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)	Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals
AUTHORS	Rehal, Sunita; Morris, Tim; Fielding, Katherine; Carpenter, James; Phillips, Patrick

VERSION 1 - REVIEW

REVIEWER	David Gillespie Cardiff University, UK
REVIEW RETURNED	31-May-2016

GENERAL COMMENTS	I have several minor comments/queries that require addressing prior to me being able to recommend this paper for publication. Abstract: • We reviewed articles for non-inferiority – what do you mean by this? Suggest revising it • “Most trials declared non-inferiority” What is meant by this? Introduction: • I think the opening paragraph could have been stronger. The description of an equivalence trial was better than that of the non-inferiority trial. Can a better description, or a brief elaboration of what you mean by “acceptably worse”, be included? • In Table 1, the intention-to-treat description in the row describing the SPIRIT guidelines mentions “randomisation as a mechanism to avoid election bias”. Should this be “selection bias”? Results: • As part of your quality grading system, you include whether or not the type I error rate is consistent with the significance level of the confidence interval. However, you then conduct an analysis that investigates the association between your grading system and the conclusion of non-inferiority without specifying a significance level yourself. I’m also not convinced that there is a trend, even if there is an observed difference between groups. Also, how was the “other” category treated? If it was ignored, this needs to be stated. If it was included, is the Cochran-Armitage test applicable for a 3 x 4 table? Discussion: • Needs restructuring – the recommendations should be contained within a subsection, rather than scattered throughout • I’m not sure there can be a consistent definition of a per-protocol population. This will almost always be trial-specific. A per-protocol population should adhere to the guiding principles, which are fairly consistent in the guidance documents (received their allocated intervention and no protocol violations), but the precise definition of

	the population from trial to trial may differ and should be stated clearly • There is no mention that the direction of the suggested association between quality of reporting and conclusion of non-inferiority is contrary to what was hypothesised • Missing data potentially introduces selection bias, and the direction of this bias could be towards or away from a conclusion of non-inferiority. I cannot see a statement within reference 40 that says anything to the contrary. I therefore think that lines 17-20 on page 23 should be rewritten to reflect this (as it is currently written, it implies that missing data problems are particularly an issue for non-inferiority trials, as more missing data could bias results towards demonstrating non-inferiority – but this is not necessarily the case) • The strengths and limitations of your study are missing from the discussion

REVIEWER	Robine Donken National Institute for Public Health and the Environment / VU University Medical Center, The Netherlands
REVIEW RETURNED	02-Jun-2016

GENERAL COMMENTS	The paper by Sunita et al. strengthens the importance of clear and complete reporting of non-inferiority papers in general. It is of interest for statisticians and researchers involved in design and reporting, but also for those interpreting non-inferiority trials. Title: The title might indicate a study on how many times non-inferiority trials report inferior results. As the topic is reporting and performance of non-inferiority trials, the reviewer suggests a slight adjustment, to cover the whole research in more advance. P4 Lines 11-13. As suggested by the authors use of a non-inferiority design is only recommended if the intervention has some other benefit. This is especially of influence when there is a biocreep potential, i.e. a new treatment might show non-inferiority, but is also inferior. Previous research by Bernabe et al. (BMC Med Res Methodol 2013) has showed that there was room for improvement in phase IV trials on these additional benefits. Did the authors check for reporting on these other benefits? Table 1. Indicates differences between the available guidelines on non-inferiority. Please guide the reader through these differences and possible implications in the background section. Additionally the outcomes which were scored by the authors, might be influenced by which guideline was used by researchers of the different papers. P9 Line 12 The paper was updated till May 2015. Half way this period an extension of the CONSORT statement was published. This might have influenced the results as was shown in previous papers by Donken et al. (Vaccine 2015) and by Schiller et al. (Trials 2012) Please elaborate of an update of search will influence the results or not. P9 Line 12-13 Several journals mentioned here might expect authors to report based on a specific guideline before considering a manuscript for publication. Please indicate the journal-specific recommendations as this might influence the reporting rates.

	Table 2. The “categories” of diseases can be experienced as confusing. Tuberculosis and hepatitis C could be seen as infectious diseases but are a separate category. Can the authors indicate what is considered as “non-infectious disease” or “infectious disease”? Table3. The 10-12% indicate that there is other guidance than the M1/M2 method described in table 1 alone. Could the authors give some guidance on this possible other method? P21 Lines 5-7 The authors state that there is disagreement between the different guidelines on vital issues. However the authors report no specific recommendations for authors of non-inferiority studies how to handle this, or call the different institutions behind guidelines to come up with unambiguously guidelines. P24 Lines 23-24 Besides enforcing authors to justify the choice of margin before publication by the journal, isn’t this of importance before approval by a medical ethics committee is given, because it is of influence on the sample size. In the introduction the authors promise to give some guidance for trialist who are working with non-inferiority. There are some recommendations throughout the paper, but a clear overview at one point might be of additional value for the readers.

