PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Interpretation of confidence intervals in clinical trials with non-significant results: systematic review and recommendations
AUTHORS	Gewandter, Jennifer; McDermott, M; Kitt, Rachel; Chaudari, Jenna; Koch, James; Evans, Scott; Gross, Robert; Markman, John; Turk, Dennis; Dworkin, Robert

VERSION 1 - REVIEW

REVIEWER	Karla Hemming University of Birmingham
REVIEW RETURNED	23-Apr-2017

GENERAL COMMENTS	I found this paper brilliant. I am a statistical reviewer for a top speciality journal and frequently find CIs reported but not properly interpreted. I also am a statistician on trials and so frequently find myself the statistician on a paper in which there is a non significant result, the CI wide and includes the possibility of clinially important effects but the PI or others in the group want to have a nice clean conclusion (to aid publication in a top journal) and so therefore want to conclude that the intervention does not work! I am so pleased to see this point addressed here. The paper is well written and I could not spot any issues which needed fixing, except at the bottom of pp 5 where I could not follow this section of text. My next point doesnt necessarily need to be addressed but the authors might like to consider: the percentages are conservative. For example it is said that 61 of 78 articles (78%) did not provide any interpretation of hte CI. However of the 99 studies included, 21 didnt even report the CI let alone interpret it. Therefore you could say that 61 of 99 studies did not interpret the CI. My only main concern with the paper is that the authors did not look at the interpretation of statistically significant CIs too. I suspect this could be another paper. But, it is interesting: how many times when saying a result is signficant do the authors determine if it is also clinically significant (I guess hardly ever).

REVIEWER	Andrew Copas MRC Clinical Trials Unit at University College London, UK
REVIEW RETURNED	25-Apr-2017

GENERAL COMMENTS	I found this an interesting article that raised some important issues

	and is generally well written. I have one major concern about the article and then two minor suggestions. My main concern about the article is that the concept of the minimum clinically meaningful treatment effect (MCMTE) is rather subtle and complex, more so than is reflected in the article. I think this does undermine the article somewhat, the concept certainly needs to be discussed in a more nuanced way in my opinion. My view is that in many scenarios the MCMTE for a patient is zero, or rather it is just greater than the null in the 'beneficial direction'. When declaring a statistically significant result after all we only see whether the 95% CI is above the null (beneficial) and not whether it is above an important level of benefit. Whether the MCME for a patient is null or not may depend on side-effects and other harms and benefits relating to the treatments compared. The MCMTE for a funder or a researcher may be much larger because research is often expensive and there may also be competing treatments which can alternatively be investigated. The MCMTE for the clinical community may also be greater than null because it may not be easy to change practice. Because of this complexity I feel that article readers will very often disagree on what is the MCMTE and therefore whether a non-significant trial result is either inconclusive or negative. For this reason I don't feel certain that there will be much value in researchers pre-specifying their MCMTE and using this in interpretation. Having said this I am fully supportive of authors describing how they interpret the 95% CI for the treatment effect (looking at secondary outcomes also) but any MCMTE they are using for this would need to justified including discussion of whether the value is from the patient or funder/researcher/clinical perspective. The authors do already mention in their discussion that if trialists are aware of disagreement in the clinical community on the MCMTE then this should be described, which I think is a very good recommendation. Minor comments 1. Relating to my main comment I think the authors should change that part of the discussion (page 11) where they discuss how they view the treatment effect used in a sample size calculation in relation to the MCMTE. I think the effect used in the calculation could be greater than the MCMTE, for example the anticipated effect, or be less than the MCMTE because trials need to be affordable. 2. I think the authors should include a description of how a meta-analysis of the current trial and previous trials of the same treatments should fit in to the interpretation of the results.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Karla Hemming
Institution and Country: University of Birmingham Please state any competing interests: None declared

Comment 1 - I found this paper brilliant. I am a statistical reviewer for a top specialty journal and frequently find CIs reported but not properly interpreted. I also am a statistician on trials and so

frequently find myself the statistician on a paper in which there is a non-significant result, the CI wide and includes the possibility of clinically important effects but the PI or others in the group want to have a nice clean conclusion (to aid publication in a top journal) and so therefore want to conclude that the intervention does not work! I am so pleased to see this point addressed here.

