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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Vitamin C therapy for patients with sepsis or septic shock: a protocol for a systematic review and a network meta-analysis
AUTHORS	Fujii, Tomoko; Belletti, Alessandro; Carr, Anitra; Furukawa, Toshi; Luethi, Nora; Putzu, Alessandro; Sartini, Chiara; Salanti, Georgia; Tsujimoto, Yasushi; Udy, Andrew; Young, Paul; Bellomo, Rinaldo

VERSION 1 – REVIEW

REVIEWER	Faheem Guirgis University of Florida College of Medicine - Jacksonville
REVIEW RETURNED	02-Sep-2019

GENERAL COMMENTS	Summary Thank you for the opportunity of reviewing this protocol. It covers the hot topic of vitamin C, thiamine and steroid therapy for septic shock, and proposes to determine the most effective interventions by performing a network meta analysis. The comparison of multiple different drugs and dosing regimens of each drug adds a level of complexity to the study and could pose problematic. However, the authors have foreseen this issue and proposed a network meta analysis accounting for all drugs and dose ranges. I do, however, suggest that this article may require a more in-depth statistical review than I am able to provide. Other comments include the variability in the definitions and criteria for sepsis and septic shock, as well as standardization of the time point for mortality as the primary outcome of the study. Overall, however, I think this is very nicely done protocol and will add to the body of literature in this area. Major comments This is an important area of research and relevant to the current field. More studies of combination therapy of Vit C, thiamine and steroids are definitely needed, including this review. Under types of participants, I think it is important for the authors to use objective definitions for defining sepsis and septic shock, or at least for disease severity. Confounders of previous trials have included varying disease severity (think CORTICUS vs Annane trial for hydrocortisone in septic shock) and other factors regarding patient heterogeneity. With this in mind, a table comparing organ failure severity (SOFA) or critical illness severity (Apache) would be helpful when comparing studies. Under types of outcome measures, the primary outcome of mortality at the longest follow up within one year of randomization is problematic. There are widely variable changes in mortality over the 1st year from sepsis onset. If measured in the early period (for

	example, up to 90 days post-sepsis), the survival is much higher than in subsequent months up to 1-year, depending on critical illness severity, and whether the patient rapidly recovered or developed chronic critical illness. This point is made nicely in the recent paper by Gardner and colleagues.(1) For this reason, I think it would be stronger for the authors to pick a static time point (or time points) that are available for all studies to provide for a more apples to apples comparison. In addition, I do not think there will be enough studies in their search (though I could be wrong) that the differences in mortality endpoints will "wash out". The authors should mention in the body of the ms that they followed the PRISMA guidelines/checklist. Currently this information is only presented in a table. Under outcome data – again I suggest changing the primary outcome to static time interval I also suggest that for the secondary outcome that SOFA score be reported as medians and interquartile ranges, as SOFA score (or any score) is not a continuous variable and not normally distributed. Rank-sum should be used for univariate comparisons. I am not that familiar with the statistics for network meta-analyses. This article may require a statistical review, though in general the methods seem sound. Minor comments None References 1. Gardner AK, Ghita GL, Wang Z, et al. The Development of Chronic Critical Illness Determines Physical Function, Quality of Life, and Long-Term Survival Among Early Survivors of Sepsis in Surgical ICUs*. Crit Care Med. 2019;47(4):566-573. doi:10.1097/CCM.0000000000003655

REVIEWER	Harri Hemila University of Helsinki, Finland
REVIEW RETURNED	04-Sep-2019

GENERAL COMMENTS	Manuscript ID bmjopen-2019-033458, entitled "Vitamin C therapy for patients with sepsis or septic shock: a protocol for a systematic review and a network meta-analysis." Reviewer comments Harri Hemilä 2019-9-4 I dont consider that this project is useful science. First, there are several previous meta-analyses on vitamin C and sepsis or critically ill. One recent publication listed seven meta-analyses https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6664573/ Only one of those seven is cited in this protocol (ref 11), which gives a misleading impression about the number of previous closely overlapping meta-analyses.

