PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Optimal duration of antibiotic treatment for community-acquired pneumonia in adults: a systematic review and duration-effect meta-analysis
AUTHORS	FURUKAWA, YUKI; Luo, Yan; Funada, Satoshi; Onishi, Akira; Ostinelli, Edoardo; Hamza, Tasnim; Furukawa, Toshi; Kataoka, Yuki

VERSION 1 – REVIEW

REVIEWER	Furlan, Ludovico University of Milan, Department of Clinical Sciences and Community Health
REVIEW RETURNED	08-Mar-2022

GENERAL COMMENTS	Dear Editor, thank you for the opportunity to review the manuscript entitled “Optimal duration of antibiotic treatment for community- acquired pneumonia in adults: a systematic review and duration-effect meta-analysis” submitted by Furukawa et al. for possible publication in BMJ Open. The article assesses through a systematic review with meta-analysis of current evidence from RCT the optimal treatment duration with antibiotics for community- acquired pneumonia (CAP) in adults. The authors use a dose-effect meta-analysis, that, in their intentions, permits to evaluate the best antibiotic duration strategy according to available data. Based on their results the authors conclude that 3-9 days treatment duration is likely to be non-inferior to 10 days duration (if clinical stability is achieved). They also suggest that a shorter duration of 3-5 days may also be safe. Certainty of results is significantly influenced by the paucity of available data. The issue assessed in the article is of great interest, with possible significant impact in clinical practice and would thus be of great interest for readers. The authors should be praised for the extensive number of analyses made. The article is also well written. Nevertheless, I have several comments that the authors may consider. First, I wonder If definition of reference standard of antibiotic treatment duration may be considered appropriate. The authors decided to set 10 days as the reference, according to what

	reported from previous studies on the use of antibiotics in CAP in clinical practice. Nevertheless, previous meta-analysis had already demonstrated comparable outcomes with less than seven days compared with more than seven days antimicrobial therapy (Li et al. Am J Med 2007; Dimopoulos et al. Drugs 2008; Tansarli et al. Antimicrob Agents Chemother 2018; Furlan et al. Int and Emerg Med 2018). This is, indeed, in line with the cut-off used in most of the include studies, even the most recent ones. I thus wonder if the authors should evaluate assessing effectiveness with a shorter reference cut-off. The authors may also consider citing previously published studies and discussing differences with meta-analyses. The manuscript is well written but I really had a hard time trying to interpret the results of the dose-effect meta-analysis (figure 2). I do not have adequate expertise in such analysis, but the authors may consider providing clarification on how the model works for the benefit of readers not used to complex or new/uncommon statistical methods. As an example Y axis in figure 2 has different scale for values below and above the reference OR 1.0, is that related on how the "model" works? Are results reported in a logarithmic scale? Also rate of clinical improvement at 10 days is reported to be 68% but it is not clear to me how the authors obtain that number, is it an estimate derived from the "model"? Based on data reported in table 1 I would consider such number to be 64% for aggregate data (i.e. 257/399). Authors may consider clarifying how results should be interpreted from a clinical point of view. In particular, how figure 2 should be interpreted. - Duration is considered as a continuous variable even if all studies evaluated outcomes in two groups treated with different duration of the same antibiotic. - I also wonder if differences among included patients may limit the validity of the study results. Indeed, almost 20% of patients were treated as outpatients, that may have different characteristics compared to hospitalized patients. - Confidence interval of OR seems to get smaller closer to 10 days treatment. According to figure 2, OR for nine days treatment duration seems to have a relative smaller confidence interval compared to three days treatment duration. Nevertheless, none of the studies included evaluated a 9 days treatment strategy, it is thus surprising that "accuracy" of OR is somehow higher for such duration. - It also seems that the shorter the duration of the antibiotic treatment the higher the cure rate. This result is surprising considering that none of the single included studies has shown such an effect. The authors may consider clarifying such finding. Minor comments: - In most included studies outcome of interest is not only "clinical improvement" but also "clinical resolution or cure".

REVIEWER	Eljaaly, Khalid King Abdulaziz University
REVIEW RETURNED	14-Mar-2022

GENERAL COMMENTS	The meta-analysis is of clinical relevance and this question is a hot topic recently. I have only minor comments for this meta-analysis: Design (both in abstract and manuscript methods section): Remove “to present” Results & methods: “Dose-effect”: Did you mean “Duration-effect” Instead? Figure I: Which other methods did you use (Web search)? Just trying to ensure the reproducibility of this systematic review. Table 1: There is an issue with the alignment of duration with each study. Would make it clear each study is comparing what vs what. Discussion: Would clearly mention that mixing outpatients with inpatients is a limitation.

REVIEWER	Ceccato, Adrian Parc Taulí Research and Innovation Institute
REVIEW RETURNED	01-Sep-2022

GENERAL COMMENTS	The authors did a systematic review and duration-effect meta-analysis about the duration of antibiotic treatment for CAP. The authors apply a novel statistical method for evaluating dose-response. I have no previous experience with this method, so I suggest sending the study to expert statistical reviewers. In general, the results don't change current knowledge about the topic. The studies included in the meta-analysis have a moderate to high risk of bias, and the conclusion obtained included a wide range of days (3-9). My main concerns about the study included: - The primary outcome (clinical improvement on day 15) is not hard, and its definition may differ between studies. - The controlled arms of the studies were not homogeneous. The authors chose arbitrary 10 days as reference. -Few patients had high severity scores, and critically ill patients were excluded, so the probability of success was higher. It is still unknown how short antibiotic therapy would impact the sickest patients. I suggest to the authors, -Improve the discussion, including the limitations mentioned above. - Discuss in more detail the results of the sensitivity analyses. - Include patients admitted to ICU. - The figs. are hard to understand. They are not the usual figs observed in meta-analysis and require better explanations of the results. The non-inferiority margin is not present in all the figs.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Ludovico Furlan, University of Milan, IRCCS Foundation Maggiore Policlinico Hospital
Comments to the Author:
Dear Editor,

thank you for the opportunity to review the manuscript entitled “Optimal duration of antibiotic treatment for community- acquired pneumonia in adults: a systematic review and duration-effect meta-analysis” submitted by Furukawa et al. for possible publication in BMJ Open.

The article assesses through a systematic review with meta-analysis of current evidence from RCT the optimal treatment duration with antibiotics for community-acquired pneumonia (CAP) in adults. The authors use a dose-effect meta-analysis, that, in their intentions, permits to evaluate the best antibiotic duration strategy according to available data. Based on their results the authors conclude that 3-9 days treatment duration is likely to be non-inferior to 10 days duration (if clinical stability is achieved). They also suggest that a shorter duration of 3-5 days may also be safe. Certainty of results is significantly influenced by the paucity of available data.

The issue assessed in the article is of great interest, with possible significant impact in clinical practice and would thus be of great interest for readers.
The authors should be praised for the extensive number of analyses made. The article is also well written.

RESPONSE
We thank Dr Furlan for the positive comments.

Nevertheless, I have several comments that the authors may consider.

