

PEER REVIEW HISTORY

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 a 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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Antibiotics for uncomplicated skin abscesses: systematic review and network meta-analysis
AUTHORS	Wang, Wen; Chen, Wenwen; Liu, Yanmei; Siemieniuk, Reed; Li, Ling; Martínez, Juan Pablo Díaz; Guyatt, Gordon; Sun, Xin

VERSION 1 – REVIEW

REVIEWER	Fahimi, Jahan University of California, San Francisco I have included my name in my comments for the author
REVIEW RETURNED	06-Oct-2017

GENERAL COMMENTS	<p>Thank you for the opportunity to review this meta-analysis of the effect of antibiotics for the treatment of skin abscesses.</p> <p>I had previously published a systematic review and meta-analysis on the same topic, and in that paper we alluded to the ongoing larger trials that were not yet published. As we suspected, and this meta-analysis confirmed, the addition of high quality randomized trial data would swing the results towards showing a clinical benefit.</p> <p>This meta-analysis is very well-executed, clearly reported, and thorough. I have no concerns about the study and think it should be published. I will only offer my clinical perspective as an emergency physician that treats many such infections -- perhaps the authors may consider these comments for the discussion as they see fit:</p> <p>These cases are extremely heterogenous -- in location of infection, severity, patient comorbidities, history of MRSA infections, and patient risk factors. While the evidence seems to support benefit for even the least complicated cases, treatment success rates are high with or without antibiotics for most healthy subjects. The decision to treat really should be considered on a case-by-case basis with attention to potential adverse effects, medication interactions, and patient preference. Even with the latest evidence, I don't uniformly prescribe antibiotics, but have certainly lowered my threshold for doing so.</p> <p>Again, I applaud the authors for a really well done analysis.</p>
-------------------------	--

REVIEWER	Spotila, Jennifer patient advocate I have included my name in my comments for the author
REVIEW RETURNED	06-Oct-2017

GENERAL COMMENTS	<p>Given how common skin infections are, this paper provides important information about the effectiveness of antibiotics as an adjuvant to surgical incision and drainage, and the impact of different antibiotic options. These data could be helpful to patients in making choices about treatment options for skin infections.</p> <p>Because this paper is a systematic review, its conclusions are inherently limited by the data from the included studies. There were only 14 randomised controlled trials (3,541 subjects) that met inclusion criteria for this review. The authors note that four were published more than 30 years ago. The authors note that treatment and antibiotic resistance has evolved substantially since then, but they do not discuss the extent to which this limited the conclusions that can be drawn from the analysis.</p> <p>This review did incorporate patient participation in identifying outcomes important for decision-making. However, limitations in the data from the included studies were not sufficiently discussed to enable patients and their doctors to make treatment decisions. This is particularly true regarding the increased risk of gastrointestinal side effects with antibiotic usage. There was no characterization of these side effects beyond the presence or absence of symptoms, such as diarrhea. No data were reported regarding the severity of these side effects or how long they lasted. There is also no discussion of the risk of developing <i>C. difficile</i>, especially after treatment with clindamycin.</p> <p>In making a decision about antibiotic therapy as an adjuvant to surgical incision and drainage, patients will consider not only the likelihood of side effects but the likely severity of those side effects. Patients also need to consider whether the antibiotic therapy puts them at risk of complications such as <i>C. difficile</i>. Risk/benefit analysis in selecting a treatment is more accurate when patients have this type of information. If the included studies do not characterize those data, then the authors may want to consider discussing this in the Strengths and Limitations section.</p>
-------------------------	--

REVIEWER	Spineli, Loukia Medizinische Hochschule Hannover, Institut für Biometrie
REVIEW RETURNED	08-Oct-2017

GENERAL COMMENTS	<p>Thank you for giving me the opportunity to conduct this review. I read this work with great interest. This is a systematic review of 14 randomized controlled trials comparing various antibiotics with each other or no antibiotic in children or adult patients with uncomplicated skin abscesses. A series of meta-analyses on specific patient-relevant outcomes indicated that patients with uncomplicated skin abscesses may benefit from prescription of antibiotics but may experience adverse events, such as diarrhea and nausea, among others. Using a network meta-analysis for the outcome 'treatment failure', the authors found that TMP-SMX is more likely to lead to abscess recurrence compared to clindamycin, whereas it is less likely to cause diarrhea. Using the GRADE tool, the authors inferred that the collated evidence is of moderate to high quality.</p> <p>Having patients and patient representatives to be involved during the set-up of the review comprises a great strength of the present work as well as having pre-specified important components for the moderator analyses.</p>
-------------------------	---

	<p>However, it is not clear to me whether the primary analysis aim of the authors was actually network meta-analysis but eventually they decided to lump the antibiotics and compare with no antibiotic in all pre-specified outcomes, since network meta-analysis was possible only for the outcome 'treatment failure'. This needs to be clarified already in the Methods or at least explicitly discussed in the Discussion section. Further in network meta-analysis, the authors could provide the readers with rankograms and SUCRA plots not necessarily with the aim to determine a formal hierarchy of the studied interventions but mostly to investigate the potency of the interventions in each rank (and the uncertainty around it) using the available collated evidence for treatment failure. My comments below focus mainly on the statistical analysis strategy.</p> <p>Methods (line 12, p. 7) Since a network meta-analysis has been also implemented, the authors should cite the NMA-PRISMA which is an extension of the conventional PRISMA.</p> <p>Literature search The authors provide a comprehensive description of the literature search; however, nothing is mentioned on any attempts made to retrieve relevant unpublished material.</p> <p>Data analysis - Meta-analyses 1. Clinical and methodological heterogeneity may manifest to statistical heterogeneity to a greater or lesser extent. Testing for statistical heterogeneity is deemed actually redundant. Measuring the extent of and investigating plausible sources of statistical heterogeneity shall be opted, instead. The latter has been carefully planned but the former has not been properly attempted. 2. The authors used the I² statistic, Cochran's Q-statistic and (my guess is) the DerSimonian and Laird (DSL) heterogeneity estimator (which is not mentioned in the Methods neither in the Results – apart from the forest plots). An important limitation to consider for the Q-statistic and by extent, for the I² statistic and the DSL (both being functions of the Q-statistic) is that they require the studies to be sufficiently large in order to approximate normality for the study-specific treatment effects as well as being similarly sized [1]. This is not the case in your review since the majority of the studies have serious size limitations. As a result, DSL cannot properly estimate the heterogeneity variance (DSL will underestimate the heterogeneity variance, instead) and I² will be considerably misleading [1, 2] (Limitation 1). For example, in the analysis 16.1.1 (Figure 2), there is obvious statistical heterogeneity in the results. However, both τ^2 and I² are substantially low and Q-statistic gives evidence against heterogeneity. This should be expected since the majority of the studies are small and this is reflected in the particularly wide and overlapping confidence intervals. If you increase the size of the studies by 10, both τ^2 and I² (and Q-statistic) will increase. Lastly, the uncertainty in estimating the heterogeneity parameter is not provided (Limitation 2) and neither is it accounted in the confidence interval of the meta-analysis odds ratio (Limitation 3) and therefore, the validity of the inference can be compromised (using the Hartung-Knapp modification could nicely tackle this limitation as this method has been advocated for sustaining the nominal type I error). The authors chose to pool the results using Mantel-Haenszel weighting which is an advantageous approach in the light of low events which is the case in the current review for the majority of the studies.</p>
--	---