	Ioannidis has been criticizing the explosion in the number of meta-analyses, with many of them being uninformative and not showing any novelty above the previously published meta-analyses on the same topics. https://www.ncbi.nlm.nih.gov/pubmed/27620683 https://www.ncbi.nlm.nih.gov/pubmed/29663048 It is not clear to me that this protocol leads to useful novelty in the meta-analysis. Second, in a ClinicalTrials.com search I can see that there are 21 trials on vitamin C and sepsis that are recruiting, not yet recruiting, or active but not recruiting. https://clinicaltrials.gov/ct2/results?term=%22vitamin+C%22&cond=sepsis&Search=Apply&recrs=b&recrs=a&recrs=d&age_v=&gndr=&type=&rslt= It is not evident from the protocol whether the authors plan to wait for the results of all the ongoing trials. If they publish their meta-analysis after, say, 5-10 of the ongoing trials are published, then their meta-analysis becomes outdated within a few years after a dozen more trials will become published. The authors motivate their protocol as: “The findings from this systematic review with network meta-analysis has the potential to inform future clinical trials” However, the findings of this meta-analysis cannot inform the planning of the above mentioned 21 trials. It is not clear that the findings could inform further trials. May be the researchers planning a new trial are more interested in all the RCTs published so far when they are planning, instead of an old meta-analysis. Third, the authors do not consider any kind of sample size analysis. They refer to the Marik before-after study (ref 12), but they fail to refer to several negative before-after studies on vitamin C and sepsis patients, see eg https://www.ncbi.nlm.nih.gov/pubmed/30970560 https://www.ncbi.nlm.nih.gov/pubmed/30654592 In any case, if we take the estimates from the Marik study, I can illustrate my point. Marik reported that hospital mortality was 8% in vitamin C + hydrocortisone+thiamine period (after), and 40% in the no treatment period (before). We can use web based calculators to estimate the number of participants that are needed for an RCT to meaningfully test such an effect. https://clincalc.com/stats/samplesize.aspx For alpha 0.05 and beta 0.2 and expected incidences 8% and 40% we can calculate that 54 participants are needed to test the effect (N = 27 to both arms). However, if we hypothesize that each of the substances has quite a similar effect and we want to test whether the addition of a second and a third substance further helps the patient, then we have a different context. Let us assume that untreated level is 40%, vitamin C alone level is 30%, vitamin C+thiamine level is 20%, and all the three reach the 8% level.

If we are interested in the testing of vitamin C alone (30%) against no treatment (40%), the same calculator tells us that we need 712 participants.
If we are interested in the testing of vitamin C alone (30%) against vitamin C+thiamine (20%), the same calculator tells us that we need 586 participants.
If we are interested in the testing of vitamin C+thiamine (20%) against the combination of all three (8%), the same calculator tells us that we need 260 participants.

The above proportions are pure guessing, but quite equal distances give the lowest needed sample sizes. If we assume, say, that by far the greatest proportion of effect comes from vitamin C (12%) so that the addition of the two other substances leads to the 8% level, we need 1764 participants to test such a hypothesized difference.

Although my above calculations are simplistic, they show the problem. Also, as I note above, there are before-after studies that did not find any benefit of vitamin C treatment, so the estimates based on Marik study can be greatly optimistic.

It is also well known that testing interactions needs many times more participants compared with testing the main effects.
Thus, testing whether vitamin C+thiamine is not just the sum of vitamin C effect and thiamine effect, needs lots of participants (some 4 fold). eg
“The results of the simulation study verify that the sample size required to detect a 2 × 2 interaction in a mixed-effects linear regression model is fourfold that to detect a main effect of the same magnitude.”
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2678722/

In figure 1, the authors have 16 treatments. To test the differences or lack of differences of such a large number of treatments may need >10000 participants. The authors do not consider at all the possibility to find out so many trials that they could estimate the size of effect of the listed 16 treatments.
They should consult a statistician to properly consider that number of participants that are needed to estimate 16 treatment effects.
Does it make any sense to have such an ambitious goal?

Methodologically, the protocol is quite ordinary.