REVIEWER	Dr Ben Ewald University of Newcastle, NSW Australia
REVIEW RETURNED	23-Jun-2016

GENERAL COMMENTS	Overall This systematic survey of non inferiority studies is a good assessment of the state of the art, and points out where analysis and reporting could be improved. It is very dry and methodological so will of limited interest to a general readership. It should be required reading for authors of non inferiority studies, but I wonder if BMJ open is the right journal. It could be made more relevant to a general readership by including examples where the weaknesses of reporting might adversely influence clinical decisions. The only real weakness of this study is the very limited double checking of data extraction. 8 of 168 papers were double read, leaving doubts about reproducibility. Details Page 9, para 2, Of course a data extraction form was used, and I assume ball point pens ? This seems excessive description. Page 9 para 3. Another cause of false non inferiority is the use of non discriminatory or unresponsive outcome measures. The quality of the trial could include the quality of the outcome measure. Page 10, line 5. It is a real weakness of this review that so little checking of data extraction was done. Double checking 8 papers out of 168 is inadequate. This is recognised as a limitation but I think should be rectified before publication. Page14 line 56; It would be interesting to illustrate with some examples the point that justification was ambiguously worded. Maybe in the discussion. P22 line 14. Setting the MCID is a difficult judgement for clinicians, no matter how many you ask, and I suspect if a researcher asks 100

	clinicians from 20 institutions (eg by a survey) the answer will be less valid than the researcher who engages in detailed discussion with a smaller number of clinicians who really come to grips with the problem. Thinking about the Dabigatran vs warfarin trial, and asking clinicians how many extra strokes are acceptable the answer is likely to be no extra strokes. Clinicians have different priorities to researchers. This question would be well illustrated by quoting examples of a well justified and a poorly justified non-inferiority margin. Typos and formatting P4 line 29 focus instead of focuses. Page4, from line 26- this paragraph would be clearer formatted as a list. Page 21 line 16. Compliment is used when complement was meant, but even complement is a clumsy way of saying this. Why not say that methods for non inferiority designs are yet to be optimised?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: David Gillespie
Institution and Country: Cardiff University, UK Competing Interests: None declared

I have several minor comments/queries that require addressing prior to me being able to recommend this paper for publication.

Abstract:
• We reviewed articles for non-inferiority – what do you mean by this? Suggest revising it
Response:
This has been rephrased as “We searched for non-inferiority trials”.

• “Most trials declared non-inferiority” What is meant by this?
Response:
This has been rephrased as “Most trials concluded non-inferiority”.

Introduction:
• I think the opening paragraph could have been stronger. The description of an equivalence trial was better than that of the non-inferiority trial. Can a better description, or a brief elaboration of what you mean by “acceptably worse”, be included?
Response:
Thank you for your comment. We have rephrased our description slightly and given further explanation: “Non-inferiority trials assess whether a new intervention is not much worse when compared to a standard treatment or care. These trials answer whether we are willing to accept a new intervention that may be clinically worse, yet still be beneficial for patients while having another advantage, such as less intensive treatment, lower cost or fewer side effects.”

• In Table 1, the intention-to-treat description in the row describing the SPIRIT guidelines mentions “randomisation as a mechanism to avoid election bias”. Should this be “selection bias”?
Response:
Yes, thank you.

Results:
• As part of your quality grading system, you include whether or not the type I error rate is consistent with the significance level of the confidence interval. However, you then conduct an analysis that investigates the association between your grading system and the conclusion of non-inferiority without specifying a significance level yourself. I'm also not convinced that there is a trend, even if there is an observed difference between groups. Also, how was the "other" category treated? If it was ignored, this needs to be stated. If it was included, is the Cochran-Armitage test applicable for a 3 x 4 table?

Response:
Thank you for your comment. We used a 5% significance level and this has been added in the methods section accordingly "at the 5% significance level".
We have modified "trend" to "difference" to better explain a p-value of 0.05 is a suggestive difference.
The 'other' category was ignored and this is footnoted underneath table 7. For added clarity we have stated the same within the main text as follows:

"Trials that were classed as having 'other' conclusion about whether non-inferiority was met were excluded from the analysis."