	<p>Unfortunately, RevMan does not offer the opportunity to select among a comprehensive set of heterogeneity estimators, neither does it offer the Hartung-Knapp modification nor does it estimate the uncertainty around the heterogeneity estimator and the I² statistic. Therefore, the results of these meta-analyses should be interpreted with great caution and this should be explicitly mentioned in the Discussion section (Recommendation 1). Alternatively, the authors can run the meta-analyses in R using the ‘metafor’ package (Recommendation 2). Then they may use the restricted maximum likelihood (REML) estimator alongside the Hartung-Knapp modification. In that way, they can provide the readers also with confidence intervals both for the heterogeneity variance and the I² statistic. A better alternative strategy would have been to model the exact distribution of the data (binomial – no correction for zero cells is needed) in a Bayesian framework using content-relevant published predictive distributions as a prior for the heterogeneity estimator [3]. In that way, (i) heterogeneity variance can be properly estimated and (ii) the uncertainty in estimating that parameter is inherently accounted for leading to more reliable inferences (Recommendation 3).</p> <p>Data analysis – Network meta-analyses</p> <p>1. The information provided for the set-up of the network meta-analysis is not sufficient enough. The authors do not mention the exact prior distributions they used for the location and dispersion parameters. Since only half the possible pairwise comparisons are informed directly and the direct comparisons are informed only by 2 to 4 trials, I will recommend to opt for a weakly informative prior for the heterogeneity parameter in the primary analysis and a less informative prior as a sensitivity analysis. The latter is necessary in order to investigate the extent to which the heterogeneity variance is sensible to prior specifications. Debray et al [4] offer a comprehensive table (Table 4, p. 303) with the prior specifications on the heterogeneity parameter that have been used or proposed. For instance, you could select HN(0, 1) I(0,) for the primary analysis and Unif(0, 5) for the sensitivity analysis.</p> <p>2. Furthermore, the authors do not report what assumption they used for the heterogeneity variance. A common-heterogeneity seems to be a feasible assumption since you have only 12 trials available to estimate all mixed effect plus 5 comparison-specific heterogeneity parameters (in case you had opted for that assumption) and hence, your model might not achieve a successful convergence. While this assumption facilitates model convergence and estimation of heterogeneity, it can be suboptimal when heterogeneity varies substantially among the comparisons [5]. Therefore, a series of random-effect meta-analysis on each direct comparison is needed in order to judge the applicability of the common-heterogeneity assumption.</p> <p>Results – Effects of antibiotics versus no antibiotics</p> <p>I would recommend to present the point estimate and 95% confidence interval of the τ^2 instead of I² as the former actually captures the extent of heterogeneity measured in the same scale with the meta-analysis treatment effect.</p> <p>Results – Comparative effects on treatment failure</p> <p>Alike the above comment, I would recommend to present the point estimate and 95% credible interval of the τ^2. If the common-heterogeneity assumption has been used, then τ^2 will be the same under ‘pooled (network)’.</p>
--	--

	<p>Avoid presenting I² without an uncertainty measure as it might be misleading. For example, an estimated I² at 17% that extends from 0% to 50% cannot be claimed to indicate low heterogeneity. Lastly, by supplementing the forest plot (Figure 8) with the predictive intervals of the mixed effects, you can visualize in addition the implications of heterogeneity on the mixed effects.</p> <p>Other comments</p> <p>Please drop the 'overall' diamond at the end of each forest plot.</p> <p>Network plot: Since 'no antibiotic' versus 'Clindamycin' has 3 trials, whereas 'Clindamycin' versus 'TMP-SMX' has only 2 trials, the former should have thicker line than the latter.</p> <p>Line 18, p. 15. It is 6 trials, not 7. Or is it 5 trials according to Figure 8?</p> <p>Regarding the intention-to-treat analysis, I guess you assumed treatment failure for the missing participants?</p> <p>References</p> <ol style="list-style-type: none"> 1. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. <i>Br J Math Stat Psychol</i>. 2007 May;60(Pt 1):29-60. 2. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I² in assessing heterogeneity may mislead. <i>BMC Med Res Methodol</i>. 2008;8:79. doi: 10.1186/1471-2288-8-79. 3. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. <i>Stat Med</i>. 2015;34(6):984-98. doi: 10.1002/sim.6381. 4. Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. <i>Res Synth Methods</i>. 2015 Dec;6(4):293-309. doi: 10.1002/jrsm.1160. 5. Thorlund K, Thabane L, Mills EJ. Modelling heterogeneity variances in multiple treatment comparison meta-analysis--are informative priors the better solution? <i>BMC Med Res Methodol</i>. 2013 Jan 11;13:2. doi: 10.1186/1471-2288-13-2.
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

1. This meta-analysis is very well-executed, clearly reported, and thorough. I have no concerns about the study and think it should be published. I will only offer my clinical perspective as an emergency physician that treats many such infections -- perhaps the authors may consider these comments for the discussion as they see fit: These cases are extremely heterogenous -- in location of infection, severity, patient comorbidities, history of MRSA infections, and patient risk factors. While the evidence seems to support benefit for even the least complicated cases, treatment success rates are high with or without antibiotics for most healthy subjects. The decision to treat really should be considered on a case-by-case basis with attention to potential adverse effects, medication interactions, and patient preference. Even with the latest evidence, I don't uniformly prescribe antibiotics, but have certainly lowered my threshold for doing so.

[Response]

Thank you so much for your appreciation and wonderful suggestion. We fully agree with you that the decision to use antibiotics or not should be made on a case-by-case basis – especially in light of the similar magnitude of risk for adverse events.

As suggested, we have revised the discussion according to the comments, please see our response to the comment # 3 from the editors.

Reviewer: 2

1. Because this paper is a systematic review, its conclusions are inherently limited by the data from the included studies. There were only 14 randomised controlled trials (3,541 subjects) that met inclusion criteria for this review. The authors note that four were published more than 30 years ago. The authors note that treatment and antibiotic resistance has evolved substantially since then, but they do not discuss the extent to which this limited the conclusions that can be drawn from the analysis.

[Response] The results of treatment failure within 1 month (OR 0.56, 95% CI 0.34 to 0.93) and recurrence within 1 month (OR 0.45, 95% CI 0.27 to 0.74) were similar when omitting studies published more than 30 years ago. We discussed this in our manuscript as follows: "Four of the RCTs were published more than 30 years ago and surgical treatments as well as antibiotic resistance patterns have changed. The results and interpretation did not change when these trials were excluded from the analyses."

2. This review did incorporate patient participation in identifying outcomes important for decision-making. However, limitations in the data from the included studies were not sufficiently discussed to enable patients and their doctors to make treatment decisions. This is particularly true regarding the increased risk of gastrointestinal side effects with antibiotic usage. There was no characterization of these side effects beyond the presence or absence of symptoms, such as diarrhea. No data were reported regarding the severity of these side effects or how long they lasted. There is also no discussion of the risk of developing *C. difficile*, especially after treatment with clindamycin.

[Response]

Two large trials that included clindamycin (n = 2028) monitored for but did not find any cases of *Clostridium difficile* infection. Unfortunately, no information was provided about the severity of antibiotic/clindamycin-associated diarrhoea, but in our experience, the severity can vary between a being a mild nuisance to severe. Rare adverse events and long-term adverse effects are unlikely to be observed in RCTs; other rare but important adverse effects include anaphylaxis and toxic epidermal necrolysis (especially with TMP-SMX). The rate of developing Stevens-Johnson syndrome/ toxic epidermal necrolysis due to TMP/SMX is approximately 3-5/100,000 exposed patients (Arch Dermatol.1990 Jan;125(1):43-47; N Engl J Med.1995 Dec ;333(24):1600-1607) Chan 1990, Roujeau 1995). We have added the result of *C. difficile* infection in our manuscript as follows: Two large trials (n=2051) monitored for *C. difficile* infection (CDI) with routine clinical monitoring: no CDI occurred in any treatment arm. We have also added the safety information to the limitation of the discussion section, as below: "Moreover, rare adverse events are unlikely to be observed in RCTs. Important but rare adverse events include anaphylaxis, *C. difficile* infection (especially with clindamycin³⁸), and Stevens-Johnson syndrome or toxic epidermal necrolysis(especially with TMP SMX³⁹)."

3. In making a decision about antibiotic therapy as an adjuvant to surgical incision and drainage, patients will consider not only the likelihood of side effects but the likely severity of those side effects. Patients also need to consider whether the antibiotic therapy puts them at risk of complications such as *C. difficile*. Risk/benefit analysis in selecting a treatment is more accurate when patients have this type of information. If the included studies do not characterize those data, then the authors may want to consider discussing this in the Strengths and Limitations section.

[Response]

Thanks for the comments. As suggested, we discussed this in the strengths and limitations section and conclusion, please see our response to the comment # 2 and below:

The decision whether or not to use antibiotics should take into account local MRSA resistance patterns, individual patient clinical factors (e.g. severity of infection, immunocompromised state), and individual values and preferences (e.g. a strong desire to avoid diarrhoea).