However, protocols for systematic reviews have much less rationale than protocols for RCTs.
For example, for an RCT it is possible to decide that only patients older than 18 years will be included.
However, such a decision may be complicated for meta-analyses.
Eg. the authors write: (p 6, line 3)
“Patients with sepsis or septic shock aged 18 years or older will be included.“

What shall they do if there is a publication with patients aged from 15 to 30 and they cannot get the IPD?
The above statement does not tell us whether they reject the trial, or whether they consider that 15 years is “close enough” to 18

	years. If the latter, what then about a trial that had 10 to 25 year olds? p 5, line 49 “Quasi-randomised trials (such as alternating patients) will be excluded.” They do not give any justification to this exclusion. If there is allocation by the day of birth so that even and odd lead to different groups, how do the authors think there could arise systematic bias? There is no justification for that exclusion. It is also possible to carry out sensitivity analyses so that researchers exclude trials for various reasons and then look if the estimates are markedly changed. p 6, line 36 “We assume any patient who meets the inclusion criteria is, in principle, equally likely to be randomised to any of the interventions in the synthesis comparator set“ This is not a reasonable assumption. If there is one trial in India on vitamin C vs placebo, and another trial in USA on vitamin C vs thiamine, and a third trial in South Africa on vitamin C and hydrocortisone, there is no justification to assume that the patients were equally likely to be randomized to any of the four treatment groups of the three RCTs. They should state that their assumption most likely is not valid, given the great variety in patients, hospitals, treatment over the world, and the dynamics of hospital treatments: ie the treatments are evolving all the time so that the context in a single hospital can be substantially different over a time period of say 5 years. p 11, line 5 There is no justification to use OR. For common outcomes, RR is much better measure of effect. For common outcomes, OR exaggerates the effects. see eg https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1114216/ p 11, line 6 “standardised mean difference (SMD) for continuous outcomes“ SMD has its origins in psychology, but it is very poor in the analysis of biological phenomena that have relevant scales. For example, the Cochrane Handbook is critical about SMD: “However, the method assumes that the differences in standard deviations among studies reflect differences in measurement scales and not real differences in variability among study populations. This assumption may be problematic in some circumstances where we expect real differences in variability between the participants in different studies… The overall intervention effect can also be difficult to interpret as it is reported in units of standard deviation rather than in units of any of the measurement scales used in the review ”

	https://handbook-5-1.cochrane.org/chapter_9/9_2_3_2_the_standardized_mean_difference.htm Also “Presenting results as a standardized mean difference, the longest standing and most widely used approach, was poorly understood and perceived as least useful.“ https://www.ncbi.nlm.nih.gov/pubmed/26504102 Why dont the authors plan to use the relative scale for the continuous outcomes, eg https://www.ncbi.nlm.nih.gov/pubmed/21447428 https://www.ncbi.nlm.nih.gov/pubmed/28494765 p 11, line 26 “We assume that patients who fulfill the inclusion criteria for studies considered for this systematic review are potentially eligible for any of the interventions that we plan to compare.” There is no justification to this kind of assumption, see above.

REVIEWER	David N. Hager, MD, PHD Johns Hopkins University
REVIEW RETURNED	07-Sep-2019

GENERAL COMMENTS	et al. Vit C therapy for pts with sepsis or septic shock: a protocol of a systematic reviw and network meta-analysis. Thank you for allowing me to review this manuscript. It is well written and clearly outlines the rationale for the planned analyses, including the search strategy, inclusion and exclusion criteria, and the outcomes of interest. I have only a couple of comments: 1. The authors classify studies in terms of very high, high, and low dose vitamin C (not low is not described as low… it is not really low but relative to the others it is) OR based on mg/kg dosing. I think the former strategy is very prone to great heterogeneity based on variable weigths of enrolled patients. As a result, a small patient randomized to 6g/day could effectively be in the high dose group were they in a mg/kg study. This seems like a limitation of the former classification that should be mentioned. 2. The primary outcome will be all-cause mortality at the longest follow-up within one year. It seems lumping a study with a 30 day outcome like VICTAS in with a study that looks at 6 month follow up will not be an accurate representation of the data. I suggest you look at an endpoint that is similar between studies for a more fair comparison. 3. Figure 1 demonstrates a huge number of comparisons…. It seems you are highly likely to end up with many significant findings due to the number of comparisons alone…. How do you determine if these differences are by chance are real? Admittedly, I am not an expert in NMA, but some discussion of this would be of benefit to other non-experts that read your manuscript. 4. A discussion of the limitations of NMA as well as potential findings of your analysis should be included in the manuscript.