Discussion:
• Needs restructuring – the recommendations should be contained within a subsection, rather than scattered throughout
Response:
Thank you for your comment. We have placed our recommendations in a table for quick reference (table 8).

• I'm not sure there can be a consistent definition of a per-protocol population. This will almost always be trial-specific. A per-protocol population should adhere to the guiding principles, which are fairly consistent in the guidance documents (received their allocated intervention and no protocol violations), but the precise definition of the population from trial to trial may differ and should be stated clearly
Response:
We agree that the definition of per-protocol is trial specific and that authors ought to be clear who is included in that analysis. We were surprised that the definitions of a per-protocol analysis were so variable. With regards to treatment, some defined the per-protocol population as "received treatment", "completed treatment", "received the correct treatment", "received at least one dose of medication" and more. While these criteria are all treatment related there are some subtle differences here that could change the conclusions made. We have modified this point in our discussion as follows:
"Most authors included treatment related exclusions such as "received treatment", "completed treatment" or "received the correct treatment". Such differences in definitions may be superficially small but could in fact make critical differences to the results of a trial."

• There is no mention that the direction of the suggested association between quality of reporting and conclusion of non-inferiority is contrary to what was hypothesised
Response:
Thank you for this comment. We have included this within the discussion:
"We anticipated that poor reporting of articles would bias towards concluding non-inferiority, however, the poorly reported trials were less likely to demonstrate non-inferiority. This is somewhat reassuring. Nevertheless, it is essential…"

• Missing data potentially introduces selection bias, and the direction of this bias could be towards or away from a conclusion of non-inferiority. I cannot see a statement within reference 40 that says anything to the contrary. I therefore think that lines 17-20 on page 23 should be rewritten to reflect this (as it is currently written, it implies that missing data problems are particularly an issue for non-inferiority trials, as more missing data could bias results towards demonstrating non-inferiority – but this is not necessarily the case)
Response:
This is a fair comment, but given that missing data are often related to patients who are doing poorly, then if there are more dropouts on the research arm, the inferiority of this arm may be masked in complete cases. In a superiority study, this effect would bias away from the alternative hypothesis for showing a difference between treatments, but in a non-inferiority trial it would bias towards the alternative hypothesis of showing no difference between treatments.

We have replaced the reference with Wiens et al (2013; statistics in biopharmaceutical research 5:383-393).

• The strengths and limitations of your study are missing from the discussion
Response:
The strengths and limitations of the study have been included after the discussion as follows:
"This research demonstrates the inconsistency in the recommendations for non-inferiority trials provided by the available guidelines, which was also reflected within this review. We have provided several recommendations using examples for researchers wishing to use the non-inferiority design and have highlighted the importance of missing data and using sensitivity analyses specific to non-inferiority trials. There are also some limitations in this review. Firstly, a justification of the choice of the margin was recorded as such if any attempt was made to do so. And so one could argue that inadequate attempts were counted as a 'justification', however there was good agreement between reviewers when independently assessed. Secondly, only one reviewer extracted information from all articles and therefore assessments may be subjective. However, there was good agreement when a random 5% of papers were independently assessed, and the categorisation of the justification of the non-inferiority margin was also independently assessed in all papers where a justification was given. Thirdly, an update of the CONSORT statement for non-inferiority trials was published during the period of the search in 2012, which could improve the reporting of non-inferiority trials over the next few years. However, the first CONSORT statement for non-inferiority trials published in 2006 was released well before the studies included in our search and we have found that reporting of non-inferiority trials remains poor."

Reviewer: 2
Reviewer Name: Robine Donken
Institution and Country: National Institute for Public Health and the Environment / VU University Medical Center, The Netherlands Competing Interests: None declared

The paper by Sunita et al. strengthens the importance of clear and complete reporting of non-inferiority papers in general. It is of interest for statisticians and researchers involved in design and reporting, but also for those interpreting non-inferiority trials.

Title: The title might indicate a study on how many times non-inferiority trials report inferior results. As the topic is reporting and performance of non-inferiority trials, the reviewer suggests a slight adjustment, to cover the whole research in more advance.
Response:
We have amended the title to:

“Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals”.

P4 Lines 11-13. As suggested by the authors use of a non-inferiority design is only recommended if the intervention has some other benefit. This is especially of influence when there is a biocreep potential, i.e. a new treatment might show non-inferiority, but is also inferior. Previous research by Bernabe et al. (BMC Med Res Methodol 2013) has showed that there was room for improvement in phase IV trials on these additional benefits. Did the authors check for reporting on these other benefits?
Response:
This is not something we checked.