Reviewer: 3

1. Having patients and patient representatives to be involved during the set-up of the review comprises a great strength of the present work as well as having pre-specified important components for the moderator analyses. However, it is not clear to me whether the primary analysis aim of the authors was actually network meta-analysis but eventually they decided to lump the antibiotics and compare with no antibiotic in all pre-specified outcomes, since network meta-analysis was possible only for the outcome 'treatment failure'. This needs to be clarified already in the Methods or at least explicitly discussed in the Discussion section. Further in network meta-analysis, the authors could provide the readers with rankograms and SUCRA plots not necessarily with the aim to determine a formal hierarchy of the studied interventions but mostly to investigate the potency of the interventions in each rank (and the uncertainty around it) using the available collated evidence for treatment failure. [Response]

Thank you for the detailed review and suggestions. We set out to address two important clinical questions: 1) in patients with uncomplicated skin abscesses, what is the impact of antibiotic therapy plus incision and drainage (I&D) compared to I&D alone; 2) what are the impacts of different antibiotic options. For the first clinical question, we prespecified a pairwise meta-analysis approach. For the second question, we prespecified a network meta-analysis (NMA) of RCTs to compare effects of alternative antibiotics. The pairwise meta-analyses and the exploration of a few prespecified subgroup analyses was the primary approach to address the first question. We have clarified in Methods section, see follows: For our primary comparison of antibiotics vs. no antibiotics, we conducted pairwise meta-analyses.

We appreciate the possibility that network meta-analyses can provide rankograms and SUCRA plots. We chose not to conduct or present these analyses because they do not incorporate certainty of evidence and can therefore be misleading. For example, a study at high risk of bias showing a large effect for an intervention would misleadingly increase the rank of the antibiotic. Further, we think that they would add unnecessary length to what is already a long paper. 2. Since a network meta-analysis has been also implemented, the authors should cite the NMA-PRISMA which is an extension of the conventional PRISMA.

[Response]

Thank you for the comments. As suggested, we have cited the NMA-PRISMA. We have tracked the key elements in reports of network meta-analyses according to the PRISMA network analysis checklist.

3. The authors provide a comprehensive description of the literature search; however, nothing is mentioned on any attempts made to retrieve relevant unpublished material.

[Response] According to your comment, we additionally searched ClinicalTrials.gov to identify additional eligible studies. The search yield 245 potentially relevant reports, however, none met inclusion criteria. We updated the information of our manuscript, please see figure 1, appendix1 and below:

We searched Medline, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) from inception to 17 August 2017 to identify relevant studies, without language restrictions. We combined database-specific subject headings (such as MeSH terms) and free-text terms regarding "skin abscess" and "anti-infective agents" to search for potentially eligible studies. Supplementary Appendix 1 presents the full search strategy. We also searched ClinicalTrials.gov to identify any unpublished studies and reviewed the reference lists of the included RCTs.

4. Clinical and methodological heterogeneity may manifest to statistical heterogeneity to a greater or lesser extent. Testing for statistical heterogeneity is deemed actually redundant. Measuring the extent of and investigating plausible sources of statistical heterogeneity shall be opted, instead. The latter has been carefully planned but the former has not been properly attempted.

[Response]

To explore sources of inconsistency, we pre-specified five subgroup hypotheses. The guideline panel proposed these subgroup hypotheses according to literature review and clinical experience. We further explored whether the effects differed between classes of antibiotics with network meta-analysis.

5. The authors used the I² statistic, Cochran's Q-statistic and (my guess is) the DerSimonian and Laird (DSL) heterogeneity estimator (which is not mentioned in the Methods neither in the Results – apart from the forest plots). An important limitation to consider for the Q-statistic and by extent, for the I² statistic and the DSL (both being functions of the Q-statistic) is that they require the studies to be sufficiently large in order to approximate normality for the study-specific treatment effects as well as being similarly sized [1]. This is not the case in your review since the majority of the studies have serious size limitations.

As a result, DSL cannot properly estimate the heterogeneity variance (DSL will underestimate the heterogeneity variance, instead) and I² will be considerably misleading [1, 2] (Limitation 1). For example, in the analysis 16.1.1 (Figure 2), there is obvious statistical heterogeneity in the results. However, both τ^2 and I² are substantially low and Q-statistic gives evidence against heterogeneity. This should be expected since the majority of the studies are small and this is reflected in the particularly wide and overlapping confidence intervals. If you increase the size of the studies by 10, both τ^2 and I² (and Q-statistic) will increase.

Lastly, the uncertainty in estimating the heterogeneity parameter is not provided (Limitation 2) and neither is it accounted in the confidence interval of the meta-analysis odds ratio (Limitation 3) and therefore, the validity of the inference can be compromised (using the Hartung-Knapp modification could nicely tackle this limitation as this method has been advocated for sustaining the nominal type I error). The authors chose to pool the results using Mantel-Haenszel weighting which is an advantageous approach in the light of low events which is the case in the current review for the majority of the studies.

Unfortunately, RevMan does not offer the opportunity to select among a comprehensive set of heterogeneity estimators, neither does it offer the Hartung-Knapp modification nor does it estimate the uncertainty around the heterogeneity estimator and the I² statistic. Therefore, the results of these meta-analyses should be interpreted with great caution and this should be explicitly mentioned in the Discussion section (Recommendation 1). Alternatively, the authors can run the meta-analyses in R using the 'metafor' package (Recommendation 2). Then they may use the restricted maximum likelihood (REML) estimator alongside the Hartung-Knapp modification. In that way, they can provide the readers also with confidence intervals both for the heterogeneity variance and the I² statistic. A better alternative strategy would have been to model the exact distribution of the data (binomial – no correction for zero cells is needed) in a Bayesian framework using content-relevant published predictive distributions as a prior for the heterogeneity estimator [3]. In that way, (i) heterogeneity variance can be properly estimated and (ii) the uncertainty in estimating that parameter is inherently accounted for leading to more reliable inferences (Recommendation 3).

[Response] We agree that there are limitations with the DerSimonian and Laird (DL) method, which does not consider uncertainty in the heterogeneity estimator. Based on some early modeling and small retrospective analyses (BMC Med Res Methodol. 2014 Feb 18;14:25.), the Hartung-Knapp-Sidik-Jonkman (HKSJ) method may decrease the type 1 error rate, however this is irrelevant in this context because avoid a hypothesis testing/p-value approach. The HKSJ approach has not been prospectively evaluated and we believe that there is not enough experience with it to advocate for its routine use. Indeed, we have found serious flaws with the approach when attempting to use it. For example, the HKSJ method can sometimes lead to implausibly narrow confidence intervals (Ann Intern Med. 2016 May 3;164(9):636-7). It might also lead to inappropriately wide confidence intervals and by how much the HKSJ approach, compared to the DL method, increases the risk of type 2 errors is uncertain. Despite our serious reservations, we have conducted a sensitivity analysis using HKSJ method at the reviewer's suggestion; please see as below and table F in Appendix 3. We conducted the following sensitivity analyses to examine the robustness of effect estimates: analyses using alternative effect measures (odds ratio versus relative risk), statistical models (fixed versus random effects), pooling methods (Peto versus M-H), alternative methods for random effects meta-analysis (DerSimonian and Laird [DL] versus Hartung-Knapp-Sidik-Jonkman [HKSJ]), and alternative assumptions about missing data, as well as analyses omitting trials published before 1990 and trials with patients treated by primary suture rather than open drainage and, for treatment failure, excluding trials that considered recurrences as treatment failure.

The confidence intervals for abscess treatment failure, late recurrence, hospitalization, gastrointestinal side effects and nausea excluded no effect with the DL method but not the HKSJ method (tables F in appendix 3).

6. The information provided for the set-up of the network meta-analysis is not sufficient enough. The authors do not mention the exact prior distributions they used for the location and dispersion parameters. Since only half the possible pairwise comparisons are informed directly and the direct comparisons are informed only by 2 to 4 trials, I will recommend to opt for a weakly informative prior for the heterogeneity parameter in the primary analysis and a less informative prior as a sensitivity analysis.

The latter is necessary in order to investigate the extent to which the heterogeneity variance is sensible to prior specifications. Debray et al [4] offer a comprehensive table (Table 4, p. 303) with the prior specifications on the heterogeneity parameter that have been used or proposed. For instance, you could select $HN(0, 1)I(0,)$ for the primary analysis and $Unif(0, 5)$ for the sensitivity analysis.

[Response] We prespecified that we would use uninformative priors for the primary analysis, which approximate frequentist confidence intervals and thus allow direct comparisons with the primary pairwise meta-analysis. As suggested, we have conducted a sensitivity analysis using weakly informative priors.