First, I wonder If definition of reference standard of antibiotic treatment duration may be considered appropriate. The authors decided to set 10 days as the reference, according to what reported from previous studies on the use of antibiotics in CAP in clinical practice.
Nevertheless, previous meta-analysis had already demonstrated comparable outcomes with less than seven days compared with more than seven days antimicrobial therapy (Li et al. Am J Med 2007; Dimopoulos et al. Drugs 2008; Tansarli et al. Antimicrob Agents Chemother 2018; Furlan et al. Int and Emerg Med 2018). This is, indeed, in line with the cut-off used in most of the include studies, even the most recent ones. I thus wonder if the authors should evaluate assessing effectiveness with a shorter reference cut-off.
The authors may also consider citing previously published studies and discussing differences with meta-analyses.

RESPONSE
We thank the reviewer for the opportunity to clarify this point.
Typically, in a dose-effect meta-analysis, the reference dose is set to the zero (placebo) or the minimal dose examined to make the interpretation easier. We set 10-day treatment as the reference, because this study had the nature of non-inferiority test of shorter duration range. However, we agree with Dr Furlan that many readers would like to know if shorter treatment duration is non-inferior to their own standard duration. We therefore made a league table. (eAppendix10) The league table shows that whatever combination (from three to ten days) we choose, the shorter treatment is likely to be non-inferior to the longer treatment.

3-day _ _ _ _ 1.48
(0.93-2.34) _ _
1.09
(0.95-1.25) 4-day _ _ _ _ _ _
1.19
(0.90-1.57) 1.09
(0.95-1.25) 5-day _ 1.10
(0.74-1.64) _ _ 1.21
(0.89-1.64)
1.29
(0.86-1.93) 1.18
(0.91-1.54) 1.08
(0.96-1.23) 6-day _ 0.63
(0.27-1.49) _ _
1.36
(0.86-2.15) 1.25
(0.91-1.72) 1.15
(0.96-1.38) 1.06
(1.00-1.13) 7-day _ _ 1.84
(0.47-7.25)
1.39
(0.93-2.09) 1.28
(0.97-1.69) 1.18
(1.00-1.38) 1.08
(0.97-1.21) 1.02
(0.92-1.13) 8-day _ _
1.42
(0.99-2.03) 1.30
(1.01-1.68) 1.19
(0.97-1.46) 1.10
(0.88-1.38) 1.04
(0.83-1.30) 1.01
(0.89-1.15) 9-day _
1.44
(1.01-2.05) 1.32
(0.98-1.77) 1.21
(0.90-1.63) 1.12
(0.79-1.58) 1.05
(0.74-1.50) 1.03
(0.80-1.33) 1.01
(0.89-1.15) 10-day
Results of the duration-effect meta-analysis are shown in the bottom-left area. Results of the pairwise meta-analyses of direct comparisons are shown in the upper-right area. Data are odds ratios (95% confidence interval) of the upper-left treatment duration compared with the bottom-right treatment duration. Non-inferior results (lower bound of the 95% confidence interval higher than 0.65) are shown in light green colour.
[eAppendix10]

We appreciate the reviewer's suggestion to expand our discussion for the previous literature. As suggested, we updated the citation and discussed more about the difference of this study compared to the previous studies as follows:

This is in line with the previous pairwise meta-analyses that showed shorter duration was non-inferior to longer duration. [Dimopoulos et al, 2008; Tansarli et al, 2018; Furlan et al, 2018] We updated the systematic review and found four trials that were not included in the previous studies. This allowed us to focus on trials that used the same antibiotics with the same daily dosage. The previous studies included trials using different antibiotics or different daily dosages, so the results may not have reflected the differences in treatment durations alone. Moreover, they subcategorised the treatment durations and may have thus lost some statistical power to detect meaningful differences among durations. We overcame this limitation by examining the duration of antibiotic treatment range in days as a continuous variable and found that three to nine-day treatment is likely to be non-inferior to 10-day treatment. [Page 23, line 349-358; DISCUSSION]

The manuscript is well written but I really had a hard time trying to interpret the results of the dose-effect meta-analysis (figure 2).
I do not have adequate expertise in such analysis, but the authors may consider providing clarification on how the model works for the benefit of readers not used to complex or new/uncommon statistical methods.

RESPONSE
We appreciate the time and effort spent by Dr. Furlan. Indeed, the dose-effect meta-analysis is still relatively new and readers would need more explanation. We described the duration-effect meta-analysis in more detail as below.

In the duration-effect meta-analysis, we assumed that the relative efficacy of a certain treatment duration (〖duration〗_i) against another (〖duration〗_j) can be expressed in the log-odds ratio (log⁡〖OR_ij〗) and that it is a function of both durations (log⁡〖OR_ij〗=f(〖duration〗_i; 〖duration〗_j)). We fitted restricted cubic splines with three knots to the dataset obtained by the systematic review because this model has shown sufficient flexibility to capture different shapes.[36] Given the clinical and methodological heterogeneity likely present in the included studies, we used the random effects model. We used 3 knots, equally spaced across the duration range (25%, 50%, 75%). Typically, in dose-effect meta-analyses, the reference dose is assigned to the zero or the minimal dose to make interpretation easier. As this duration-effect meta-analysis aimed to test the non-inferiority of shorter treatment duration, we decided to use the maximum duration as the reference to make interpretation easier.
[Page 11-12, line 204-215; METHODS/Duration-effect meta-analysis]

As an example, Y axis in figure 2 has different scale for values below and above the reference OR 1.0, is that related on how the “model” works? Are results reported in a logarithmic scale?

RESPONSE
Yes, the Y-axis uses a logarithmic scale. This is the same as the forest plot of conventional pairwise meta-analysis for binary outcomes. We computed in log ORs but presented the result in table 2 in original ORs by exponentiating the obtained log ORs. We added an explanation in the results section as below:

The x-axis of the figures represents the treatment duration in days. The y-axis represents the odds ratio of the outcome in a logarithmic scale, just as in the forest plot of conventional pairwise meta-analysis using binary outcomes.
[Page 18, line 286-288; RESULTS/Duration-effect meta-analysis]

We also noticed that the different range of Y-axis of figures (0.1 to10, 0.2 to 5, 0.5 to 2…) might have been confusing. We set the Y-axis of all graphs to be 0.1 to 10 (See e.g. Figures 2 and 3). We believe

that this allows readers to understand the widths of confidence intervals more easily (narrow for clinical improvement on day 15 and day 30, wide for all-cause mortality).

Also rate of clinical improvement at 10 days is reported to be 68% but it is not clear to me how the authors obtain that number, is it an estimate derived from the “model”? Based on data reported in table 1 I would consider such number to be 64% for aggregate data (i.e. 257/399).

RESPONSE
We thank Dr Furlan for the careful read of our manuscript. We calculated this number by metaprop function of meta R package. It is a random effects meta-analysis of single proportions to calculate the overall proportion. This calculation is independent of the model used in the main duration-effect meta-analysis. We revised the sentence to make it clear that we used meta-analytical method to calculate the percentage, instead of arithmetic additions and divisions.

The weighted average percentage of clinical improvement rate on day 15 in the 10-day treatment arms was 68%, based on a single proportion meta-analysis of the included studies. [Page 20, line 317-319; RESULTS/Duration-effect meta-analysis]

Authors may consider clarifying how results should be interpreted from a clinical point of view. In particular, how figure 2 should be interpreted.