Response - Thank you very much!

Comment 2 - The paper is well written and I could not spot any issues which needed fixing, except at the bottom of pp 5 where I could not follow this section of text.

Response – Thank you for highlighting the need for clarification in this section. We attempted to clarify our meaning in the modified text from p. 5 and reproduced below.

A coding manual was developed to evaluate the frequency with which CIs were reported for the treatment effects in RCTs (Appendix 2). In the subset of articles that reported results that were not statistically significant for the primary outcome measure, coders were asked to evaluate whether the CI for the treatment effect indicated that the data were consistent with the absence of a clinically relevant treatment effect or that the results were inconclusive (i.e., the coders compared the CI for the treatment effect to a clinically relevant treatment effect declared by the authors at any point in the manuscript or the treatment effect specified in the sample size calculation if no clinically relevant treatment effect was described by the authors). Articles were excluded from this subset if they reported results that were both significant and not significant for the primary outcome measure (i.e., when multiple analyses were reported for the primary outcome measure). Articles were, however, included in this subset even if they reported a statistically significant treatment effect in a subgroup analysis or in analyses that were identified as sensitivity analyses because these analyses were considered secondary.

Comment 3 - My next point doesn't necessarily need to be addressed but the authors might like to consider: the percentages are conservative. For example it is said that 61 of 78 articles (78%) did not provide any interpretation of the CI. However of the 99 studies included, 21 didn't even report the CI let alone interpret it. Therefore you could say that 61 of 99 studies did not interpret the CI.

Response – Thank you for this important point. We have added results that use 99 for the denominator. Please see p. 8 (text reproduced below) for these modifications.

Eight-two (83%) of the 99 articles with statistically nonsignificant results did not provide any interpretation of the treatment effect using CIs. Sixty-one (78%) of the 78 articles that reported confidence intervals did not interpret them. In the 17 (17%) articles that did provide an interpretation of the treatment effect using CIs, the interpretations were of 5 types:

Comment 4 - My only main concern with the paper is that the authors did not look at the interpretation of statistically significant CIs too. I suspect this could be another paper. But, it is interesting: how many times when saying a result is significant do the authors determine if it is also clinically significant (I guess hardly ever).

Response – We agree that a discussion of whether the effect size from a trial with statistically significant results is also clinically meaningful is important in primary reports of RCTs and very much appreciate the suggestion to consider conducting another study to examine this issue. However, we are hesitant to suggest that CIs should be used as a basis in such a discussion for the following reason. For trials in which the treatment effect is statistically significant, it is often the case that the CI for the treatment effect will contain values of the treatment effect that would not be considered to be clinically meaningful. Such trials are invariably interpreted as being "positive", yet they often cannot

rule out small treatment effects. This is a byproduct of conventional trial design and hypothesis testing: a statistically significant result allows one only to conclude that the treatment effect is different from the null hypothesis value. Because the null hypothesis value is almost always zero rather than a treatment effect that is considered clinically meaningful, we are not sure how useful it will be to suggest that authors discuss whether the lower bound of the CI of the treatment effect includes differences that are not clinically meaningful, which is why we did not investigate whether this type of interpretation occurred.

Reviewer: 2
Reviewer Name: Andrew Copas
Institution and Country: MRC Clinical Trials Unit at University College London, UK Please state any competing interests: None declared

Please leave your comments for the authors below

I found this an interesting article that raised some important issues and is generally well written. I have one major concern about the article and then two minor suggestions.