We have added the following to the methods: The primary network meta-analysis was conducted with uninformative priors with a uniform distribution, $Unif(0, 5)$. We also conducted a sensitivity analysis with weakly informative priors ($HN(0, 1)I(0,)$). In the results, we have included the following: The results and interpretation of the network meta-analysis did not change when we used weakly informative priors instead of than uninformative priors (data not shown). The results of the sensitivity analysis are included below, for completeness of documentation but are not included in the paper to avoid confusion.

Table. Sensitivity analysis using weakly informative priors for treatment failure within 1 month.

	No	Early	Late	TMP-SM	Clindamyc
	antibiotics	cephalosporin	cephalosporin	X	in
No					
No antibiotics					
antibiotics					
Early	52 (-33, 227)	Early cephalosporin			
cephalosporin					
Late	32 (-54, 250)	-19 (-106, 99)	Late cephalosporin		
cephalosporin					
	-34 (-50, -12)	-86 (-261, 2)	-66 (-283, 23)	TMP-SM	
TMP-SMX				X	
	-39 (-57, -11)	-91 (-266,-1)	-71 (-288, 20)	-5 (-27, 21)	Clindamyc in
Clindamycin					

Each number is a risk difference and 95% credible interval. The rows are the reference category: a risk difference <0 favours the row. Green shading = high certainty; orange shading = moderate certainty; red shading = low certainty. Based on the median treatment failure rate in the no antibiotics arms, we assume that the baseline risk of treatment failure without antibiotics is 9.0%.

7. Furthermore, the authors do not report what assumption they used for the heterogeneity variance. A common-heterogeneity seems to be a feasible assumption since you have only 12 trials available to estimate all mixed effect plus 5 comparison-specific heterogeneity parameters (in case you had opted for that assumption) and hence, your model might not achieve a successful convergence. While this assumption facilitates model convergence and estimation of heterogeneity, it can be suboptimal when heterogeneity varies substantially among the comparisons [5]. Therefore, a series of random-effect meta-analysis on each direct comparison is needed in order to judge the applicability of the common-heterogeneity assumption.

[Response]

Thank you. We assumed a common heterogeneity parameter and now include that in the methods: We assumed common heterogeneity within the network.

We carefully assessed inconsistency within each of the direct estimates within the network. To do so, we calculated the I² for heterogeneity for each pairwise comparison (see figure 8). Then, using the GRADE approach, we considered whether or not we should rate down the certainty in the estimates for inconsistency (see Appendix Table B).

8. I would recommend to present the point estimate and 95% confidence interval of the τ^2 instead of I² as the former actually captures the extent of heterogeneity

measured in the same scale with the meta-analysis treatment effect.

[Response]

Thank you for the great statistical input. We have chosen to present the I² rather than Tau² because I² is more familiar to a general audience, including the clinicians to whom this paper is targeted. This is consistent with Cochrane recommendations (see Cochrane Handbook chapter 9.5.2 Identifying and measuring heterogeneity).

9. Alike the above comment, I would recommend to present the point estimate and 95% credible interval of the τ^2 . If the common-heterogeneity assumption has been used, then τ^2 will be the same under 'pooled (network)'. Avoid presenting I² without an uncertainty measure as it might be misleading. For example, an estimated I² at 17% that extends from 0% to 50% cannot be claimed to indicate low heterogeneity. Lastly, by supplementing the forest plot (Figure 8) with the predictive intervals of the mixed effects, you can visualize in addition the implications of heterogeneity on the mixed effects.

[Response]

As stated above, we have carefully considered inconsistency in our analysis. While the global heterogeneity of a network can be informative, we agree that it can be misleading, regardless of the measurement unit used, when considering the certainty of the evidence at a pairwise level. For that reason, we take the GRADE approach, in which we evaluate heterogeneity for each pairwise comparison using both statistical tests as well as careful visual observation of the forest plots. We also consider clinical and methodological inconsistency.

10. Please drop the 'overall' diamond at the end of each forest plot.

[Response]

As suggested, we drop the 'overall' diamond at the end of forest plot, in the cases where subgroup effects are most trustworthy.

11. Network plot: Since 'no antibiotic' versus 'Clindamycin' has 3 trials, whereas 'Clindamycin' versus 'TMP-SMX' has only 2 trials, the former should have thicker line than the latter.

[Response]

Great catch. The initial network counted the 2 RCTs that were reported in a single paper as 1. We have updated the network plot.

12. Line 18, p. 15. It is 6 trials, not 7. Or is it 5 trials according to Figure 8?

[Response]

It is 7 trials. As we mentioned in our manuscript "One report included two

independent RCTs, and the other reported results of a factorial trial that also compared two surgical approaches and reported results separately for each approach".

13. Regarding the intention-to-treat analysis, I guess you assumed treatment failure for the missing participants?

[Response]

Our primary approach is the complete case analysis, in order to minimize assumptions about outcomes from the missing participants, as suggested by GRADE guidance (J Clin Epidemiol. 2017 Jul;87:14-22). When the results suggest a treatment effect, we now conduct sensitivity meta-analyses using worst plausible analysis to inform GRADE certainty in evidence assessments (Table E in appendix 3). For outcomes that are not robust to the worst plausible analysis, we rated down our certainty in the evidence for risk of bias. We performed analyses using four additional assumptions about the missing data for the outcomes of treatment failure and recurrence within 1 month: 1) none of the participants lost to follow-up had the event; 2) all the participants lost to follow-up had the event; 3) none of those lost to follow-up in the treatment group had the event and all those lost to follow-up in the control group did (best case scenario); 4) all participants lost to follow-up in the treatment group had the event and none of those in the control group did (worst case scenario),

Reviewer 4

1. These studies cannot be pooled quantitatively since the definitions of “clinical failure” and clinical cure differ markedly from trial to trial. In the trial by Talan the endpoint definition changes from one time point to another (early vs later) in the same trials making it challenging if not impossible to evaluate what is being affected by the interventions. A unit-less effect size called “clinical failure” is meaningless since it pools different outcomes of different importance together. One of the criticisms of meta-analysis is the combining of disparate studies and that is the case here, where not only were there different outcomes but there were different co-interventions such a type of packing etc., which have been shown in prior trials to influence outcomes independent of antibiotics. In addition, current trials use a composite of worsening of the original lesion and development of new lesions. Few patients have worsening of the original lesion and most the outcomes are in the “new infection” category. This again is lost in the composite endpoint and the pooling of the evidence.

[Response]

We appreciate your comments, which helped us improve the manuscript. We agree with you that the definition of treatment failure varied among included trials. To make this abundantly clear and to allow readers to assess this outcome definition for themselves, we now include a table that includes the definitions for treatment failure that the trials used. Although the trials used slightly different definitions for treatment failure, all measured the same thing – whether treatment of the abscess is cured or not. Because the diagnosis of a skin abscess is based on clinical findings, it is thus subject to subjectivity. Therefore, there we appreciated the trialists’ use of formal criteria for treatment failure. Almost all trials defined treatment failure as a lack of resolution of signs or symptoms of infection. We think the heterogeneity of definition between trials is acceptable. This is further corroborated by low or very low statistical measures of heterogeneity. We also fully agree that treatment failure being a composite outcome causes challenges in interpretation – especially because it is difficult to place importance on the outcome. A limitation of composite outcomes is that we are unsure of the underlying distribution of outcomes that inform the composite (e.g. number who failed because of increasing pain/erythema vs. hospitalization). While we do believe that the analyses involving treatment failure provide useful information about the magnitude of expected benefit, we also provide detailed analyses of other key non-composite endpoints, including abscess recurrence at 1 month and at 1-3 months.

Overall, we certainly agree with you that the slightly different definitions for treatment failure is a limitation, but for that outcome only. We have listed details of the outcomes definitions in table C of appendix 2 and discussed in the strengths and limitations section. It read as:

In addition, the definition of outcomes varied among included trials.

2. It is not clear how having a single patient liaison addresses the issues of what is important to patients validates endpoint selection. The composite endpoints used in these trials focus on clinician reported outcomes that are not direct measures of benefit to patients, and the use of composites makes it challenging if not impossible to infer effects on direct patient benefit (e.g. decreased pain and other symptoms, improved function in patients' daily lives). A recent study of qualitative interviews of patients with skin abscess across a broad range of patients showed that the current outcome measure miss many of the outcomes that patients consider important (ISPOR 2016).