RESPONSE
We appreciate the opportunity to make our paper clinically more relevant. Odds ratio is an excellent measure to summarise, but it is not readily interpretable. We therefore translated ORs into absolute event rates using the weighted average control event rates discussed above. (Table 2) We presented how many arms were included in the single-proportion meta-analyses, but we noticed this was confusing and adds not much information. We omitted these data.

Odds ratios need to be translated into absolute event rates so that the results can be interpreted from the clinical point of view. The weighted average percentage of clinical improvement rate on day 15 in the 10-day treatment arms was 68% (based on a single proportion meta-analysis of the included studies). Using this average, we computed the absolute clinical improvement rates at the following durations as follows: 3-day treatment 75% (95%CI: 68 to 81%), 5-day treatment 72% (66 to 78%), and 7-day treatment 69% (61 to 76%). (Table 2) [Page 20, line 316-322; RESULTS/Duration-effect meta-analysis]

Table 2 Primary and secondary outcomes for 3, 5, 7 and 10-day treatment

Outcome	Treatment duration (days)			
	3	5	7	10 (Reference)
Clinical improvement at day 15	OR 1.44 [1.01-2.05]	1.21 [0.90-1.63]	1.05 [0.74-1.50]	1.00 (reference)
Rate	75% [68-81%]	72% [66-78%]	69% [61-76%]	68% -
All-cause mortality	OR 1.11 [0.28-4.35]	0.93 [0.34-2.58]	0.84 [0.23-3.09]	1.00 (reference)
Rate	3% [1-11%]	3% [1-7%]	2% [1-8%]	3% -
Serious adverse events	OR 0.73 [0.27-1.96]	0.80 [0.51-1.24]	0.86 [0.40-1.85]	1.00 (reference)
Rate	15% [6-31%]	16% [11-22%]	17% [9-30%]	19% -
Clinical improvement at day 30	OR 1.24 [0.86-1.78]	1.16 [0.82-1.63]	1.09 [0.74-1.60]	1.00 (reference)
Rate	81% [74-86%]	80% [74-85%]	79% [73-84%]	77% -

- Duration is considered as a continuous variable even if all studies evaluated outcomes in two groups

treated with different duration of the same antibiotic.

RESPONSE
In conventional pairwise meta-analyses, duration needed to be arbitrarily dichotomised into a binary variable (short vs long). In contrast, we treated duration as a continuous variable. This enabled us to use dose-effect meta-analysis, which considers the results of individual trials in a two-dimensional space. The duration-effect meta-analysis fits a flexible curve to the dataset in a two-dimensional space, thus borrowing strength of the comparisons from adjacent treatment durations. In other words, the duration-effect MA takes advantage of the clinical natural assumption that the effect sizes are more closely associated between 7 days vs 10 days than between 5 days vs 10 days. In conventional meta-analyses that arbitrarily divides the treatment durations into subcategories, the above two comparisons might be pooled in the same meta-analysis (thus ignoring potential differences between 5 days vs 10 days and 7 days vs 10 days) or used in different meta-analyses (thus losing the important information that these effect sizes may be correlated).

In order to understand how the duration-effect meta-analysis summarised the included trials, we compared the result of each included trial, pairwise meta-analyses and the duration-effect meta-analysis as follows (eAppendix7 and 10).

The relative efficacy of 7-day vs 10-day was examined in a single trial (Siegel 1999) and the result was OR 1.81 (95%CI; 0.47 to 7.25), whereas the duration-effect meta-analysis showed OR 1.05 (0.74 to 1.50).

5-day vs 10-day was examined in three trials (Leophonte2002, Stralin2014, Uranga2016). The pairwise meta-analysis showed OR 1.21 (0.89 to 1.64), whereas the duration-effect meta-analysis showed OR 1.21 (0.90 to 1.63).

6-day vs 8-day was examined in a single trial (Aliberti2017) and the result was OR 0.68 (0.27 to1.49). The duration-effect meta-analysis showed OR 1.08 (0.97 to 1.21). The direction was the opposite, but this is understandable because this treatment range was the only treatment range in which shorter duration showed a worse outcome, in contradistinction to adjacent studies.

5-day vs 7-day was examined in two trials (Tellier2004, File2007). The pairwise meta-analysis showed OR 1.10 (0.74 to 1.64) whereas duration-effect meta-analysis showed OR 1.15 (0.96 to 1.38).

3-day vs 8-day was examined in two trials (ElMoussaoui2006, Dinh2021). The pairwise meta-analysis showed OR 1.48 (0.93 to 2.34). Duration-effect meta-analysis showed OR 1.39 (1.93 to 2.09).

All in all, the duration-effect meta-analysis combined the results of included trials in a reasonable way and clarified the overall duration-effect relationship.

- I also wonder if differences among included patients may limit the validity of the study results. Indeed, almost 20% of patients were treated as outpatients, that may have different characteristics compared to hospitalized patients.

RESPONSE
We agree with Dr Furlan that the inclusion of both the outpatients and inpatients may result in clinical heterogeneity and may limit the validity of the study results. We have therefore conducted a series of sensitivity analyses focusing on inpatients and the results were generally in line with the primary analyses. We added an explanation in the limitations section as follows:

We included both the outpatients and inpatients, which may limit the validity of the study results. However, the overall statistical heterogeneity was low and sensitivity analyses excluding trials with outpatients generally confirmed the primary analyses (eAppendix11). [Page 25, line 382-385; DISCUSSION/Limitations]

- Confidence interval of OR seems to get smaller closer to 10 days treatment. According to figure 2, OR for nine days treatment duration seems to have a relative smaller confidence interval compared to three days treatment duration. Nevertheless, none of the studies included evaluated a 9 days treatment strategy, it is thus surprising that “accuracy” of OR is somehow higher for such duration.

RESPONSE
We appreciate the opportunity to explain this point. Yes, 95%CI gets narrower when the duration range was examined by many trials. There can be another reason why 95%CI gets narrower; when it is close to the reference point (10-day treatment, in this case). 10-day treatment is the reference point in this case and the OR is naturally 1, with zero 95%CI. A hypothetical treatment duration, say, 9.99-day treatment would be almost identical to the 10-day treatment and thus the OR would be almost 1.00 with very very narrow 95%CI. This is why 9-day treatment had relatively narrow 95%CI. We added an explanation as follows:

The 95% CI band becomes narrower when the duration range was examined by many trials or when it gets closer to the reference point. [Page 19, line 298-300; RESULTS/Duration-effect meta-analysis]

Please see the league table (eAppendix10, attached above in this response letter), which also shows narrower CIs for the comparisons of close durations.

- It also seems that the shorter the duration of the antibiotic treatment the higher the cure rate. This result is surprising considering that none of the single included studies has shown such an effect. The authors may consider clarifying such finding.