VERSION 1 – AUTHOR RESPONSE

Response to the reviewers' comments
Reviewer 1: Prof Faheem Guirgis
Thank you for your time in reviewing this manuscript. Please kindly find our itemised responses below.

Major comments
1. Under types of participants, I think it is important for the authors to use objective definitions for defining sepsis and septic shock, or at least for disease severity. Confounders of previous trials have included varying disease severity (think CORTICUS vs Annane trial for hydrocortisone in septic shock) and other factors regarding patient heterogeneity. With this in mind, a table comparing organ failure severity (SOFA) or critical illness severity (Apache) would be helpful when comparing studies.
Response: We agree with this comment. We will include a table outlining the reported organ dysfunction scores (SOFA or MODS) and severity of critical illness (APACHE or SAPS) for each trial population. "Characteristics of included studies and information flow in the network" under "Data synthesis" in the method section has now been revised accordingly.

2. Under types of outcome measures, the primary outcome of mortality at the longest follow up within one year of randomization is problematic. There are widely variable changes in mortality over the 1st year from sepsis onset. If measured in the early period (for example, up to 90 days post-sepsis), the survival is much higher than in subsequent months up to 1-year, depending on critical illness severity, and whether the patient rapidly recovered or developed chronic critical illness. This point is made nicely in the recent paper by Gardner and colleagues.(1) For this reason, I think it would be stronger for the authors to pick a static time point (or time points) that are available for all studies to provide for a more apples to apples comparison. In addition, I do not think there will be enough studies in their search (though I could be wrong) that the differences in mortality endpoints will "wash out".
Response: We appreciate this comment and the reference, which nicely describes long term survival; up to 12months post sepsis. As Gardner et al. illustrate, the mortality of patients with sepsis increases up to 3 months, and then plateaues in both patients with chronic critical illness and those who rapidly recover. Thus, we have revised the definition of the primary outcome to "all-cause mortality at the longest follow up within one year but 90 days or longer after randomisation." To provide information on long term outcomes, we will add a sensitivity analysis for all-cause mortality at the longest reported follow-up.

3. The authors should mention in the body of the ms that they followed the PRISMA guidelines/checklist. Currently this information is only presented in a table.
Response: We have added a "Study design" section at the beginning of the METHODS section and have clarified that the SR is and will be compliant to PRISMA-P and PRISMA extension for NMA, respectively.

4. Under outcome data – again I suggest changing the primary outcome to static time interval
Response: Thank you. Please see above. We have modified the definition of the primary outcome as suggested.

5. I also suggest that for the secondary outcome that SOFA score be reported as medians and interquartile ranges, as SOFA score (or any score) is not a continuous variable and not normally distributed. Rank-sum should be used for univariate comparisons.
Response: We agree that the SOFA score will have a skewed distribution. However, we would like to use means and SDs in meta-analysis as they are required to calculate the pooled effect estimate mathematically. We will also provide medians and IQRs, if there will be sufficient data to report them.

6. I am not that familiar with the statistics for network meta-analyses. This article may require a statistical review, though in general the methods seem sound.
Response: Thank-you.

Minor comments: None
References
1. Gardner AK, Ghita GL, Wang Z, et al. The Development of Chronic Critical Illness Determines Physical Function, Quality of Life, and Long-Term Survival Among Early Survivors of Sepsis in Surgical ICUs*. Crit Care Med. 2019;47(4):566-573. doi:10.1097/CCM.0000000000003655