[Response]

The guideline panel included three people with lived experience of skin abscesses (two adult patients and one parent of a child patient). They helped to identify the outcomes most important to patients and with interpreting the results based on what they expected the typical patient values and preferences to be, as well as the variation between patients. We also counted on the experiences of our clinician panel members and co-authors, who have extensive experience interacting with a variety of patients with skin abscesses.

Treatment failure was identified as being important to patients, but other outcomes (e.g. pain, hospitalization) are probably even more important to most people. For completeness, we present all of the patient-important outcomes.

3. The data on outcomes that are serious and clearly important to patients is present in too few studies to draw conclusions. One cannot state that hospitalization is decreased since it was only evaluated in two trials and patients may be hospitalized merely due to the size of the abscess at baseline and the need for inpatient drainage. This was the case in the observational study by Lee where all children with abscess >5cm were admitted at baseline and this has led to the erroneous notion that abscesses >5cm are somehow "more severe" yet these patients did not have worse outcomes in that same study. There is no difference in deaths and sepsis occurred at similar frequency therefore there is no evidence that antibiotics decrease these serious events, yet the paper by Daum in the NEJM begins by talking about the rare but serious consequences of

[Response]

We fully agree with the author that a limitation of the evidence is that not all trials reported all of the patient important outcomes. This is an unfortunate reality when summarizing evidence from RCTs in every area of medicine. Using the GRADE approach, we transparently assess certainty in the evidence for each outcome. For some outcomes such as sepsis or hospitalizations, there was limited evidence from a low number of events because few studies reported this outcome. Thus we rated down

certainty in the evidence for due to serious imprecision around the absolute effect. In addition, we have added the safety information to the discussion of the manuscript, please see our response to the comment # 2 from reviewer 2.

4. The older studies noted as not “underpowered” since all current trials in skin abscesses that are non-inferiority trials use a non-inferiority margin of >10%. This means that effect sizes on “failure” (no matter how defined) of less than 10% should be considered as clinically inconsequential. If this thinking applies in NI trials then it should apply in superiority trials as well in that any difference less than 10% is not clinically meaningful, meaning the confidence intervals around the point estimate of the treatment difference should exclude 10% (this would be needed in any case to justify current NI trials). For example, the trial by Schmitz was powered for a 15% difference which would be needed to rule out a minimal difference of 10%. Therefore, this trial is not “underpowered to detect a clinically meaningful difference. Detecting any difference greater than zero is not clinically meaningful in this disease. The current trial by Talan were powered to detect a difference of at least 7.5% for the LOWER BOUND of the confidence interval and this is specified in the protocol that is an appendix in the NEJM and was pointed out by the Data Safety Monitoring Board for this trial. Therefore, these trials fail to exclude this magnitude of difference yet were interpreted a “positive” trials since the range of the confidence interval excluded zero, implying that any difference greater than zero is clinically meaningful when these trials were a) not designed in this way and b) current NI trials used 9 to approve every antibiotic for skin infections (where trials include up to a third of patients with skin abscesses) use a 10% NI margin. This discussion is missing from this meta-analysis, which only addresses the irrelevant question of whether “antibiotics” are better than nothing [Response]

Many of the trials were, as the reviewer points out, non-inferiority trials. As far as we can tell, the non-inferiority thresholds used were chosen arbitrarily and without patient input, and, as the reviewer points out, were inconsistent. Instead of using arbitrary cutoffs for ‘importance’ and using hypothesis testing to evaluate such cutoffs, we take the approach of quantifying the magnitude of effect, and its uncertainty therein for, for each outcome. This allows each patient to decide for themselves whether or not the desirable consequences outweigh the undesirable consequences. We agree with the reviewer that the desirable effects of antibiotics are modest and closely balanced with undesirable consequences; we highlight this in both the abstract and discussion.

5. The notion that all antibiotics are the same is also incorrect so pooling the data provides misleading evidence. The recent trial by Daum et al showed differences in later effects with clindamycin noted in this meta-analysis but diluted by the pooling of effects.

[Response]

We conducted subgroup analyses to explore whether antibiotics with MRSA-activity performed better than those without, and whether the effect of TMP-SMX differed from clindamycin. Considering individual antibiotics may have different effects, we further conducted a network meta-analysis of RCTs to compare effects of alternative antibiotics. We summarized the GRADE evidence profile of alternative antibiotics in table 4-6. The Daum study which suggested that compared to TMP-SMX, clindamycin probably results in lower risk of recurrence at 1 month have been summarized in table 4, however we have low certainty that this difference is true because of imprecision (the confidence interval includes no effect) and because the study was inconsistent with indirect evidence from other trials that compared TMP-SMX to placebo and showed TMP-SMX was in fact effective at preventing recurrence.

For adverse events, we agree that the incidence and severity of adverse events is almost certainly different with different antibiotics. We now report the safety outcomes separately for each antibiotic (clindamycin and TMP-SMX) in table 3.

The incidence and severity of adverse events is likely to differ between antibiotics,

thus we analysed the safety outcomes separately for each antibiotic (clindamycin and TMP-SMX). Both TMP-SMX (four trials, 9, 10, 24, 26 OR 1.28, 95% CI 1.04 to 1.58, I²=0%;

21 more (3 more to 43 more) per 1000 patients; moderate quality) and clindamycin (one trial, 9 OR 2.29, 95% CI 1.35 to 3.88; 95 more (28 more to 187 more) per 1000 patients; moderate quality) were associated with increased risk of overall gastrointestinal side effects. Clindamycin increases the risk of diarrhoea (one trial, 9 OR 2.71, 95% CI 1.50 to 4.89; 96 more (30 more to 193 more) per 1000 patients; high quality), while TMP-SMX probably does not (three trials, 9, 10, 26 OR 0.92, 95% CI 0.70 to 1.22, I²=0%; moderate quality) (Table 3). Two large trials^{9,10} (n=2051) monitored for *C. difficile* infection (CDI) with routine clinical monitoring: no CDI occurred in any treatment arm. TMP-SMX probably increases the risk of nausea (TMP-SMX OR 1.49, 95% CI 0.98 to 2.25, I²=11%; moderate quality), while clindamycin may not (OR 0.96, 95% CI 0.31 to 3.02; moderate quality). TMP-SMX

does not appear to have an important effect on the risk of sepsis (one trial, 10 OR 7.24, 95% CI 0.14 to 364.86; moderate quality) or death (two trials, 9, 10 OR 0.98, 95% CI

0.06 to 15.68; no difference (4 fewer to 4 more) per 1000; high quality) because both outcomes were so rare. The risk of anaphylaxis is uncertain (TMP-SMX OR 2.32, 95% CI 0.67 to 8.06; clindamycin OR 2.17, 95% CI 0.62 to 7.58; low quality, table 3).

6. Many of the cited trials have no evaluation at all of adverse events, therefore there is a bias since there is more information on efficacy and less information on adverse effects.

[Response]

We approached adverse events like all outcomes, as discussed above. Clindamycin

clearly causes diarrhoea and gastrointestinal symptoms, while TMP-SMX probably causes gastrointestinal symptoms and nausea. Rare adverse effects such as *Clostridium difficile* infection and Stevens-Johnson syndrome/toxic epidermal necrolysis (occurring in approximately 3-5/10,000-100,000 patients) are unlikely to be observed in RCTs. This is an important limitation of RCTs and we have discussed it in strengths and limitations section, please see our response to the comment # 2 from reviewer 2.

7. The current trials by Daum and Talan both have problems with missing data that seriously affect their conclusions and should rank them as having “high” risk of bias – not low. The fact that these trials have the largest sample size does not obviate the problem that both trials had large number of “administrative failures” defined as subjects who did not attend the test of cure visit. Such endpoints are obviously not of value to patients. All these patients were defined as “failures” in the intention to treat analysis and there was an imbalance in that more “administrative failures” were in the placebo group. This is an inherent bias against placebo. In addition, the administrative that are numerically as larger or larger than actual failures due to direct measured outcomes on the disease itself. The effect of missing data on these trials is to make the conclusions unclear and do not rule in benefit of the interventions in skin abscess. For instance, a “worst case” analysis presented at IDWeek by Talan et al showed that in 3 of 4 analyses the drug is not better than placebo and in one of those analyses placebo is statistically superior to drug on improving outcomes. This paper does point out that only 6 of 14 trials they evaluated had “infrequent” missing data. But the issue is not “frequency” but how the missing data affects conclusions. Sensitivity analyses on missing data completely change the conclusions of these studies in that the drugs lack effect or placebo performs better in these analyses.

