PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Testing for non-inferior mortality: a systematic review of non-inferiority margin sizes and trial characteristics
AUTHORS	Pong, Sandra; Urner, Martin; Fowler, Robert; Mitsakakis, Nicholas; Seto, Winnie; Hutchison, Jamie; Science, Michelle; Daneman, Nick

VERSION 1 – REVIEW

REVIEWER	Rod Taylor University of Glasgow, Scotland, UK
REVIEW RETURNED	06-Oct-2020

GENERAL COMMENTS	This manuscript presents a systematic the reporting on non-inferiority RCTs The paper is well written, appears to have used a robust methodology and present some interesting new findings. There three issues that I would recommend require attention from the authors 1. Addition of a 'table 1' – in addition to the text reporting of study characterises in the results, I think the inclusion of Table 1 summarising the key study characteristics of the included trials would be helpful for the reader e.g. in addition to those presented in the text, I would add date of publication/continent of publication etc 2. Adjusted results – I am not sure a separate table 2 is warranted to present adjusted results. Instead I would suggest these adjusted results are incorporated into the current table 1 3. Interpretation of the results – given that the adjusted analyses show no evidence to support subgroups, I would suggest that the authors reword their current conclusions on the evidence of impact of timing since FDA guidance and industry-funding.

REVIEWER	Dominic Leung Liverpool Hospital University of New South Wales Australia
REVIEW RETURNED	13-Oct-2020

GENERAL COMMENTS	Pong and coworker evaluated the size and variability of non-inferiority margins used in non-inferiority trials of medications with primary outcomes involving mortality and to examine the association between trial characteristics and non-inferior margin size. A total of 111 articles were included in the systematic review. They found the median non-inferior margin was an absolute risk difference of 9% and the median relative risk of 1.5. They performed a multivariate (linear) regression analysis with medical specialty, inclusion of pediatric patients, mortality as sole or part of the primary outcome, presence of industry funding and found that only medical specialty

	was significantly associated with absolute non-inferiority margin size. Therefore, they conclude that the absolute and relative non-inferiority margins are large. I have the following comments: 1 While the objective of the systemic review was clearly stated, the hypotheses that were going to be tested were unclear. The therapeutic interventions in the trials included were medications and the endpoints included mortality, the specific interventions ie the new medications, were expected to be completely different. The conditions these medications were supposed to treat were completely different. As a result and more importantly, the treatment effects of the standard treatments tested in these trials would vary significantly. It is difficult to justifying lumping all such trials together and describe the size and variability of the non-inferior margins 2 One of the most important factors to consider in defining the non-inferior margin is the effectiveness of the standard treatment on the condition being tested. This important factor is not addressed in the manuscript. For example, if the standard treatment is higher effective against a particular condition, one can accept a higher loss of effectiveness of any new treatment in return to some perceived benefits of the new treatment. The 50% rule is often applied, ie 50% of the benefits of standard treatment is to be preserved, to define the non-inferiority margin. In this manuscript, in my opinion, it is more important to examine how many of the trials examined specified how the non-inferiority margins were defined and whether the efficacy of standard treatments was taken in considerations. 3 An advantage of using relative risk in defining the non-inferior margins is that an assumption of the estimated risks of event on standard treatment is not needed. The authors correctly pointed out that (Page 16, lines 25-30) “Since a relative non-inferiority margin accounts for the estimated baseline risk of outcome, it would be a more conservative choice over an absolute margin to conclude non-inferiority should the event rate in the control group be lower than expected” In fact, when one looks at the actual event rates and the estimated event rate in non-inferiority trials, the actual event rates were almost always lower than the estimated event rate (especially in the standard treatment arms). With an actual event rate actually lower than the estimated rate, the sample size required is going to be larger and the power of the trial lower, ie easier to declare non-inferiority. In this manuscript, it is also important to compare the observed (actual) event rate and the estimated event rate. A figure showing the relationship between actual and estimated risks would be more informative than Figure 2 given. 4. On page 13 lines 13-16, the authors stated the finding that trials with mortality as part of a composite primary outcome had significantly higher relative non-inferiority margins compared to those with mortality as a single primary outcome. This is entirely expected. A composite endpoint is associated with a higher event rate. Therefore, a standard treatment of that condition is likely to be more efficacious. With more efficacious standard treatments, applying the 50% rule is likely to result in a higher relative risk as non-inferiority margins. 5. I have concerns about the multivariate linear regression model