RESPONSE
We agree with Dr. Furlan that this finding is surprising and worth discussing further. We believe that this could be because the duration-effect meta-analysis treated duration as a continuous variable and borrowed strength from adjacent comparisons. Thus, for example, when 5-day treatment is non-statistically-significantly superior to 10-day treatment, and 3-day treatment is non-statistically-significantly superior to 8-day treatment, 3-day treatment could turn out to be statistically significantly superior to 10-day treatment. However, we believe that this finding needs to be further examined in future trials. We added the following in the discussion section.

Possibility of 3-day treatment being superior to 10-day treatment should be carefully interpreted, as none of the included trials, previous meta-analyses or the pairwise meta-analysis of the included trials (eAppendix7, post hoc analysis) showed the superiority of shorter treatment duration. This could be explained by the fact that most of the combinations of treatment durations examined (7 days vs 10 days, 5 days vs 10 days, 5 days vs 7 days, 3 days vs 8 days) suggested better efficacy of shorter durations, if not statistically significant alone. (eAppendix7, post hoc analysis) The duration-effect meta-analysis combined these findings, leading to the possible superiority of the shortest duration examined (3 days) over the longest duration examined (10 days). Further research focusing on the shorter duration range is warranted to confirm this finding. [Page 24, line 363-373/DISCUSSION]

Minor comments:
- In most included studies outcome of interest is not only “clinical improvement” but also “clinical resolution or cure”.

RESPONSE
We thank the reviewer for pointing this out. We added a description about this in the main text and a summary table in the appendix.
Clinical improvement was often described as “clinical cure” or “clinical success” and was often defined as resolution of fever and improvement of symptoms related to pneumonia without further antibiotics. More detailed definitions of clinical improvement in each included study is listed in the appendix. (eAppendix6) [Page 14, line 253-257; RESULTS]
Study Definition
Siegel et al, 1999 “Patients were classified as a cure if the pneumonia was successfully treated within the constraints of the study protocol, including resolution of fever and leukocytosis and substantial improvement in chest radiograph by day 42”
Léophonte et al, 2002 “The main criteria defining success were apyrexia on D10 (temperature 37.5◦C) and no other antibiotic treatment before D10. The secondary criteria were absence of clinical signs on D10, cure (normalized clinical status and radiological imagery on D30/D45), and no other antibiotic treatment before D30/D45.”
Tellier et al, 2004 “Clinical cure was defined as either the return to the pre-infection state (i.e. all pneumonia-related signs and symptoms had disappeared and chest X-ray findings had shown improvement) or improvement in related post-infectious stigmata, such that residual symptoms if any did not require additional treatment and were accompanied by improvement or lack of progression based on chest X-ray.”
El Moussaoui et al, 2006 “Cure—resolution or improvement of symptoms and clinical signs related to pneumonia without the need for additional or alternative antibiotic therapy”
File et al, 2007 “Clinical response was based on subjective symptoms and objective signs of auscultatory findings (rales, rhonchi, wheezing and breath sounds) and was defined as success (sufficient improvement or resolution of the signs and symptoms of CAP recorded at baseline such that no additional antibacterial therapy was required at the end of therapy or follow-up)”
Strålin et al, 2014 “Clinical cure”
Uraga et al, 2014 “The primary outcomes were clinical success rate at day 10 and late follow-up (day 30) since admission, defined as resolution or improvement in signs and symptoms related to pneumonia without further antibiotics, and CAP-related symptoms at day 10 measured with the 18-item CAP symptom questionnaire, a specific and validated patient-reported outcome measure on which higher scores indicate more severe symptoms (range, 0-90).”
Aliberti et al, 2017 “Early failure was the primary composite study outcome occurring within 30 days following CAP diagnosis and including any of the following conditions: 1) pneumonia related complications (e.g., lung abscess, empyema); 2) clinical failure during hospitalization (definition in the online data supplement); 3) a new antibiotic course after discontinuation of antibiotic therapy prescribed for the pneumonia, 4) re-hospitalization from any reason; 5) death from any reason.”
Dinh et al, 2021 “Cure was defined by the following criteria: apyrexia (temperature ≤37·8°C); resolution or improvement of clinical signs or symptoms (coughing frequency or severity, sputum production, dyspnoea, crackles); and no additional antibiotic treatment (for community-acquired pneumonia or any reason) since the last follow-up visit.”

Reviewer: 2
Dr. Khalid Eljaaly, King Abdulaziz University
Comments to the Author:
The meta-analysis is of clinical relevance and this question is a hot topic recently.
I have only minor comments for this meta-analysis:

RESPONSE
We appreciate Dr. Khalid Eljaaly’s positive response.

Design (both in abstract and manuscript methods section):
Remove “to present”

RESPONSE
We dropped “to present” both from the abstract and main text as suggested.

Results & methods: “Dose-effect”: Did you mean “Duration-effect” Instead?

RESPONSE
We thank Dr. Khalid Eljaaly for the careful consideration of our manuscript. We changed the “Dose-effect meta-analysis” to “Duration-effect meta-analysis” in the methods section.

Figure I:
Which other methods did you use (Web search)? Just trying to ensure the reproducibility of this systematic review.

RESPONSE
We appreciate the opportunity to clarify this point. It was a document published by the European Medicines Agency that we found by web searching, and we searched the sponsor’s protocol code number that we found on EUCTR. We reorganised and revised the list of included studies section in the eAppendix as follows:

4.1. List of studies included in the analyses

Aliberti2017
- Aliberti S, Ramirez J, Giuliani F, et al. Individualizing duration of antibiotic therapy in community-acquired pneumonia. Pulm Pharmacol Ther 2017; 45: 191–201.
- NCT01492387

Dinh2021
- Dinh A, Ropers J, Duran C, et al. Discontinuing β-lactam treatment after 3 days for patients with community-acquired pneumonia in non-critical care wards (PTC): a double-blind, randomised, placebo-controlled, non-inferiority trial. Lancet 2021; 397: 1195–203.
- NCT01963442

ElMoussaoui2006
- El Moussaoui R, Borgie C, Broek P, et al. Effectiveness of discontinuing antibiotic treatment after three days versus eight days in mild to moderate-severe community acquired pneumonia: randomised, double blind study. BMJ 2006; 332: 1355.

File2007
- File TM, Mandell LA, Tillotson G, et al. Gemifloxacin once daily for 5 days versus 7 days for the treatment of community-acquired pneumonia: a randomized, multicentre, double-blind study. J Antimicrob Chemoth 2007; 60: 112–20.
- European Medicines Agency. Withdrawal assessment report for factive. 2009. (https://www.ema.europa.eu/en/documents/withdrawal-report/withdrawal-assessment-report-factive_en.pdf; Last accessed on 25 September 2022) *
- EUCTR2004-002619-10-CZ

Uranga2016
- Uranga A, España PP, Bilbao A, et al. Duration of Antibiotic Treatment in Community-Acquired Pneumonia: A Multicenter Randomized Clinical Trial. JAMA Intern Med. 2016; 176: 1257.
- Uranga A, Artaraz A, Bilbao A, et al. Impact of reducing the duration of antibiotic treatment on the long-term prognosis of community acquired pneumonia. BMC Pulm Med. 2020;20(1):261.