[Response]

Thank you for the insightful comment. Missing data is a key issue for systematic reviews, to which we have dedicated a great deal of thought (e.g. *J Clin Epidemiol.* 2017 Jul;87:14-22). We agree with the reviewer that counting all participants who were lost to follow-up as treatment failures is inappropriate and almost certainly misleading. We do not make this assumption and instead use the complete case analysis (only those known to have had an event were considered to have had an event).

We also now examine the robustness of effect estimates to various assumptions about what happened to the patients who were lost to follow-up by performing sensitivity analyses using five alternative assumptions about missing data: none of the missing participants had an event, all missing participants had an event, best case scenario, worst case scenario, and worst plausible scenario (Table E in appendix 3). Sensitivity analyses using alternative assumptions resulted in a change in interpretation for several key outcomes including treatment failure, hospitalization, early and late

recurrence, pain, and additional surgical procedures, which were not robust to worst-plausible analyses. For the outcomes that were not robust to the most extreme plausible assumptions, we rated down certainty in the evidence for risk of bias due to missing participant outcome data according to the GRADE guideline. We supplemented the results of sensitivity analyses in our manuscript, it reads as:

For the results of the primary analysis suggested statistically significant treatment effect, sensitivity analyses using plausible assumptions about missing data were not robust to the worst plausible analysis (Table F in appendix 3).

8. The statement: "The difference in results is attributable to two recent large, low risk of bias trials, the addition of which raises the evidence to moderate or high quality for most outcomes" as stated on page 19 is misleading and incorrect. These trials are at high risk of bias for the reasons explained above. The issue of missing data and administrative failures driving the results calls the results of both trials into question. Implying to practicing clinicians that the data is more robust than it actually is misleads clinicians to prescribe drugs when they are not necessary based on presumed benefits that are not clear, and on outcomes that are not direct measure of patient benefit

[Response]

To address the reviewer's comment, we have removed the "low risk of bias trials" from the sentence. As mentioned above, we conducted sensitivity analyses using alternative assumptions about missing data. We rated down our certainty in the evidence for risk of bias for outcomes that were not robust to the worst plausible analysis.

9. The numbers needed to treat to harm in the two most recent studies are similar to the numbers needed to treat so just as many patients will be harmed with headache, nausea, diarrhea as would be helped by preventing an unclear "clinical failure". In the Daum trial clindamycin's adverse effects are statistically significantly increased compared to placebo and there were more serious adverse effects in the TMP-SMX group including a case of sulfa hypersensitivity which was stated to "resolve spontaneously" but can be fatal. Antibiotics are known to cause C diff colitis which has increased markedly in recent years and a recent study in JAMA showed it was responsible for the greatest increase in infection related deaths. The notion that because there were few to no cases in these studies that drugs like clindamycin (shown over years to cause the greatest proportion of C diff cases) don't result in C diff in abscess patients does not comport with years of evidence external to this trial.

[Response]

We completely agree that rare but serious antibiotic-related side effects should be taken into account when deciding whether or not to prescribe an antibiotic to any patient, and particularly so in this case where the benefits are modest. As suggested,

we have discussed this in the strengths and limitations section and conclusion, please see our response to the comment # 2 and comment # 3 from reviewer 2. A strength of our paper is that it highlights the magnitude of effect for each outcome, which will help patients decide for themselves whether or not the benefits are worth the risk of rare but serious adverse effects.

10. Given the framing of antibiotic resistance as a “world crisis” there needs to be some rationality in prescribing and withholding drugs in patients who have disease that spontaneously resolves with no evidence that drugs decrease serious complications. There is no difference in deaths or complications and the rates of these events are low to zero with placebo. It is misleading to imply to clinicians that because a pooled effect size greater than zero on disparate indirect outcomes exists, that it is rational to prescribe antibiotics in this setting. These trials do not address the effects of antibiotics in sicker or immunocompromised patients so extrapolating from these studies to other populations is not warranted.

[Response]

We conducted this systematic review to inform a BMJ Rapid Recommendation which aims to make rapid and trustworthy recommendations regarding new research that might change clinical practice. We have now removed any statements implying practice suggestions in this meta-analysis and instead only advocate for shared decision making, leaving the decision to use antibiotics or not to the individual patients and for guideline panel(s).

11. In summary, the data from trials in skin abscess remain inconclusive due to issues of bias and missing data and unclear outcomes and would recommend rejecting this analysis as it would misleadingly imply to clinicians that the data is stronger than it is.

[Response]

Thank you for all of the helpful comments. We hope that by addressing all of the points, in addition to our approach of transparently rating the quality of evidence for each outcome with the GRADE approach, we present a fair assessment of the quality of evidence for each outcome that we considered.

Reviewer 5

Major comment

1. Overall – When considering the different outcomes between TMP-SMX and clindamycin, one must consider the pathogens and resistance patterns. TMP-SMX has relatively poor efficacy against *Streptococcus*. Clindamycin should not be used against isolates with a positive “D-test.”

If pathogens and resistance patterns are known, then the study authors could perform sub-group analyses taking into consideration this clinical information. This would provide clinicians with even more useful information.

[Response]

Thanks very much for the comments. We agree with you that pathogens and resistance patterns are important information for clinicians – particularly so because there appears to be a misunderstanding by some clinicians in thinking that MRSA might not be the most common pathogen (it is). Many of the trials did report the percentage of patients with MRSA and MSSA and we have listed in the proportions table 1.

With respect to antibiotic resistance patterns, we now say: “The most common pathogen cultured was MRSA, the proportion of which ranged from 43.5% to 87.8%. None of the trials reported resistance rates of clindamycin and TMP-SMX”.

2. Discussion – Strengths and Limitations – The authors point out that they included four RCTs published more than 30 years ago.

It would be helpful to point out that those RCTs were conducted prior to the rise in CA-MRSA. Given that their chief finding in this meta-analysis is that anti-MRSA antibiotics worked better, they should alert the reader that those older studies were less likely to include patients with MRSA. This is a key argument that must be made. It can help explain why the newer studies found benefit with antibiotics and the older studies did not.

[Response]

Unfortunately none of the older studies reported the prevalence of MRSA; nor could they have because testing for methicillin resistance instead of penicillin resistance became common practice well after their publications. In one study that included two RCTs, the authors say that most of the patients had infections caused by penicillin-resistant *S. aureus*. Given the propensity for MRSA, more so than MSSA, to cause abscesses, we wonder if MRSA was the most important pathogen in patients with abscesses, even then. Whether the proportion of abscesses caused by MRSA in the trials has changed is speculative and we have therefore performed a sensitivity analysis that excludes these trials. There was no apparent differences in effect estimates between the newer and older trials, and they probably failed to reject their null hypotheses because they were underpowered (the effect size is quite small), and/or they used ineffective antibiotics (two studies used antibiotics without activity against MRSA).

Minor Comments

3. Abstract

a. Results – The last line says that TMP-SMX has a higher risk of “early recurrence,” but the data that are cited refer to outcomes at 1 month. I would not

call this early recurrence.

[Response]

We used recurrence at 1 month instead of early recurrence. It reads as below:

Compared to clindamycin, TMP-SMX has similar effects on treatment failure (NMA OR 1.08, 95% CI 0.69 to 1.75; 10 more [53 fewer to 41 more] per 1000 patients, high quality), lower risk of diarrhea (OR 0.29, 95% CI 0.16 to 0.55, 109 fewer [132 fewer to 66 fewer] per 1000 patients, high quality), but probably has higher risk of recurrence at 1 month (OR 2.14, 95% CI 1.11 to 4.12, 67 more [7 more to 163 more] per 1000 patients, moderate quality).

4. Introduction

a. The rise of CA-MRSA as a cause of SSTIs in the 1990s should be discussed to give some context as to why older and newer studies might have different findings. [Response]

Thank you for the comment. We agree that while changing microbiology is probably important, whether it is an effect modifier in this context is speculative. Further, we were unable to identify conclusive evidence that MRSA is a relatively more common cause of skin abscesses (not all SSTIs) than it was in the 1990s. The results of older studies are not substantively different than newer studies, suggesting that the reason recent studies detected a difference was that they had more power to detect a difference than the older studies.

5. Methods

a. Data Extraction – The authors state that pathogens were collected, but I do not see any report or analysis of these data in the Results or Discussion sections.