Reviewer 2: Prof Harri Hemila

1. First, there are several previous meta-analyses on vitamin C and sepsis or critically ill. One recent publication listed seven meta-analyses
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6664573/
Only one of those seven is cited in this protocol (ref 11), which gives a misleading impression about the number of previous closely overlapping meta-analyses.
Ioannidis has been criticizing the explosion in the number of meta-analyses, with many of them being uninformative and not showing any novelty above the previously published meta-analyses on the same topics.
https://www.ncbi.nlm.nih.gov/pubmed/27620683
https://www.ncbi.nlm.nih.gov/pubmed/29663048
It is not clear to me that this protocol leads to useful novelty in the meta-analysis.
Response: We acknowledge that a number of SRs have already been conducted in this area, and the problems this can generate. Indeed, one of the investigators (AC) wrote the commentary paper referred to, while another (RB) published an opinion paper related to Ioannidis's work, further criticizing the explosion in the number of SR's. https://www.ncbi.nlm.nih.gov/pubmed/29663047
We are quite sceptical of and take a negative view of unnecessarily repeated systematic reviews. However, we submit that this SR with NMA will be essential in this area, considering the significant global interest in vitamin C for metabolic resuscitation, driven largely by a single-centre before-after study reported in 2017 (Marik study, ref 11 in the manuscript).
The study reported a dramatic decrease in mortality with a combination of vitamin C, thiamine and hydrocortisone for patients with severe sepsis or septic shock. A number of randomised controlled trials are being conducted to assess the effect of this combination therapy in patients with sepsis.
As there is so much enthusiasm for adopting the combination therapy proposed in the Marik study, a critical question has remained unanswered, that is, which component of the combination therapy is effective? This can be un-packed into two further questions: 1. What is the independent effect of each drug in the combination therapy, i.e. vitamin C, thiamine, and corticosteroid? And 2. Is there any synergistic effects of the combination? To answer this uncertainty, a conservative systematic review with pair-wise meta-analysis is not a suitable design, as simple pair-wise comparison cannot answer either of the questions above. Thus, we have designed this systematic review with component network meta-analysis which is the most comprehensive approach to answer these key questions.
We have revised the manuscript to clarify the rationale of our SR with component NMA in the Introduction.

2. Second, in a ClinicalTrials.com search I can see that there are 21 trials on vitamin C and sepsis that are recruiting, not yet recruiting, or active but not recruiting.
https://clinicaltrials.gov/ct2/results?term=%22vitamin+C%22&cond=sepsis&Search=Apply&recrs=b&recrs=a&recrs=d&age_v=&gndr=&type=&rslt=

It is not evident from the protocol whether the authors plan to wait for the results of all the ongoing trials. If they publish their meta-analysis after, say, 5-10 of the ongoing trials are published, then their meta-analysis becomes outdated within a few years after a dozen more trials will become published.
The authors motivate their protocol as: “The findings from this systematic review with network meta-analysis has the potential to inform future clinical trials”
However, the findings of this meta-analysis cannot inform the planning of the above mentioned 21 trials. It is not clear that the findings could inform further trials. May be the researchers planning a new trial are more interested in all the RCTs published so far when they are planning, instead of an old meta-analysis.
Response: We appreciate this comment. We are aware that many ongoing trials on this topic are registered in clinical trial registries. However, we do not expect that all of these trials will be published in the next 2 years, as meta-epidemiological studies consistently show that less than half of registered trials are eventually published. Moreover, we cannot tell which ones these will be.
https://www.ncbi.nlm.nih.gov/pubmed/22214755 https://www.ncbi.nlm.nih.gov/pubmed/19901971
For these reasons, Cochrane proposes a new approach called the “Living systematic review,” by which the body of evidence should be continuously updated by including new important findings as soon as possible. This approach allows evidence users to access up-to-date information.
Furthermore, the living network meta-analysis has been shown to produce strong evidence against the null hypothesis earlier than conventional pair-wise meta-analysis in a meta-epidemiological study. This benefit is crucial especially from an ethical point of view as researchers should not expose patients to any risk in a futile clinical trial.
https://www.ncbi.nlm.nih.gov/pubmed/29490922
As we state in the Acknowledgement section, we are conducting one of the clinical trials potentially eligible for this SR/NMA; the VITAMINS trial, ClinicalTrials.gov NCT03333278. Also, we indicate that several trials will be published in due course. We plan to close the dataset and analyse the available data when we complete Day 90 follow up in the VITAMINS trial. This will occur in October 2019. We have added this information in the ‘Ethics and dissemination’ section of the manuscript.