Leophonte2002
- Léophonte P, Choutet P, Gaillat J, et al. Efficacité comparée de la ceftriaxone dans un traitement de dix jours versus un traitement raccourci de cinq jours des pneumonies aigues communautaires de l'adulte hospitalisé avec facteur de risque. Médecine Et Maladies Infect 2002; 32: 369–81.

Siegel1999
- Siegel RE, Alicea M, Lee A, Blaiklock R. Comparison of 7 Versus 10 Days of Antibiotic Therapy for Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia. Am J Ther 1999; 6: 217–22.

Stralin2014
- Strålin K, Rubenson A, Lindroth H, et al. Betalactam treatment until no fever for 48 hours (at least 5 days) versus 10 days in community-acquired pneumonia: randomized, non-inferiority, open study. Pneumonia 2014; 3: 246–81.
- ISRCTN14523624

Tellier2004
- Tellier G, Niederman MS, Nusrat R, et al. Clinical and bacteriological efficacy and safety of 5 and 7 day regimens of telithromycin once daily compared with a 10 day regimen of clarithromycin twice daily in patients with mild to moderate community-acquired pneumonia. J Antimicrob Chemoth 2004; 54: 515–23.
- Tellier G, Chang JR, Asche CV, Lavin B, Stewart J, Sullivan SD. Comparison of hospitalization rates in patients with community-acquired pneumonia treated with telithromycin for 5 or 7 days or clarithromycin for 10 days. Curr Med Res Opin. 2004;20(5):739-747.

4.2. List of ongoing trials

NCT03609099
- NCT03609099. Adequate Duration of Antibiotic Treatment in Community-acquired Pneumonia With High Risk Class and Adequate Initial Clinical Response (2017-001406-15).
NCT04089787
- NCT04089787. Shortened Antibiotic Treatment of 5 Days in Community-Acquired Pneumonia (CAP5).

* found during web search using the sponsor's protocol code number.
[eAppendix4. List of all included papers]

Table 1:
There is an issue with the alignment of duration with each study. Would make it clear each study is comparing what vs what.

RESPONSE
We appreciate the opportunity to improve readability of our manuscript. We added horizontal lines in Table 1. We believe it is now clearer each study compared what vs what.

Also, we corrected a mistake in the number of participants in the 10-day treatment arm of Stralin2014 (103.5 -> 104).

Discussion:
Would clearly mention that mixing outpatients with inpatients is a limitation.

RESPONSE
We thank Dr. Khalid Eljaaly for the opportunity to make our manuscript more balanced. Indeed, mixing outpatients with inpatients was a limitation. We therefore conducted sensitivity analyses focusing on inpatients and they were generally in line with the primary analyses. We stated in the limitation section as follows:

Forth, baseline severity of the included studies varied. We included both the outpatients and inpatients, which may have concealed important heterogeneity in the study results. However, sensitivity analyses excluding trials with outpatients generally confirmed the primary analyses (eAppendix11) and the overall statistical heterogeneity was low. [page 25, line 381-385. DISCUSSION/ Limitations]

Reviewer: 3
Dr. Adrian Ceccato, Hospital Clinic de Barcelona
Comments to the Author:
The authors did a systematic review and duration-effect meta-analysis about the duration of antibiotic treatment for CAP. The authors apply a novel statistical method for evaluating dose-response. I have no previous experience with this method, so I suggest sending the study to expert statistical reviewers.
In general, the results don't change current knowledge about the topic. The studies included in the meta-analysis have a moderate to high risk of bias, and the conclusion obtained included a wide range of days (3-9).

RESPONSE
We thank Dr. Adrian Ceccato for reviewing our paper carefully.

My main concerns about the study included:
- The primary outcome (clinical improvement on day 15) is not hard, and its definition may differ between studies.

RESPONSE
As pointed out by Dr. Ceccato, the definition of clinical improvement slightly differed between studies. Please note however that the definition of “clinical improvement” was the same within each study, and the relative effect size (OR) of achieving that outcome tends to remain similar when different yet related outcome definitions are used. (Furukawa TA, et al. 2011. http://www.ncbi.nlm.nih.gov/pubmed/21062670) In addition, to account for potential heterogeneity, we used the random effects model which assumes that the study effect sizes follow different yet related distributions. We now added an explanation about it in the main text and the summary table in the appendix as follows:

Clinical improvement was often described as “clinical cure” or “clinical success” and was often defined as resolution of fever and improvement of symptoms related to pneumonia without further antibiotics. More detailed definitions of clinical improvement in each included study is listed in the appendix. (eAppendix6)

[page 14, line 253-257; RESULTS]

Study Definition
Siegel et al, 1999 “Patients were classified as a cure if the pneumonia was successfully treated within the constraints of the study protocol, including resolution of fever and leukocytosis and substantial improvement in chest radiograph by day 42”
Léophonte et al, 2002 “The main criteria defining success were apyrexia on D10 (temperature 37.5◦C) and no other antibiotic treatment before D10. The secondary criteria were absence of clinical signs on D10, cure (normalized clinical status and radiological imagery on D30/D45), and no other antibiotic treatment before D30/D45.”
Tellier et al, 2004 “Clinical cure was defined as either the return to the pre-infection state (i.e. all pneumonia-related signs and symptoms had disappeared and chest X-ray findings had shown improvement) or improvement in related post-infectious stigmata, such that residual symptoms if any did not require additional treatment and were accompanied by improvement or lack of progression based on chest X-ray.”
El Moussaoui et al, 2006 “Cure—resolution or improvement of symptoms and clinical signs related to pneumonia without the need for additional or alternative antibiotic therapy”
File et al, 2007 “Clinical response was based on subjective symptoms and objective signs of auscultatory findings (rales, rhonchi, wheezing and breath sounds) and was defined as success (sufficient improvement or resolution of the signs and symptoms of CAP recorded at baseline such that no additional antibacterial therapy was required at the end of therapy or follow-up)”
Strålin et al, 2014 “Clinical cure”
Uraga et al, 2014 “The primary outcomes were clinical success rate at day 10 and late follow-up (day 30) since admission, defined as resolution or improvement in signs and symptoms related to pneumonia without further antibiotics, and CAP-related symptoms at day 10 measured with the 18-item CAP symptom questionnaire, a specific and validated patient-reported outcome measure on which higher scores indicate more severe symptoms (range, 0-90).”
Aliberti et al, 2017 “Early failure was the primary composite study outcome occurring within 30 days following CAP diagnosis and including any of the following conditions: 1) pneumonia related complications (e.g., lung abscess, empyema); 2) clinical failure during hospitalization (definition in the online data supplement); 3) a new antibiotic course after discontinuation of antibiotic therapy prescribed for the pneumonia, 4) re-hospitalization from any reason; 5) death from any reason.”
Dinh et al, 2021 “Cure was defined by the following criteria: apyrexia (temperature ≤37·8°C); resolution or improvement of clinical signs or symptoms (coughing frequency or severity, sputum production, dyspnoea, crackles); and no additional antibiotic treatment (for community-acquired pneumonia or any reason) since the last follow-up visit.”

[eAppendix6]

- The controlled arms of the studies were not homogeneous. The authors chose arbitrary 10 days as reference.