	and suggest review by biostatisticians. I do not understand how the explanatory variables were chosen. In the manuscript, medial specialty, inclusion of pediatric patients, mortality as a sole or part of a composite endpoint, presence of industry funding were entered as explanatory variables. Although there is no common consensus as to how explanatory variables are to be chosen, biological plausibility and results on univariate testings are often taken into consideration. I agree industry funding and mortality as an endpoint (as explained in point 4 above) may have some plausibility, but I find the inclusion of medical specialty and pediatric patients as explanatory variables hard to understand. Furthermore, there was only one references for most of the explanatory variables (eg pediatrics, industry funding, mortality outcome). 6. I would recommend inclusion of the multivariable linear regression equations with R and p values of the equations stated. This will help the reader in interpreting how much the variability in the dependent variable can be explained by the independent (explanatory) variables and the significance of the equations.

REVIEWER	Palash Ghosh Indian Institute of Technology Guwahti, India
REVIEW RETURNED	27-Jan-2021

GENERAL COMMENTS	Review report of the manuscript 'Testing for non-inferior mortality: a systematic review of non- inferiority margin sizes and trial characteristics': The article describes the size and variability of non-inferiority margins in non-inferiority trials with primary outcomes involving mortality and examining the association between trial characteristics and non-inferiority margin size. The use of multiple linear regression for the given primary outcome may not be appropriate here. It needs a correction. I hope my following comments may help the authors improve the "statistical methods and analyses" part of the article. 1. It is not easily clear from the text why the "absolute non-inferiority margin" is expressed in terms of percentage ("%") in Table 1, Figures 2 and 3? Perhaps, the authors mean absolute "percentage" difference of risk. It can be clearly written in the text with appropriate context to avoid any confusion. 2. For secondary objective, the authors have used multiple linear regression (MLR). For inference, MLR assumes the response variable is normally distributed. However, in this manuscript, the response variable is "non-inferiority margin" expressed in "percentage," which is unlikely to follow a normal distribution. It is evident from Figure 3A that "Absolute non-inferiority margin (%)" is following a skewed distribution rather than a normal distribution. The same is true for the other response variable, "Relative non-inferiority margin," in Figure 3B. Given the above reasons, the analyses presented in Tables 2 and 3 may not be appropriate. I suggest the authors take a log-transformation of the response variable(s) and then check the normality assumption. In that case, the authors can redo the analyses considering log- transformed response variable(s) once the assumption is verified. 3. In the entire manuscript, there are few (rounding-off/range)

	errors. Here, I point out some of them: a) The second last line, page 11: "…. non-inferiority margin observed was 1.5"; it should be a range 1.26-1.5. Otherwise, authors may report the midpoint of the interval. b) The second last line in the second last paragraph of page 11: "…. with a skewed distribution and distinct peaks observed at 5, 10 and 15%"; the middle number should be 9 instead of 10. c) The first line of page 13: absolute non-inferiority margin is reported as 3.5%. However, the same is 3.6% in Table 1.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Rod Taylor, University of Glasgow
Comments to the Author:
This manuscript presents a systematic the reporting on non-inferiority RCTs
The paper is well written, appears to have used a robust methodology and present some interesting new findings.
There three issues that I would recommend require attention from the authors
1. Addition of a 'table 1' – in addition to the text reporting of study characterises in the results, I think the inclusion of Table 1 summarising the key study characteristics of the included trials would be helpful for the reader e.g. in addition to those presented in the text, I would add date of publication/continent of publication etc

We have added a table describing the study characteristics of the included trials, including the date of publication and location(s) of the studies as requested by Reviewer 1. Due to the large size of the table containing 111 studies, we have included this in Appendix B.

2. Adjusted results – I am not sure a separate table 2 is warranted to present adjusted results. Instead I would suggest these adjusted results are incorporated into the current table 1

We performed regression analyses of the absolute non-inferiority margins and relative non-inferiority margins and presented these results (Tables 2 and 3) separately from the current Table 1 summary of characteristics of included trials. This was done intentionally because we wanted to distinguish the descriptive characteristics of the included trials from the results of analyses of the association between trial characteristics and size of the non-inferiority margins (absolute and relative). We would prefer to keep the current Table 1 summary of trial characteristics separate. However, we can combine the current Tables 2 and 3 together into a single table if preferred by the Reviewer and Editor.

3. Interpretation of the results – given that the adjusted analyses show no evidence to support subgroups, I would suggest that the authors reword their current conclusions on the evidence of

impact of timing since FDA guidance and industry-funding.