Comment 1 - My main concern about the article is that the concept of the minimum clinically meaningful treatment effect (MCMTE) is rather subtle and complex, more so than is reflected in the article. I think this does undermine the article somewhat, the concept certainly needs to be discussed in a more nuanced way in my opinion. My view is that in many scenarios the MCMTE for a patient is zero, or rather it is just greater than the null in the 'beneficial direction'. When declaring a statistically significant result after all we only see whether the 95% CI is above the null (beneficial) and not whether it is above an important level of benefit. Whether the MCMTE for a patient is null or not may depend on side-effects and other harms and benefits relating to the treatments compared. The MCMTE for a funder or a researcher may be much larger because research is often expensive and there may also be competing treatments which can alternatively be investigated. The MCMTE for the clinical community may also be greater than null because it may not be easy to change practice. Because of this complexity I feel that article readers will very often disagree on what is the MCMTE and therefore whether a non-significant trial result is either inconclusive or negative. For this reason I don't feel certain that there will be much value in researchers pre-specifying their MCMTE and using this in interpretation. Having said this I am fully supportive of authors describing how they interpret the 95% CI for the treatment effect (looking at secondary outcomes also) but any MCMTE they are using for this would need to justified including discussion of whether the value is from the patient or funder/researcher/clinical perspective. The authors do already mention in their discussion that if trialists are aware of disagreement in the clinical community on the MCMTE then this should be described, which I think is a very good recommendation.

Response – Thank you for highlighting the need for a broader discussion of the nuances of the MCMTE and its interpretation and for your suggestions. We have added a paragraph to the Discussion to address these points. Please see p. 10 and reproduced below. We have also removed the statement regarding pre-specification of the MCMTE in protocols for interpretation. Finally, we modified the reporting recommendations (Table 2; p. 19) to reflect these changes (please see text reproduced below).

It must be acknowledged, of course, that the magnitude of a treatment effect that would be considered clinically meaningful can differ depending on many factors, including the setting of the trial and perspective of the reader.16-19 For example, if a treatment has very few side effects or no treatments currently exist for the condition, the minimal clinically meaningful treatment effect is likely to be relatively small compared to a treatment with greater safety risk. This may be especially true

from an individual patient's perspective. On the other hand, the minimal clinically meaningful treatment effect may be larger from a funder or researcher's perspective when considering whether to support or pursue a line of research. It is important that these potential differences in perspective are acknowledged when interpreting CIs and that the authors present the rationale for the minimal clinically meaningful treatment effect that they used to interpret the results of the trial.

Minor comments

Comment 1 - Relating to my main comment I think the authors should change that part of the discussion (page 11) where they discuss how they view the treatment effect used in a sample size calculation in relation to the MCMTE. I think the effect used in the calculation could be greater than the MCMTE, for example the anticipated effect, or be less than the MCMTE because trials need to be affordable.

Response – This is an important point, thank you. We have added an example of a potential reason why effect sizes other than the clinically meaningful treatment effect might be used for sample size calculations on p. 12 and reproduced below.

Furthermore, the treatment effects used to determine the sample size of a trial are not necessarily what one would consider to be the minimal clinically meaningful treatment effect that investigators might still be pleased to demonstrate.19 For example, investigators may justify the sample size using a larger effect than would be considered minimally clinically meaningful if an effect of this magnitude were anticipated based on existing data.

Comment 2 - I think the authors should include a description of how a meta-analysis of the current trial and previous trials of the same treatments should fit in to the interpretation of the results.

Response – Although an analysis like the one suggested above would certainly be useful to include in RCT reports to improve interpretation of the intervention's effects, inclusion of such a meta-analysis in primary reports of RCTs is currently very rare. We believe that it would be beyond the scope of our systematic review to suggest that such an interpretation should be a requirement for primary reports of RCTs and therefore would prefer not to add a discussion of this topic to our paper.

VERSION 2 – REVIEW

REVIEWER	Karla Hemming University of Birmingham
REVIEW RETURNED	05-Jun-2017

GENERAL COMMENTS	I love this paper. Andrew Copas raised a very interesting and valid point. But, I think the authors have addressed this point. This paper should help provide evidence that authors need to interpret as well as present CIs.

REVIEWER	Andrew Copas MRC Clinical Trials Unit and University College London, UK
REVIEW RETURNED	08-Jun-2017

GENERAL COMMENTS	I believe this article is much improved and my concerns have been well addressed. I have no further comments and congratulate the

	authors on this very good article.