Please see table 1, which includes the proportion of patients with MSSA and with MRSA.

b. Data Extraction – The authors do not say that resistance information was collected. If not, then it should be noted as a limitation of the study. Also, the authors might want to comment on whether or not this information was available in the eligible studies.

[Response]

We planned to collect the information of pathogens and resistance patterns, however, none of the trials reported sufficient information about resistance patterns to be useful (especially CA-MRSA resistance to clindamycin or TMP-SMX). We can therefore only report the proportion of patients with MRSA and with MSSA due to limited data, please see it in table 1 and results section:

The most common pathogen cultured was MRSA, the proportion of which ranged from 43.5% to 87.8%.

In the results, we have also added:

None of the trials reported resistance rates of clindamycin and TMP-SMX.

6. Results

a. Comparative Effects of TMP-SMX versus Clindamycin on Other Outcomes – The finding about higher risk of abscess recurrence with TMP-SMX is based on a single study; therefore, this recommendation does not have strong enough evidence to warrant inclusion in the abstract's conclusion.

[Response]

We agree – and in fact, we have lowered the certainty in evidence to low from moderate for this outcome, for this comparison (because of imprecision and inconsistency with trials of TMP-SMX vs. placebo). We also removed it from the abstract.

7. Discussion

a. Strengths and limitations – It would be better to give specifics on how clinicians can use local microbiology and resistance patterns to make good choices (i.e., the prevalence of CA-MRSA, Streptococcus spp., and positive D-tests).

[Response]

Thanks for the comment. We completely agree that there are important individual patient factors to consider when choosing whether or not to use an antibiotic, and when choosing which antibiotic to use. We have revised our manuscript to discuss some of these considerations and to highlight the importance of context-dependent decision-making. Please see our response to the comment # 3 from reviewer 2.

VERSION 2 – REVIEW

REVIEWER	David A. Talan, MD UCLA, USA Author of one of the reviewed trials, otherwise no competing interests.
REVIEW RETURNED	12-Dec-2017

<p>GENERAL COMMENTS</p>	<p>General comments: This is a systematic review and meta-analysis of antibiotics for skin abscesses; the title should include "adjunctive" (antibiotics) and "drained" (abscesses) as indicated in the objective, since all trials reviewed included incision and drainage. After a broad search and imposing screening criteria, the meta-analysis includes both trials that compared antibiotics to no antibiotics or placebo, and those which compared antibiotics to each other. The authors conclude that their analysis suggests that TMP/SMX and clindamycin confer modest benefit for several important outcomes, but this is offset by similar risk of side effects.</p> <p>First, it is unclear to this reviewer how including antibiotic comparison studies is relevant to the question of benefits and risks associated with "adjunctive" antibiotics. Perusal of the 6 antibiotic comparison trials included is notable for 5 that do not report microbiology, 4 (Bucko a,b, Giordano, Keiichi) that compare similar agents (cephalosporins), any of which would have activity against MSSA (which would predominate at the time of the trials), and one that uses a comparison antibiotic that may not be active against MSSA (azithromycin) but no susceptibility results are provided. Only one trial (Miller) reports adequate microbiology and compares active drugs of different classes, but it has no placebo arm.</p> <p>Regarding studies to directly address the objective, the trials (summarized in Table 1) are dominated by two recent large and well-designed US multi-center, randomized, double-blind trials (Talan 1265 and Daum 796 participants), with comprehensive microbiology and demonstration that the antibiotics used had in vitro activity against the predominant cultured pathogen, i.e., MRSA (and MSSA as well). Other placebo/no antibiotic trials included in the meta-analysis done in the MRSA era with microbiology and using active antibiotics were much smaller (Duong 161 and Schmitz 212), only accounting for a 15% contribution. Regarding other placebo-controlled/no antibiotic trials, the authors decided to include one trial (Rajendran) in which MRSA caused 88% of infections but a non-active antibiotic was used (cephalexin), and a few small trials from the pre-MRSA era, two not blinded and without microbiology reported (Macfie a, b), one of which suture-closed drained abscesses, a non-standard technique (Macfie a), and one placebo-controlled trial (Llera) without microbiology done 30 years ago that likely used an active antibiotic, presuming these lesions were due to MSSA at that time.</p> <p>To this reviewer, inherent in the question of ANTibiotic (emphasis added) efficacy is a reasonable likelihood that the antibiotic is actually active in killing the likely pathogen, and for this we accept in vitro susceptibility testing. Therefore, the authors might reconsider their selection criteria, which potentially only leaves the Talan, Daum, Duong, and Schmitz studies, all of which were done in the MRSA era. Of note, both Talan (Ann Emerg Med 2017) and Daum reported subgroup analyses indicating outcome benefits associated with antibiotics compared to placebo for all <i>S. aureus</i> infections, both MRSA and MSSA. Limiting studies to blinded trials and presuming MSSA infection pre-millennium, one could also consider including the Llera study.</p> <p>Based on the independent mega-trials by Talan and Daum, using different populations and outcome definitions, but demonstrating very similar results, many feel that the question of the benefit of adjunctive (active) antibiotics is settled science and that the more interesting question is how the benefits balance against the costs and risks.</p>
--------------------------------	---

	<p>Benefits among primary outcomes include less treatment failures requiring another antibiotic course and/or another drainage procedure (and repeat healthcare visits), and secondary outcomes of fewer recurrent new site infections, hospitalizations, and household infections. These secondary outcomes have greater weight because of their general consistency with primary outcome results and each other, but remain to be formally studied as primary outcomes (and may never be), which is not addressed. Demonstrated risks are mainly slightly more mild self-resolving antibiotic-related side effects. Despite this paper's conclusion suggesting otherwise ("modest benefit...OFFSET by SIMILAR risk"[emphasis added]), the question of relative cost/risk/benefit is not well addressed by this research, and there are other methods to do this type of analysis.</p> <p>Further, at the outset, it is stated that patients were involved in the development of this paper. However, while wading into what would otherwise be shared-decision making, it is unclear how the patient advisors to this study might weigh the potential benefits (e.g., avoiding additional doctor's visits, antibiotics, and surgeries, and possible invasive infections) vs. cost (a course of TMP/SMX costs ~\$5USD vs. office/ED visit \$100s/hospitalization \$1000s) vs. risks (mild GI upset, rare C. difficile, and rarer Stevens Johnson Syndrome).</p> <p>Specific comments:</p> <p>Introduction: P. 6, para 2 - The 2 recent large trials did not suggest benefits may be conferred, they actually demonstrated statistically superior outcomes.</p> <p>P. 10, para 2 - It's unclear why one would hypothesize smaller effects of TMP/SMX vs. clindamycin or larger effects in children vs. adults.</p> <p>P. 12, para. 3 - The largest trial is not ref. #9, Daum (786 participants), rather ref. #10, Talan (1265 participants), which did not limit the maximal abscess size (other than amendable to outpatient care). Also, this is incorrect - "None of the trials reported resistance rates of clindamycin or TMP/SMX. Both the Talan and Daum trials did. While "two trials included proportion of patients with diabetes," some stated exclusion of diabetics (Daum, Llera, maybe others) so, in fact, they reported 0% diabetics. One might also consider mention of whether the trial described and used standardized methods for incision and drainage.</p> <p>p. 20, para. 2. - It is unclear the basis for stating that antibiotics may confer an even smaller benefit in patients who present to their GPs, especially since Daum and Talan trials (and subgroup analyses) demonstrated similar improved outcomes for smaller abscesses as well. Also, while mention is made of possible rare severe adverse drug reactions, similarly, it could be mentioned that the mega-trial studies were not powered to detect (and did not show) differences in rates of potentially antibiotic-presentable subsequent serious invasive infections, which is particularly a concern with MRSA.</p> <p>Table 1 includes papers with reference #s; it would also be helpful for other tables to include reference #s (instead "in 8 studies," etc.)</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

1. General comments: This is a systematic review and meta-analysis of antibiotics for skin abscesses; the title should include "adjunctive" (antibiotics) and "drained" (abscesses) as indicated in the objective, since all trials reviewed included incision and drainage. After a broad search and imposing screening criteria, the meta-analysis includes both trials that compared antibiotics to no antibiotics or placebo, and those which compared antibiotics to each other. The authors conclude that their analysis suggests that TMP/SMX and clindamycin confer modest benefit for several important outcomes, but this is offset by similar risk of side effects.