We have added the following clarifying sentences to the Discussion section:
-"However, the difference was not significant when relative non-inferiority margins were compared between trials with and without industry funding." (page 17, paragraph 3--referring to industry-funding)

-"This was significant only for relative non-inferiority margins, but not for absolute non-inferiority margins." (page 18, paragraph 1--referring to FDA guidance)

Reviewer: 2

Dr. Dominic Leung, University of New South Wales
Comments to the Author:
Pong and coworker evaluated the size and variability of non-inferiority margins used in non-inferiority trials of medications with primary outcomes involving mortality and to examine the association between trial characteristics and non-inferior margin size. A total of 111 articles were included in the systematic review. They found the median non-inferior margin was an absolute risk difference of 9% and the median relative risk of 1.5. They performed a multivariate (linear) regression analysis with medical specialty, inclusion of pediatric patients, mortality as sole or part of the primary outcome, presence of industry funding and found that only medical specialty was significantly associated with absolute non-inferiority margin size. Therefore, they conclude that the absolute and relative non-inferiority margins are large.
I have the following comments:
1. While the objective of the systemic review was clearly stated, the hypotheses that were going to be tested were unclear. The therapeutic interventions in the trials included were medications and the endpoints included mortality, the specific interventions ie the new medications, were expected to be completely different. The conditions these medications were supposed to treat were completely different. As a result and more importantly, the treatment effects of the standard treatments tested in these trials would vary significantly. It is difficult to justifying lumping all such trials together and describe the size and variability of the non-inferior margins

We acknowledge that the trials included in our review covered a wide range of pharmacological interventions for a variety of medical conditions. We intentionally did not limit our trials to specific medication classes or medical specialties because our focus was specifically on the size of absolute and relative non-inferiority margins used in trials that included mortality in their primary outcomes. To account for the variety of medical conditions, we adjusted for medical specialty in our multivariable regression analyses to test whether there was an association between specialty and non-inferiority margin. We believe that the diverse range of medications and conditions covered by the trials included in this review increases the overall generalizability of our results to all non-inferiority trials that include mortality in their primary outcomes.

We have clarified our hypothesis statement in the Introduction section.

-"We hypothesized that non-inferiority margins in these trials will be large and variable; and the size of non-inferiority margins will be related to the type of patients and medical conditions studied, as well as availability of industry funding and how mortality has been included in the outcome." (page 6, paragraph 2)

2. One of the most important factors to consider in defining the non-inferior margin is the effectiveness of the standard treatment on the condition being tested. This important factor is not addressed in the manuscript. For example, if the standard treatment is higher effective against a particular condition, one can accept a higher loss of effectiveness of any new treatment in return to some perceived benefits of the new treatment. The 50% rule is often applied, ie 50% of the benefits of standard treatment is to be preserved, to define the non-inferiority margin. In this manuscript, in my opinion, it is more important to examine how many of the trials examined specified how the non-inferiority margins were defined and whether the efficacy of standard treatments was taken in considerations.

We agree with Reviewer 2 that the effectiveness of the standard treatment on the condition being tested is an important consideration in selecting the non-inferior margin in non-inferiority trials. We have added the following to our Introduction section.

-"The size of non-inferiority margins could also be influenced by the effectiveness of the standard treatment. A highly effective standard treatment could allow researchers to tolerate higher thresholds for decreased effectiveness with a new treatment." (page 5, paragraph 2)

However, this is only one of many proposed rationales for selecting a non-inferiority margin, and there are many investigators that might argue that larger non-inferiority margins should not be tolerated for comparisons against highly effective standard treatments and that any loss of efficacy for mortality should be considered equally across diseases and treatments.

We did not examine how non-inferiority margins were defined in each trial because our review was deliberately focused on the size of the non-inferiority margin used in the trials. Prior systematic reviews have already focused on the topic of non-inferiority margin justification (Wangge et al. *PLoS ONE* 2010;5:e13550; Althunian et al. *Trials* 2017;18:107; Rehal et al. *BMJ Open* 2016;6:e012594). We took into consideration how the efficacy of standard treatments affected the size of non-inferiority margins when we analyzed non-inferiority margins as relative to the standard treatment in each trial and compared them to the results when the margins were expressed as absolute differences.

3. An advantage of using relative risk in defining the non-inferior margins is that an assumption of the estimated risks of event on standard treatment is not needed. The authors correctly pointed out that (Page 16, lines 25-30)
"Since a relative non-inferiority margin accounts for the estimated baseline risk of outcome, it would be a more conservative choice over an absolute margin to conclude non-inferiority should the event

rate in the control group be lower than expected"
In fact, when one looks at the actual event rates and the estimated event rate in non-inferiority trials, the actual event rates were almost always lower than the estimated event rate (especially in the standard treatment arms). With an actual event rate actually lower than the estimated rate, the sample size required is going to be larger and the power of the trial lower, ie easier to declare non-inferiority. In this manuscript, it is also important to compare the observed (actual) event rate and the estimated event rate. A figure showing the relationship between actual and estimated risks would be more informative than Figure 2 given.