3. Third, the authors do not consider any kind of sample size analysis.
They refer to the Marik before-after study (ref 12), but they fail to refer to several negative before-after studies on vitamin C and sepsis patients, see eg https://www.ncbi.nlm.nih.gov/pubmed/30970560 https://www.ncbi.nlm.nih.gov/pubmed/30654592
In any case, if we take the estimates from the Marik study, I can illustrate my point.
Marik reported that hospital mortality was 8% in vitamin C + hydrocortisone+thiamine period (after), and 40% in the no treatment period (before). We can use web based calculators to estimate the number of participants that are needed for an RCT to meaningfully test such an effect.
https://clincalc.com/stats/samplesize.aspx
For alpha 0.05 and beta 0.2 and expected incidences 8% and 40% we can calculate that 54 participants are needed to test the effect (N = 27 to both arms).
However, if we hypothesize that each of the substances has quite a similar effect and we want to test whether the addition of a second and a third substance further helps the patient, then we have a different context.
Let us assume that untreated level is 40%, vitamin C alone level is 30%, vitamin C+thiamine level is 20%, and all the three reach the 8% level.
If we are interested in the testing of vitamin C alone (30%) against no treatment (40%), the same calculator tells us that we need 712 participants.
If we are interested in the testing of vitamin C alone (30%) against vitamin C+thiamine (20%), the same calculator tells us that we need 586 participants.
If we are interested in the testing of vitamin C+thiamine (20%) against the combination of all three (8%), the same calculator tells us that we need 260 participants.
The above proportions are pure guessing, but quite equal distances give the lowest needed sample sizes. If we assume, say, that by far the greatest proportion of effect comes from vitamin C (12%) so

that the addition of the two other substances leads to the 8% level, we need 1764 participants to test such a hypothesized difference.
Although my above calculations are simplistic, they show the problem. Also, as I note above, there are before-after studies that did not find any benefit of vitamin C treatment, so the estimates based on Marik study can be greatly optimistic.
It is also well known that testing interactions needs many times more participants compared with testing the main effects.
Thus, testing whether vitamin C+thiamine is not just the sum of vitamin C effect and thiamine effect, needs lots of participants (some 4 fold). Eg “The results of the simulation study verify that the sample size required to detect a 2 × 2 interaction in a mixed-effects linear regression model is fourfold that to detect a main effect of the same magnitude.”
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2678722/
In figure 1, the authors have 16 treatments. To test the differences or lack of differences of such a large number of treatments may need >10000 participants. The authors do not consider at all the possibility to find out so many trials that they could estimate the size of effect of the listed 16 treatments.
They should consult a statistician to properly consider that number of participants that are needed to estimate 16 treatment effects.
Does it make any sense to have such an ambitious goal?

Response: We appreciate the comment on sample size estimation for a future trial. In this systematic review, we are planning to conduct network meta-analysis, in which direct evidence from trials comparing treatments of interest (e.g. vit C + thiamine vs. thiamine), and indirect evidence from trials comparing the treatments of interest with a common comparator (e.g. vit C + thiamine vs. placebo, and thiamine vs. placebo) are synthesised. The NMA can increase precision in the estimations through making the best use of all available direct and indirect evidence, and we can compare interventions which haven’t been directly compared in any trial. Thus, the conventional sample size calculation for a single parallel-group randomised controlled trial is not applicable. The NMA is proposed as the first step to plan a future clinical trial, and we may be able to use the conditional power calculation based on the findings in the SR/NMA to design a future clinical trial. Further details have been explained in papers by a statistician in our team (GS).
https://www.ncbi.nlm.nih.gov/pubmed/25225031 https://www.ncbi.nlm.nih.gov/pubmed/29996869
As explained in our response to comment No.3, NMA informs us more efficiently on the evidence summaries than traditional pair-wise meta-analysis, which potentially reduces research waste. The aim of the current study is to assess the currently available evidence with the NMA approach as the first step in designing a future trial and is not limited to finding any statistical significance in the comparisons. The comparison and the sample size of the future trial will be dependent on the findings from the SR/NMA.
Finally, we appreciate we may not be able to answer each interaction question. However, we expect we will be able to create estimates of what that interaction might be and such estimates will create more precise boundaries around the number of patients needed to enrol to answer them. We think achieving such estimates is useful.

4. Methodologically, the protocol is quite ordinary.
However, protocols for systematic reviews have much less rationale than protocols for RCTs. For example, for an RCT it is possible to decide that only patients older than 18 years will be included. However, such a decision may be complicated for meta-analyses. Eg. the authors write: (p 6, line 3) “Patients with sepsis or septic shock aged 18 years or older will be included.“ What shall they do if there is a publication with patients aged from 15 to 30 and they cannot get the IPD? The above statement does not tell us whether they reject the trial, or whether they consider that 15 years is “close enough” to 18 years. If the latter, what then about a trial that had 10 to 25 year olds?