RESPONSE
We thank Dr. Ceccato for the opportunity to clarify this point. As pointed out, the controlled arms were not homogeneous. They had, for example, different treatment durations. That is exactly where the duration-effect meta-analysis comes in. By assuming treatment-effect relationship, it can summarise results from studies that examined different treatment duration ranges. We acknowledge that we needed to explain more why we chose 10 days as reference. We revised the statistical analysis section as follows:

Duration-effect meta-analysis
In the duration-effect meta-analysis, we assumed that the relative efficacy of a certain treatment

duration (〖duration〗_i) against another (〖duration〗_j) can be expressed in the log-odds ratio (log⁡〖OR_ij 〗) and that it is a function of both durations (log⁡〖OR_ij 〗=f(〖duration〗_i;〖duration 〗_j)). We fitted restricted cubic splines with three knots to the dataset obtained by the systematic review because this model has shown sufficient flexibility to capture different shapes. Given the clinical and methodological heterogeneity likely present in the included studies, we used the random effects model. We used three knots, equally spaced across the duration range (25%, 50%, 75%). Typically, in dose-effect meta-analyses, the reference dose is assigned to the zero or the minimal dose to make interpretation easier.[36] As this duration-effect meta-analysis tested the non-inferiority of shorter treatment duration, we decided to use the maximum duration as the reference to make interpretation easier. Also, we set 10 days as the reference because it can be regarded as the current practice. We tested the non-inferiority with the non-inferiority margin of 10%, as previously proposed, and the superiority of the shorter duration examined against 10 days using the ITT dataset.
[Page 11-12, line 204-218; METHODS/Data collection and analysis/Duration-effect meta-analysis]

We also conducted additional analyses with different reference durations and made a league table. (eAppendix10) It confirmed the non-inferiority of shorter duration compared to longer duration, regardless of the combination we choose.

3-day _ _ _ _ 1.48
(0.93-2.34) _ _
1.09
(0.95-1.25) 4-day _ _ _ _ _ _
1.19
(0.90-1.57) 1.09
(0.95-1.25) 5-day _ 1.10
(0.74-1.64) _ _ 1.21
(0.89-1.64)
1.29
(0.86-1.93) 1.18
(0.91-1.54) 1.08
(0.96-1.23) 6-day _ 0.63
(0.27-1.49) _ _
1.36
(0.86-2.15) 1.25
(0.91-1.72) 1.15
(0.96-1.38) 1.06
(1.00-1.13) 7-day _ _ 1.84
(0.47-7.25)
1.39
(0.93-2.09) 1.28
(0.97-1.69) 1.18
(1.00-1.38) 1.08
(0.97-1.21) 1.02
(0.92-1.13) 8-day _ _
1.42
(0.99-2.03) 1.30
(1.01-1.68) 1.19
(0.97-1.46) 1.10
(0.88-1.38) 1.04
(0.83-1.30) 1.01
(0.89-1.15) 9-day _

1.44
(1.01-2.05) 1.32
(0.98-1.77) 1.21
(0.90-1.63) 1.12
(0.79-1.58) 1.05
(0.74-1.50) 1.03
(0.80-1.33) 1.01
(0.89-1.15) 10-day
Results of the duration-effect meta-analysis are shown in the bottom-left area. Results of the pairwise meta-analyses of direct comparisons are shown in the upper-right area. Data are odd ratios (95% confidence interval) of the upper-left treatment duration compared with the bottom-right treatment duration. Non-inferior results (lower bound of the 95% confidence interval higher than 0.65) are shown in light green colour.
[eAppendix10]

-Few patients had high severity scores, and critically ill patients were excluded, so the probability of success was higher. It is still unknown how short antibiotic therapy would impact the sickest patients.

RESPONSE
We completely agree with the reviewer that the results of our study may not be generalisable to the sickest patients. We added this in the limitation section.

Fifth, we did not include patients admitted to intensive care units and the results of this study may not be generalisable to those patients. [Page 25, line 385-387; DISCUSSION/Limitations]

I suggest to the authors,
-Improve the discussion, including the limitations mentioned above.

RESPONSE
As suggested, we revised the discussion. Please see the responses to the comments above and below.

- Discuss in more detail the results of the sensitivity analyses.

RESPONSE
We revised the sensitivity analyses section accordingly.
Sensitivity analyses were in line with the primary analyses. Sensitivity analyses using different locations of knots confirmed the stability of the shape of the spline curves. (eAppendix11, Figures S1) Sensitivity analyses excluding trials with an overall high risk of bias were also in agreement with the primary analyses. (eAppendix11, Figure S2.1) Sensitivity analyses excluding trials with outpatients also confirmed the main findings, suggesting the results are generalisable to inpatients, except for those admitted to the intensive care unit. (eAppendix11, Figure S2.2) Sensitivity analyses using the per protocol dataset and those including only trials that used antibiotics recommended for empirical treatment of CAP by clinical guidelines also confirmed the results. (eAppendix11, Figure S3 and S4) Exploratory sensitivity analyses showed that non-inferiority of the shorter duration was more likely to be the case in studies that randomised patients who had reached clinical stability early (eAppendix11, FigureS5)
[Page 22, line 328-339; RESULTS/Sensitivity analyses]

- Include patients admitted to ICU.

RESPONSE

We thank Dr. Ceccato for the suggestion. We agree with Dr. Ceccato that the optimal antibiotic treatment duration for those admitted to ICU is of great clinical interest. However, we have already included outpatients and we fear that including those admitted to ICU would make clinical heterogeneity even higher. Also, as suggested above, exploratory sensitivity analyses suggested that those who do not achieve clinical stability in a few days may need different treatment duration. We therefore believe that optimal treatment duration for severely ill patients like those in ICU should be evaluated separately. We acknowledge, however, that it is also a limitation. We added in the limitations section as below:

Fifth, we did not include patients admitted to intensive care units and the results of this study may not be generalisable to those patients. Optimal antibiotic duration for severely ill patients need to be further investigated. [Page 25, line 385-387; DICUSSION/Limitations]

- The figs. are hard to understand. They are not the usual figs observed in meta-analysis and require better explanations of the results. The non-inferiority margin is not present in all the figs.

RESPONSE
We appreciate the opportunity to improve the readability of our manuscript. As suggested by the reviewer, the duration-effect meta-analysis is still uncommon and the figures need more explanation. We revised the manuscript as below:
Duration-effect meta-analysis
We present the duration-effect curves in Figure 2 and Figure 3, and the tabulation of results in Table 2. The x-axis of the figures represents the treatment duration in days. The y-axis represents the odds ratio of the outcome on a logarithmic scale, just as in the forest plot of conventional pairwise meta-analysis using binary outcomes. The thin solid horizontal line represents the odds ratio = 1 and the thin dotted horizontal line in the clinical improvement figures and the all-cause mortality figure corresponds to the non-inferiority margin translated into OR. (The weighted average percentage of clinical improvement rate on day 15 in the 10-day treatment arms was 68%. The non-inferiority margin was therefore 58% and the corresponding OR was 0.65. For all-cause mortality, the numbers were 3%, 13% and OR 4.8, respectively. For clinical improvement on day 30, the numbers were 77%, 67% and OR 0.61, respectively. We did not show the non-inferiority margin in the figures for severe adverse events, because the position paper did not provide any margin for this outcome.) The thick solid line represents the dose duration-effect curve and the thick dotted lines represent its 95% CI. The 95% CI band becomes narrower when the duration range was examined by many trials or when it gets closer to the reference point. For the beneficial outcomes (clinical improvement), OR > 1 means more effective. For the harmful outcomes (all-cause mortality and serious adverse events), OR < 1 means safer.
[Page 18-19, line 285-302; RESULTS/Duration-effect meta-analysis]