We have added this figure to Appendix C to show the relationship between observed event rate and estimated risk of outcome in the control groups, as suggested by Reviewer 2.

4. On page 13 lines 13-16, the authors stated the finding that trials with mortality as part of a composite primary outcome had significantly higher relative non-inferiority margins compared to those with mortality as a single primary outcome. This is entirely expected. A composite endpoint is associated with a higher event rate. Therefore, a standard treatment of that condition is likely to be more efficacious. With more efficacious standard treatments, applying the 50% rule is likely to result in a higher relative risk as non-inferiority margins.

We agree with Reviewer 2.

5. I have concerns about the multivariate linear regression model and suggest review by biostatisticians. I do not understand how the explanatory variables were chosen. In the manuscript, medial specialty, inclusion of pediatric patients, mortality as a sole or part of a composite endpoint, presence of industry funding were entered as explanatory variables. Although there is no common consensus as to how explanatory variables are to be chosen, biological plausibility and results on univariate testings are often taken into consideration. I agree industry funding and mortality as an endpoint (as explained in point 4 above) may have some plausibility, but I find the inclusion of medical specialty and pediatric patients as explanatory variables hard to understand. Furthermore, there was only one references for most of the explanatory variables (eg pediatrics, industry funding, mortality outcome).

One of our team members and co-authors is a biostatistician (NM).

We chose to include pediatrics, industry funding and mortality outcome type as pre-specified explanatory variables because they are plausible factors to influence the choice of non-inferiority margin size. The reference which previously looked at these variables and their association with non-inferiority margins was a survey that was based on self-report by respondents. By including these factors in our review of published non-inferiority trials, we studied actual practice.

As mentioned in our response to Reviewer 2's point #1 above, we also adjusted for medical specialty in our multivariable regression analyses to account for the wide variety of medical conditions and interventions involved in the trials included in our review. In univariate analyses, there were significant differences in median non-inferiority margin sizes among different medical specialties regardless of whether the margin was expressed as an absolute difference or relative to the standard treatment.

6. I would recommend inclusion of the multivariable linear regression equations with R and p values of the equations stated. This will help the reader in interpreting how much the variability in the dependent variable can be explained by the independent (explanatory) variables and the significance of the equations.

We have included the adjusted R-squared values for each model. The p-values were previously provided and remain in the footnotes of Tables 2 and 3. Because the objective of the regression analyses was to assess whether selected trial characteristics were associated with non-inferiority margin sizes, the direction and significance of the independent variables adjusted for in the models are the main results that we are focusing on. Since we are not using the model equations to quantify the effect of the independent variables on the size of non-inferiority margins, the regression equations with the numerical coefficients written out are of less importance. We feel that the coefficient values presented currently in table form in Tables 2 and 3 provides the same information adequately.

Reviewer: 3
Dr. Palash Ghosh, Indian Institute of Technology Guwahati
Comments to the Author:
Please see the attached file.
Review report of the manuscript 'Testing for non-inferior mortality: a systematic review of noninferiority margin sizes and trial characteristics': The article describes the size and variability of non-inferiority margins in non-inferiority trials with primary outcomes involving mortality and examining the association between trial characteristics and non-inferiority margin size. The use of multiple linear regression for the given primary outcome may not be appropriate here. It needs a correction. I hope my following comments may help the authors improve the "statistical methods and analyses" part of the article.

1. It is not easily clear from the text why the "absolute non-inferiority margin" is expressed in terms of percentage ("%") in Table 1, Figures 2 and 3? Perhaps, the authors mean absolute "percentage" difference of risk. It can be clearly written in the text with appropriate context to avoid any confusion.

We have now clarified that absolute non-inferiority margins are expressed as "absolute risk differences in percentage" in the Methods section (page 9, paragraph 2)

2. For secondary objective, the authors have used multiple linear regression (MLR). For inference, MLR assumes the response variable is normally distributed. However, in this manuscript, the response variable is "non-inferiority margin" expressed in "percentage," which is unlikely to follow a

normal distribution. It is evident from Figure 3A that "Absolute non-inferiority margin (%)" is following a skewed distribution rather than a normal distribution. The same is true for the other response variable, "Relative non-inferiority margin," in Figure 3B. Given the above reasons, the analyses presented in Tables 2 and 3 may not be appropriate. I suggest the authors take a log-transformation of the response variable(s) and then check the normality assumption. In that case, the authors can redo the analyses considering log-transformed response variable(s) once the assumption is verified.

