guided treatment algorithms. Surely this is true, but it seems the authors agree that if any physician ever follows the algorithm, we would see a reduction in initiation rates and mean duration, leaving only the “how much” question remaining. Thus my original concern about framing this question as a significance test rather than an estimation exercise remains.

Response:
Thank you for this comment. We agree with the reviewer and we have removed all mention of significance testing throughout the manuscript to not confuse readers with our reporting of estimates.

2. The authors have modified some of the language in the discussion section to clarify the basis of their safety conclusions. These changes are welcome, but I would further urge them to remove phrases like “no significant difference” from any sections pertaining to safety. Non-significance, by itself, tells us nothing about safety. Most often, it is a product of insufficient power. Given that the discussion makes various safety claims, I would additionally urge inclusion of the assumed non-inferiority margins on which these claims are based in the statistical methods section, rather than being introduced in the discussion.

Response:
We agree with the reviewer request and, as outlined above, we removed mention of significance testing as well as “no significant difference” for our results. Additionally, we commented on the observed heterogeneity of our results in the context of sample size. As well, we made mention of the non-inferiority margins in the methods as suggested.

We have modified the text as follows:

Methods: “... When assessing safety outcomes, we used non-inferiority margins of 5% for infection relapse and hospitalization, and 2% for mortality.”
Discussion: “... This is important since the prolonged use of antibiotics in infants without culture-proven infection has been associated with higher risk of mortality or morbidity.\(^5\) However, while the non-inferiority margin for mortality of cohort studies was 0%, the non-inferiority margin of the two included RCTs was 5%. The heterogeneity of such results, due to the relatively small sample sizes of the RCTs (n=82 and n= 22)\(^{21} 40\), demonstrates the need for further studies of larger sample size to evaluate the safety of using CRP based algorithms to guide antibiotic treatment duration for these patients.”

3. The authors did not respond to my observation that the only RCT in adults looking at duration of antibiotic treatment actually observed a higher mortality rate in the CRP arm, with a confidence interval wide enough to breach any reasonable non-inferiority margin (-14%, 17%). The adult RCTs looking at initiation of antibiotic treatment did not see a higher mortality rate in the CRP arm, but only because there were no deaths in any of these studies (meaning they provide no information about the difference in mortality risks). These results suggest that there is a great degree of uncertainty about the effect of this intervention on mortality risk in adults, and so I remain skeptical about safety conclusions in adults that make any reference to mortality. At a minimum, I think a clear disclaimer on this point should be included in the discussion section, and ideally in the abstract as well.

Response:
Thank you for this comment. We have revised our statements in the discussion regarding safety in adults to reflect the breach of non-inferiority and urged caution in interpretation of our results as both the numbers of outcomes and samples sizes were low in studies of adults. Similarly, we have also modified the abstract to indicate the need for more data on safety.

We have modified the text as follows:

Discussion: “... Nevertheless, the non-inferiority margin for mortality in the study\(^{41}\) evaluating the use of CRP algorithms to guide duration of antibiotic treatment in this patient population was 18%, which breaches any reasonable non-inferiority margin to determine safety. Finally, due to the low number of deaths and relapses observed in neonates and adults, we should interpret the aforementioned safety results with caution.”

Abstract: “... Conclusion: The use of CRP-based algorithms seems to reduce antibiotic treatment duration in neonates, as well as to decrease antibiotic treatment initiation in adult outpatients. However, further high-quality studies are still needed to assess safety, particularly in children outside the neonatal period.”

4. The responses to my minor points were all satisfactory.

Response:
Thank you.