We added non-inferiority margin in the all-cause mortality figure. We did not add one in the severe adverse events figure, because no non-inferiority margin was found.
[Figure2, (a) All-cause mortality]

VERSION 2 – REVIEW

REVIEWER	Furlan, Ludovico University of Milan, Department of Clinical Sciences and Community Health
REVIEW RETURNED	24-Nov-2022

GENERAL COMMENTS	The authors have performed a very accurate revision of the manuscript with extensive responses to reviewers' comments. I believe that hay assessed all major issues with adequate responses and modifications of the manuscript. As already underlined in the previous revision, I do not have adequate expertise in duration-effect-meta-analysis so I would suggest revision by an experienced statistician.

REVIEWER	Ceccato, Adrian Parc Taulí Research and Innovation Institute
REVIEW RETURNED	10-Oct-2022

GENERAL COMMENTS	The ms has been improved. Two major issues were not addressed by the authors, however, the message and the topic are of interest.

REVIEWER	Troy, Jesse Duke University, Pediatrics
REVIEW RETURNED	04-Jan-2023

GENERAL COMMENTS	This paper describes a meta-analysis of duration of antibiotic treatment and clinical improvement in patients diagnosed with community acquired pneumonia. Studies were eligible for inclusion if they were randomized trials that compared 2 or more durations of treatment with the same antibiotic. The statistical model used for the analysis was a single stage mixed effects model referred to as a “dose effects” model for aggregate data where study results can be expressed as odds ratios, relative risks, or rate ratios. In the present study the duration of antibiotic treatment was considered as the “dose” variable and the authors refer to the analysis as a duration effect meta-analysis. This methodology is established and peer-reviewed, and standard software is available for implementing it. The authors clearly explain the methodology in their paper and provide appropriate citations. In light of the meta-analysis design decisions the statistical methodology is appropriate. However, after reading the paper the justification for one of the design decisions was unclear to me. Specifically, why did the authors select a binary outcome of clinical improvement at Day 15? Given the variability in timing of outcome assessment across the included studies (which ranges from 7-45 days) it seems more natural to analyze time to clinical improvement rather than improvement at a landmark date which only a few of the included studies actually estimated. Of course, this would require a different statistical approach and so it might be helpful to add some details around why a binary endpoint was selected. For example, perhaps none of the included studies reported time-to-event analyses, or perhaps the Day 15 endpoint is common in this literature for some reason. The reason for addressing this point is as follows. From the standpoint of any of the individual included trials, using a binary endpoint at Day 15 requires complete follow-up of all randomized participants through Day 15. This implies that everyone who was randomized has confirmed clinical improvement (or not) at Day 15 and there is no loss to follow-up or death prior to Day 15. Otherwise, survival analysis methods would be required. From the standpoint of the meta-analysis, use of a binary endpoint at Day 15 requires some consistency across

	studies with respect to the window around the measurement time, i.e., all included studies should have assessed the outcome as close to Day 15 as possible. Regarding this point, the actual assessment day ranged from 7 to 45 days in the included studies. Of course, clinical improvement at Day 7 doesn't imply being alive with improvement at Day 15. Similarly, clinical improvement at Day 45 doesn't imply improvement was present 1 month earlier at Day 15. It would be nice to add some discussion of these limitations as well as pointing out for the reader how (if it all) the studies with extreme deviation from the Day 15 measurement time effected the analysis. Adding the actual measurement day for each study to Table 1 would also be helpful for the reader.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Dr. Adrian Ceccato, Parc Taulí Research and Innovation Institute, CIBERES

Comments to the Author:

The ms has been improved.

Two major issues were not addressed by the authors, however, the message and the topic are of interest.

Reviewer: 3

Competing interests of Reviewer: I have not conflict of interest

RESPONSE

We thank Dr. Adrian Ceccato for reviewing our manuscript carefully and giving us the opportunity to improve our manuscript.

Reviewer: 1

Dr. Ludovico Furlan, University of Milan, IRCCS Foundation Maggiore Policlinico Hospital

Comments to the Author:

The authors have performed a very accurate revision of the manuscript with extensive responses to reviewers' comments.

I believe that hay assessed all major issues with adequate responses and modifications of the manuscript.

As already underlined in the previous revision, I do not have adequate expertise in duration-effect-meta-analysis so I would suggest revision by an experienced statistician.

Reviewer: 1

Competing interests of Reviewer: author of article assessing similar research question

RESPONSE

We thank Dr. Ludovico Furlan for the positive comments.

Reviewer: 4

Dr. Jesse Troy, Duke University

Comments to the Author:

This paper describes a meta-analysis of duration of antibiotic treatment and clinical improvement in patients diagnosed with community acquired pneumonia. Studies were eligible for inclusion if they were randomized trials that compared 2 or more durations of treatment with the same antibiotic. The statistical model used for the analysis was a single stage mixed effects model referred to as a "dose effects" model for aggregate data where study results can be expressed as odds ratios, relative risks, or rate ratios. In the present study the duration of antibiotic treatment was considered as the "dose" variable and the authors refer to the analysis as a duration effect meta-analysis. This methodology is established and peer-reviewed, and standard software is available for implementing it. The authors clearly explain the methodology in their paper and provide appropriate citations. In light of the meta-analysis design decisions the statistical methodology is appropriate.

RESPONSE

We thank Dr. Jesse Troy for reviewing our manuscript from the statistical point of view.

However, after reading the paper the justification for one of the design decisions was unclear to me. Specifically, why did the authors select a binary outcome of clinical improvement at Day 15? Given the variability in timing of outcome assessment across the included studies (which ranges from 7-45 days) it seems more natural to analyze time to clinical improvement rather than improvement at a landmark date which only a few of the included studies actually estimated. Of course, this would require a different statistical approach and so it might be helpful to add some details around why a binary endpoint was selected. For example, perhaps none of the included studies reported time-to-event analyses, or perhaps the Day 15 endpoint is common in this literature for some reason. The reason for addressing this point is as follows.

From the standpoint of any of the individual included trials, using a binary endpoint at Day 15 requires complete follow-up of all randomized participants through Day 15. This implies that everyone who was randomized has confirmed clinical improvement (or not) at Day 15 and there is no loss to follow-up or death prior to Day 15. Otherwise, survival analysis methods would be required.

RESPONSE

We appreciate this opportunity to explain why we chose the clinical improvement at Day 15 as the primary outcome. First, as the reviewer speculated, we anticipated that reporting time-to-event analyses is uncommon in this field. In fact, none of the included trials reported time-to-event analyses, and all of them prioritised binary outcomes. Second, we chose Day 15 because the Infectious Diseases Society of America (IDSA) and the US Food and Drug Administration (FDA) proposed it as the primary date to judge the relative efficacy of different antibiotic regimens (Spellberg, et al. 2008). Last but not least, this meta-analysis aimed at assessing the non-inferiority of shorter treatment duration using the non-inferiority margin proposed in the position paper mentioned above. To use the proposed non-inferiority margin (i.e. 10%), we needed to use the binary outcome.