We have log-transformed the response variable (absolute and relative non-inferiority margins), which improved the skewness of the distribution of non-inferiority margin and improved the performance and diagnostics of the regression models. We have included selected diagnostic plots in Appendix D.

3. In the entire manuscript, there are few (rounding-off/range) errors. Here, I point out some of them:

a) The second last line, page 11: "…. non-inferiority margin observed was 1.5"; it should be a range 1.26-1.5. Otherwise, authors may report the midpoint of the interval.

b) The second last line in the second last paragraph of page 11: "…. with a skewed distribution and distinct peaks observed at 5, 10 and 15%"; the middle number should be 9 instead of 10.

c) The first line of page 13: absolute non-inferiority margin is reported as 3.5%. However, the same is 3.6% in Table 1.

We have made the corrections and double-checked the numbers.

VERSION 2 – REVIEW

REVIEWER	Dominic Leung University of New South Wales, Liverpool Hospital Cardiology
REVIEW RETURNED	16-Mar-2021

GENERAL COMMENTS	I would like to thank the authors for answering the queries and for making an effort to improve the manuscript. What I mean by biostatistician review is that the editors should get an independent biostatistician review (and not by one of the team members of the authors). I do not have any further suggestion as I feel that the manuscript is as good as it can get given the methodology, aims and the analysis.

REVIEWER	Palash Ghosh Indian Institute of Technology Guwahati, Department of Mathematics
REVIEW RETURNED	21-Mar-2021

GENERAL COMMENTS	1. I could not find 'appendix D' referred by the authors for 'selected diagnostic plots' to check. 2. For Table 2 and 3, the adjusted R-square are 0.44 and 0.1, respectively. They indicate (particularly for Table 3) that MLR gives a poor fit to the data. Any inference based on the results presented in Table 3 may not be reliable. Authors can discuss the goodness of fit

	of the model in detail and interpret the results accordingly. 3. For my third comment of the first review, the authors said, 'we have made the corrections and double-checked the numbers.' It is difficult to check authors' corrections if they do not mention the page, paragraph, and line numbers in their response.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Dominic Leung, University of New South Wales
Comments to the Author:
I would like to thank the authors for answering the queries and for making an effort to improve the manuscript. What I mean by biostatistician review is that the editors should get an independent biostatistician review (and not by one of the team members of the authors). I do not have any further suggestion as I feel that the manuscript is as good as it can get given the methodology, aims and the analysis.

We appreciate the feedback and comments--thank you.

Reviewer: 3
Dr. Palash Ghosh, Indian Institute of Technology Guwahati
Comments to the Author:
1. I could not find 'appendix D' referred by the authors for 'selected diagnostic plots' to check.

We have included Appendix D. This is referred to in the results on page 14, paragraph 3.

2. For Table 2 and 3, the adjusted R-square are 0.44 and 0.1, respectively. They indicate (particularly for Table 3) that MLR gives a poor fit to the data. Any inference based on the results presented in Table 3 may not be reliable. Authors can discuss the goodness of fit of the model in detail and interpret the results accordingly.

We have added the following to the results and discussion:

-"In our regression analyses of the association between trial characteristics and non-inferiority margin sizes, log-transformation of the non-inferiority margin (outcome variable) resulted in slight improvements to the performance of the regression models. The diagnostic plots of the regression models before and after log-transformation of the absolute and relative non-inferiority margins are provided in Appendix D." (Results, page 14, paragraph 3)

-"Although there was a large amount of variability in the regression models with low adjusted R-squared values, the direction and significance of the independent variables adjusted for in the models indicated that there was an important effect of medical specialty on non-inferiority margin size." (Discussion, page 20, paragraph 1)

3. For my third comment of the first review, the authors said, 'we have made the corrections and double-checked the numbers.' It is difficult to check authors' corrections if they do not mention the page, paragraph, and line numbers in their response.

The following changes were made:

-"There was a wide range of non-inferiority margins…distinct peaks observed at 5, 9 and 15%." (page 12, line 6)

-"The most common relative non-inferiority margin observed was in the range of 1.26 to 1.5." (page 12, line 10)

-"Thrombosis trials had the lowest median absolute non-inferiority margin of 3.6%." (page 13, line 3--starting below Table 1)

-Numbers in Tables 2 and 3 have all been changed after log-transformation of the outcome variable in the regression analyses. (page 15-16)