First, it is unclear to this reviewer how including antibiotic comparison studies is relevant to the question of benefits and risks associated with "adjunctive" antibiotics. Perusal of the 6 antibiotic comparison trials included is notable for 5 that do not report microbiology, 4 (Bucko a,b, Giordano, Keiichi) that compare similar agents (cephalosporins), any of which would have activity against MSSA (which would predominate at the time of the trials), and one that uses a comparison antibiotic that may not be active against MSSA (azithromycin) but no susceptibility results are provided. Only one trial (Miller) reports adequate microbiology and compares active drugs of different classes, but it has no placebo arm.

Regarding studies to directly address the objective, the trials (summarized in Table 1) are dominated by two recent large and well-designed US multi-center, randomized, double-blind trials (Talan 1265 and Daum 796 participants), with comprehensive microbiology and demonstration that the antibiotics used had in vitro activity against the predominant cultured pathogen, i.e., MRSA (and MSSA as well). Other placebo/no antibiotic trials included in the meta-analysis done in the MRSA era with microbiology and using active antibiotics were much smaller (Duong 161 and Schmitz 212), only accounting for a 15% contribution. Regarding other placebo-controlled/no antibiotic trials, the authors decided to include one trial (Rajendran) in which MRSA caused 88% of infections but a non-active antibiotic was used (cephalexin), and a few small trials from the pre-MRSA era, two not blinded and without microbiology reported (Macfie a, b), one of which suture-closed drained abscesses, a non-standard technique (Macfie a), and one placebo-controlled trial (Llera) without microbiology done 30 years ago that likely used an active antibiotic, presuming these lesions were due to MSSA at that time.

Response: To this reviewer, inherent in the question of ANTibiotic (emphasis added) efficacy is a reasonable likelihood that the antibiotic is actually active in killing the likely pathogen, and for this we accept in vitro susceptibility testing. Therefore, the authors might reconsider their selection criteria, which potentially only leaves the Talan, Daum, Duong, and Schmitz studies, all of which were done in the MRSA era. Of note, both Talan (Ann Emerg Med 2017) and Daum reported subgroup analyses indicating outcome benefits associated with antibiotics compared to placebo for all *S. aureus* infections, both MRSA and MSSA. Limiting studies to blinded trials and presuming MSSA infection pre-millennium, one could also consider including the Llera study.

Comment: Based on the independent mega-trials by Talan and Daum, using different populations and outcome definitions, but demonstrating very similar results, many feel that the question of the benefit of adjunctive (active) antibiotics is settled science and that the more interesting question is how the benefits balance against the costs and risks. Benefits among primary outcomes include less treatment failures requiring another antibiotic course and/or another drainage procedure (and repeat healthcare visits), and secondary outcomes of fewer recurrent new site infections, hospitalizations, and household infections.

These secondary outcomes have greater weight because of their general consistency with primary outcome results and each other, but remain to be formally studied as primary outcomes (and may never be), which is not addressed. Demonstrated risks are mainly slightly more mild self-resolving antibiotic-related side effects. Despite this paper's conclusion suggesting otherwise ("modest benefit...OFFSET by SIMILAR risk"[emphasis added]), the question of relative cost/risk/benefit is not well addressed by this research, and there are other methods to do this type of analysis.

Further, at the outset, it is stated that patients were involved in the development of this paper. However, while wading into what would otherwise be shared-decision making, it is unclear how the patient advisors to this study might weigh the potential benefits (e.g., avoiding additional doctor's visits, antibiotics, and surgeries, and possible invasive infections) vs. cost (a course of TMP/SMX costs ~\$5USD vs. office/ED visit \$100s/hospitalization \$1000s) vs. risks (mild GI upset, rare C. difficile, and rarer Stevens Johnson Syndrome).

[Response] We thank the reviewer for the comments. Our systematic review addressed two clinical questions: among patients with uncomplicated skin abscesses, what is the impact of antibiotic plus incision and drainage compared to incision and drainage alone, and what is the impact of different antibiotic options. We included all RCTs that included a comparison of antibiotics versus no antibiotics or a comparison of different types of antibiotics in patients with uncomplicated skin abscesses, regardless of the type of antibiotics. We conducted both pairwise meta-analysis and a network meta-analysis (NMA) to address these two questions.

We fully understood that the antibiotic regimen may change over time, and there was substantial variability in the preferred antibiotic regimen and antibiotic resistance. We conducted subgroup analysis to explain variability in effect estimates among antibiotics with or without MRSA coverage. We also conducted extensive sensitivity analyses to examine the impact of different scenarios on the effect estimates. The results were robust. For instance, we conducted a sensitivity analysis by omitting studies published more than 30 years ago, and the results were similar: treatment failure within 1 month (OR 0.56, 95% CI 0.34 to 0.93) and recurrence within 1 month (OR 0.45, 95% CI 0.27 to 0.74). Our findings were also consistent with the reviewer's comments that antibiotics with MRSA activity had better treatment outcomes.

We agreed that benefit, risk and cost are important factors to assess whether patients should be administered to antibiotics. Our review summarized available evidence regarding benefits and risk of these antibiotics. The guideline development group will use this information, and include the considerations about costs to inform the development of The BMJ Rapid Recommendations.

For this review, patient representatives were involved in the phase of question development. They were, however, not involved in the review production and decisions.

We are confident that the approach we used can adequately address the questions posed. The design and conduct of the systematic review was a result of collaboration between guideline developers and systematic reviewers.

Specific comments:

2. Introduction: P. 6, para 2 - The 2 recent large trials did not suggest benefits may be conferred, they actually demonstrated statistically superior outcomes.

[Response] Thanks for the comments. As suggested, we revised our expression, please see below:

"Recently, two large RCTs were published, 9,10 both of which suggested that adjunctive trimethoprim and sulfamethoxazole (TMP-SMX) or clindamycin may improve cure rate compared to placebo".

3. P. 10, para 2 - It's unclear why one would hypothesize smaller effects of TMP/SMX vs. clindamycin or larger effects in children vs. adults.

[Response] We hypothesized smaller effects with TMP-SMX versus clindamycin and larger effects with children versus adults according to previous evidence, in which children had a significantly higher cure rate when using clindamycin versus TMP-SMX or placebo, and this treatment advantage was greater than that seen among adults.

4. P. 12, para. 3 - The largest trial is not ref. #9, Daum (786 participants), rather ref. #10, Talan (1265 participants), which did not limit the maximal abscess size (other than amenable to outpatient care). Also, this is incorrect - "None of the trials reported resistance rates of clindamycin or TMP/SMX. Both the Talan and Daum trials did. While "two trials included proportion of patients with diabetes," some stated exclusion of diabetics (Daum, Llera, maybe others) so, in fact, they reported 0% diabetics. One might also consider mention of whether the trial described and used standardized methods for incision and drainage.

[Response] We have corrected the reference and revised our manuscript, as below:

"The resistance rates of clindamycin 9,24,32 ranged from 7.1% to 18%, while TMP-SMX 9,10,24,26,32 ranged from 0% to 2.6%."

"Two trials 10,27 included a proportion of patients with diabetes (2.4% to 11%), and seven trials 9,24,25,26,29,32 excluded patients with diabetes".

5. p. 20, para. 2. - It is unclear the basis for stating that antibiotics may confer an even smaller benefit in patients who present to their GPs, especially since Daum and Talan trials (and subgroup analyses) demonstrated similar improved outcomes for smaller abscesses as well.

[Response] The patient characteristics and their medical conditions may differ between emergency department and GPs. GPs may involve more patients with early stage of skin infection, in which abscesses may not be well developed. Without performing incision and drainage, the effect of antibiotics would be likely limited.

6. Also, while mention is made of possible rare severe adverse drug reactions, similarly, it could be mentioned that the mega-trial studies were not powered to detect (and did not show) differences in rates of potentially antibiotic-presentable subsequent serious invasive infections, which is particularly a concern with MRSA.

[Response] As suggested, we have added:

"Important but rare adverse events include anaphylaxis, C. difficile infection (especially with clindamycin³⁸), and Stevens-Johnson syndrome or toxic epidermal necrolysis (especially with TMP-SMX³⁹). Only one trial reported rate of serious invasive infection (0.2%-0.4%), however, the trial was under-powered to detect differences of this very rare but potentially fatal event."

7. Table 1 includes papers with reference #s; it would also be helpful for other tables to include reference #s (instead "in 8 studies," etc.)

[Response] Thank you for your suggestions. We believe that the current layout is good for the systematic review.