Response: We appreciate and agree with the comment. As we do not collect IPD for this SR, we added criteria for clarity.
“Patients with sepsis or septic shock aged 18 years or older will be included. Studies that included a minority (< 10%) of patients under 18 years, or studies with a median or mean age of patients over 20 years will be included.”

5. p 5, line 49
“Quasi-randomised trials (such as alternating patients) will be excluded.” They do not give any justification to this exclusion. If there is allocation by the day of birth so that even and odd lead to different groups, how do the authors think there could arise systematic bias? There is no justification for that exclusion. It is also possible to carry out sensitivity analyses so that researchers exclude trials for various reasons and then look if the estimates are markedly changed.
Response: Quasi-randomisation such as alternating patients or the allocation of intervention according to the day of birth allows investigators to predict which group the next eligible patient will be assigned. Such inadequate allocation concealment has been reported to overestimate the effect by 30% through the selection of study participants. As such, we submit that quasi-randomised trials should be excluded. We have added this explanation to the manuscript.
https://www.ncbi.nlm.nih.gov/pubmed/7823387

6. p 6, line 36
“We assume any patient who meets the inclusion criteria is, in principle, equally likely to be randomised to any of the interventions in the synthesis comparator set“ This is not a reasonable assumption. If there is one trial in India on vitamin C vs placebo, and another trial in USA on vitamin C vs thiamine, and a third trial in South Africa on vitamin C and hydrocortisone, there is no justification to assume that the patients were equally likely to be randomized to any of the four treatment groups of the three RCTs. They should state that their assumption most likely is not valid, given the great variety in patients, hospitals, treatment over the world, and the dynamics of hospital treatments: ie the treatments are evolving all the time so that the context in a single hospital can be substantially different over a time period of say 5 years.
Response: We appreciate this question. Page 6 line 36 in the original text referred to the concept that all the interventions we are comparing in this SR, have similar indications, namely as adjunct anti-inflammatory therapy for sepsis. Under this assumption, we can imagine a mega-trial with all treatments being compared. We have added an explanation for clarification, now on page 7.
We understand the reviewer’s concern regarding the comparability of different trials. In combining the effects reported in different comparisons, i.e. vit C vs placebo trials and vit C vs thiamine trials, effect modifiers should be similarly distributed across comparisons. This is also the transitivity assumption in the NMA, which allows the effects from the different comparisons to be combined through the common intervention, i.e. vit C in this instance.
Then, the question will be whether the effect we might observe in an Indian trial could be different if we run the same trial in the USA, or, whether we will find different effects when we run the same trial now as those conducted 10 years ago; AND whether those differences, if any, arise systematically and thus not by chance. We should be cautious about the systematic error because it brings intransitivity in the comparisons and should be addressed in the study design. Thus, we have mentioned in the manuscript that we would conduct an NMA if the included studies are sufficiently eligible for any of the interventions that we plan to compare.
We agree that it is plausible to assume the effect of the adjunct therapies in sepsis can be different over time because some empirical evidence showed an increased response to placebo in recent psychology clinical trials. Thus, we have added a sensitivity analysis to examine the robustness of the findings by excluding trials published before 2010.

7. p 11, line 5

There is no justification to use OR. For common outcomes, RR is much better measure of effect. For common outcomes, OR exaggerates the effects. see eg https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1114216/
Response: We acknowledge that the OR is vulnerable to misinterpretations and that the OR always show larger values than the RR, except in a case where the effect is null. We chose the OR because the calculations to estimate the component-specific effects in component NMA are dependent on the mathematical features of the OR; in turn, the RR is not suitable because it lacks symmetry, i.e. the conclusions may differ depending on whether we focus on the risk of death or survival. Furthermore, a methodological study has found possible bias in Cochrane random-effects meta-analysis using RR due to the restricted range of RR. https://www.ncbi.nlm.nih.gov/pubmed/30854785
We have added an explanation in the manuscript.