- Spellberg B, Talbot GH, Brass EP, et al. Position paper: recommended design features of future clinical trials of antibacterial agents for community-acquired pneumonia. Clin Infect Dis. 2008;47 Suppl 3:S249-65.

We agree that it is important to follow all the participants randomised, but unfortunately it is unrealistic for clinical trials to achieve zero loss to follow up. To challenge our findings against this potential limitation, as pre-specified in our protocol, we assumed a worst-case scenario: we used the total number of randomised participants as our denominator, and considered those who had been randomised but whose assessment was missing at follow up in the original study as drop out for any reason other than death or serious adverse events and without clinical improvement:

We will use the number of randomized patients as the denominator for intention-to-treat (ITT) dataset and we will use per-protocol (PP) dataset as defined by the original study. Those who had been randomized but not accounted for in the original study will be assumed to have dropped out for some reason other than death or serious adverse events and without clinical improvement. [eAppendix1, page 3, Protocol/Primary outcome and secondary outcomes]

To further examine this potential limitation due to missing outcome data, we additionally conducted sensitivity analyses counting missing data as clinical success. Although the confidence interval bands were relatively wider, they generally confirmed that shorter treatment duration was likely to be non-inferior to 10-day treatment.

##S5.4 To test the influence of handling missing data as not improved (counting missing data as clinically improved)

[eAppendix 12, FigureS5.4]

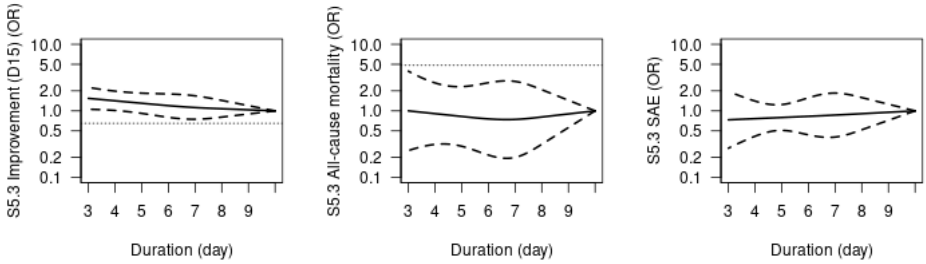
Furthermore, post-hoc sensitivity analyses which excluded trials with substantial deviation from the day 15 measurement time (eAppendix12, Figures S5.3) and those which imputed missing data as clinically improved (eAppendix12, and S5.4) also aligned with the primary analyses. [Sensitivity analyses; Page 22, line 335-338]

From the standpoint of the meta-analysis, use of a binary endpoint at Day 15 requires some consistency across studies with respect to the window around the measurement time, i.e., all included studies should have assessed the outcome as close to Day 15 as possible. Regarding this point, the actual assessment day ranged from 7 to 45 days in the included studies. Of course, clinical improvement at Day 7 doesn't imply being alive with improvement at Day 15. Similarly, clinical improvement at Day 45 doesn't imply improvement was present 1 month earlier at Day 15. It would be nice to add some discussion of these limitations as well as pointing out for the reader how (if it all) the studies with extreme deviation from the Day 15 measurement time effected the analysis. Adding the actual measurement day for each study to Table 1 would also be helpful for the reader.

RESPONSE

We appreciate this opportunity to clarify this point. Indeed, the actual measurement day ranged from 7 to 44 days. As pointed out by the reviewer, this may have introduced heterogeneity. To test the influence of trials with large deviation from the day 15 measurement time, we conducted additional sensitivity analyses excluding studies considered outliers (measurement timepoint deviated more than a week from day 15): Siegel1999 (measurement day 42-44), Stralin2014 (day 28) and Aliberti2017 (day 30). They confirmed the findings of the primary analyses.

##S5.3 To test the influence of trials with large deviation from the day 15 measurement time (excluding Siegel1999, Stralin2014, Aliberti2017. Clinical improvement on day 30 not applicable.)



[eAppendix 12, FigureS5.3]

Furthermore, post-hoc sensitivity analyses which excluded trials with substantial deviation from the day 15 measurement time (eAppendix12, Figures S5.3) and those which imputed missing data as clinically improved (eAppendix12, and S5.4) also aligned with the primary analyses. [Sensitivity analyses; Page 22, line 335-338]

As suggested, we now discuss this issue in the limitation section and revised Table 1 accordingly (we moved the risk of bias columns to the appendix to preserve the readability of Table 1).

Sixth, the actual measurement day for the primary outcome in each included study varied (7 to 44 days) and this may have introduced between-study heterogeneity. However, post-hoc sensitivity analyses excluding trials with large deviation from the day 15 measurement time were in line with the primary analyses. [DISCUSSION/Limitations; page 25, line383-386]

Study	Age, mean, y	Age, SD, y	Female, %	PSI IV+V, %	Setting	Duration, day, median	Antibiotics	No. of participants	No. of clinical improvement on day 15	Measurement day for day 15	No. of death	No. of SAE	No. of clinical improvement on day 30	Measurement day for day 30
Siegel et al, 1999	61.1	15.1	NA	NA	Inpatient	7 10	CXM	25 27	21 20	42-44	1 0	- -	21 20	42-44
Leophonte et al, 2002	64.0	18.7	25	NA	Inpatient	5 10	CRO	125 119	93 85	10	4 5	27 32	85 75	30
Tellier et al, 2004	45.8	18-87†	42	7	Both	5 7	TEL	193 195	154 157	17-21	1 2	9 5	154 157	17-21
El Moussaoui et al, 2006	57.2*	23.9*	40	12	Inpatient	3 8	AMX	57 64	50 56	10	1 0	0 0	47 49	28
File et al, 2007	45.4	16.8	42	3	Outpatient	5 7	GMI	256 256	240 234	7-9	0 1	8 14	237 221	24-30
Stralin et al, 2014	NA	NA	NA	NA	Inpatient	5 10	β-lactam	103 103.5	79 81	28	- -	- -	79 81	28
Uranga et al, 2016	65.4	18.3	37	39	Inpatient	5 10	Various	162 150	90 71	10	3 3	18 19	147 132	30
Aliberti et al, 2017	60.6*	24.8*	40	24	Inpatient	6 8	Various	125 135	111 125	30	4 1	- -	111 125	30
Dinh et al, 2021	73.2*	21.0*	41	39	Inpatient	3 8	β-lactum + placebo β-lactum + AMC	152 151	117 102	15	3 2	1 1	109 109	30

[Table 1]

Reviewer: 4

Competing interests of Reviewer: None

VERSION 3 – REVIEW

REVIEWER	Troy, Jesse Duke University, Pediatrics
REVIEW RETURNED	01-Mar-2023

GENERAL COMMENTS	Thank you for taking the time to address my comments. Congratulations on this excellent work!