8. p 11, line 6
"standardised mean difference (SMD) for continuous outcomes" SMD has its origins in psychology, but it is very poor in the analysis of biological phenomena that have relevant scales. For example, the Cochrane Handbook is critical about SMD: "However, the method assumes that the differences in standard deviations among studies reflect differences in measurement scales and not real differences in variability among study populations. This assumption may be problematic in some circumstances where we expect real differences in variability between the participants in different studies… The overall intervention effect can also be difficult to interpret as it is reported in units of standard deviation rather than in units of any of the measurement scales used in the review " https://handbook-5-1.cochrane.org/chapter_9/9_2_3_2_the_standardized_mean_difference.htm
Also "Presenting results as a standardized mean difference, the longest standing and most widely used approach, was poorly understood and perceived as least useful." https://www.ncbi.nlm.nih.gov/pubmed/26504102
Why dont the authors plan to use the relative scale for the continuous outcomes, eg https://www.ncbi.nlm.nih.gov/pubmed/21447428 https://www.ncbi.nlm.nih.gov/pubmed/28494765
Response: We acknowledge the criticism of SMD as an effect measure for continuous outcomes. We do agree that the ratio of geometric means is a good measure for continuous outcomes. The difficulty here is that our continuous outcomes include organ failure scores which can be reported in changes from the baseline scores. When changes in scores are reported, we can still pool the data with SMD; however, ROM cannot be used to pool negative values. As such, we would like to add analyses using ROM for time to cessation of vasopressor therapy and ICU length of stay in the sensitivity analysis.

9. p 11, line 26
"We assume that patients who fulfill the inclusion criteria for studies considered for this systematic review are potentially eligible for any of the interventions that we plan to compare." There is no justification to this kind of assumption, see above.
Response: Please kindly refer to our response to comment No. 6.

Reviewer: 3 Prof.David N. Hager, MD, PHD
Thank you for allowing me to review this manuscript. It is well written and clearly outlines the rationale for the planned analyses, including the search strategy, inclusion and exclusion criteria, and the outcomes of interest. I have only a couple of comments:
Thank-you. Please kindly find our responses following each comment.

1. The authors classify studies in terms of very high, high, and low dose vitamin C (not low is not described as low… it is not really low but relative to the others it is) OR based on mg/kg dosing. I think the former strategy is very prone to great heterogeneity based on variable weigths of enrolled patients. As a result, a small patient randomized to 6g/day could effectively be in the high dose group

were they in a mg/kg study. This seems like a limitation of the former classification that should be mentioned.
Response: We appreciate this comment. As we plan to do meta-analysis using aggregated data, we can only classify each arm according to the dosage as reported by the investigators. We agree this might be prone to misclassification of the dose at each patient level pharmacologically, and have added this as a limitation in the 'Limitation' section on page 13.

2. The primary outcome will be all-cause mortality at the longest follow-up within one year. It seems lumping a study with a 30 day outcome like VICTAS in with a study that looks at 6 month follow up will not be an accurate representation of the data. I suggest you look at an endpoint that is similar between studies for a more fair comparison.
Response: We agree with the comment. We have revised the definition of the primary outcome to "all-cause mortality at the longest follow up within one year but 90 days or longer after randomisation," as the mortality of patients with sepsis reportedly plateaus after 90 days up to one year. We will provide the findings at the longest follow-up as a sensitivity analysis.

3. Figure 1 demonstrates a huge number of comparisons…. It seems you are highly likely to end up with many significant findings due to the number of comparisons alone…. How do you determine if these differences are by chance are real? Admittedly, I am not an expert in NMA, but some discussion of this would be of benefit to other non-experts that read your manuscript.
Response: We appreciate this comment on multiplicity in the NMA. There is no formal and satisfactory method to account for the multiple comparisons in the systematic review as the existing methods to adjust multiplicity in a single trial are not applicable. Thus, we do not plan to adjust for multiplicity, and have added text in the 'Limitations' section to draw attention to this.

4. A discussion of the limitations of NMA as well as potential findings of your analysis should be included in the manuscript.
Response: We have added a 'Limitations' section on page 11 as suggested. Thank you very much for your comment.

VERSION 2 – REVIEW

REVIEWER	Faheem Guirgis University of Florida College of Medicine, Jacksonville USA
REVIEW RETURNED	07-Oct-2019

GENERAL COMMENTS	They have addressed my concerns

REVIEWER	David N. Hager, MD, PHD Johns Hopkins University
REVIEW RETURNED	21-Oct-2019

GENERAL COMMENTS	The responses to reviewer comments are appropriate. Appropriate limitations are now included.