Table 1. Indicates differences between the available guidelines on non-inferiority. Please guide the reader through these differences and possible implications in the background section. Additionally the outcomes which were scored by the authors, might be influenced by which guideline was used by researchers of the different papers.
Response:
Thank you for this comment. The differences between the guidelines highlight the confusion between each and therefore confusion for researchers using non-inferiority trials, which are broadly described in the introduction. We have included the main differences as follows:
• For the non-inferiority margin, we have added that “all guidelines” recommend justifying the choice of the non-inferiority margin on a clinical basis and have modified whether statistical considerations should be chosen:
“However, it is unclear whether statistical considerations should also impact on the choice of an appropriate margin as recommended by the Draft FDA 2010, ICH E10 and EMEA 2006 guidelines (table 1). Ignoring statistical evidence from meta-analyses or systematic reviews could have the potential for clinicians to choose an unrealistic margin.”
• For analyses we have added the following:
“In particular, the CONSORT 2006 guidelines describe the PP analysis as excluding patients not taking allocated treatment or otherwise not protocol-adherent, whereas the ICH E9 guidelines state that the PP analysis is a “subset of patients who complied sufficiently with the protocol, such as exposure to treatment, availability of measures and absence of major protocol violations”. These obscure definitions could lead researchers to arbitrarily exclude patients from analyses. The draft FDA guidelines recommend researchers to use an ITT and as-treated analysis, although it is unclear what is meant by ‘as-treated’ as this is not defined within the guidelines.”
• For missing data, we have added the following:
“The ICH E9 guidelines recommend using the now outdated last observation carried forward imputation method and the more recent SPIRIT guidelines recommend multiple imputation, but caution the reader that it relies on untestable assumptions”

We agree that guidelines used by authors may have influenced the choices researchers made, however all journals recommend using the CONSORT statements/checklist.

P9 Line 12 The paper was updated till May 2015. Half way this period an extension of the CONSORT statement was published. This might have influenced the results as was shown in previous papers by Donken et al. (Vaccine 2015) and by Schiller et al. (Trials 2012) Please elaborate of an update of search will influence the results or not.
Response:
Thank you for this useful comment. The results shown by Donken et al show a 6% increase in how the non-inferiority margin was determined in articles published after 2006 when compared to those published between 1996-2005. This is quite an underwhelming improvement given that these articles were published several years after the CONSORT statement. The paper by Schiller is referenced in

our review (please see page 21), and we found an improvement in justifying the choice of the non-inferiority margin, although we were generous in what we classed as 'justification'. Both papers conclude overall there was little improvement in reporting since the publication of the CONSORT statement.
In our review we included journals from 2010, four years after the first extension of the CONSORT statement for reporting of noninferiority and equivalence randomised trials (published in 2006). At the time of collecting data, this guideline was 9 years old and reporting of non-inferiority trials has continued to be poor, even after the further endorsement of the 2012 statement. It is unlikely that an update of the search to include articles in this review from the last year would improve by a significant amount if at all, but possible that the guideline could improve reporting over the next two or three years. We have added the following to our strengths and limitations in the discussion:
"Thirdly, an update of the CONSORT statement for non-inferiority trials was published during the period of the search in 2012, which could improve the reporting of non-inferiority trials over the next few years. However, the first CONSORT statement for non-inferiority trials published in 2006 was released well before the studies included in our search and we have found that reporting of non-inferiority trials remains poor."

P9 Line 12-13 Several journals mentioned here might expect authors to report based on a specific guideline before considering a manuscript for publication. Please indicate the journal-specific recommendations as this might influence the reporting rates.
Response:
All journals recommend the CONSORT statements. We have added the following to the methods:
"All journals refer authors to the CONSORT statement and checklist when reporting."

Table 2. The "categories" of diseases can be experienced as confusing. Tuberculosis and hepatitis C could be seen as infectious diseases but are a separate category. Can the authors indicate what is considered as "non-infectious disease" or "infectious disease"?
Response:
Thank you for pointing this out. These categories were chosen to group diseases that were singular. Any contagious disease that can be spread from one person to another was classed as infectious. We have changed these categories to 'other' and have re-ordered the categories to read non-infectious diseases followed by infectious diseases.

Table3. The 10-12% indicate that there is other guidance than the M1/M2 method described in table 1 alone. Could the authors give some guidance on this possible other method?
Response:
The 10-12% category is what authors stated in their article from disease specific FDA guidelines. Two articles referenced FDA for HIV drug development, one was FDA guidance for developing antiretroviral drugs for treatment and another referred to unofficial FDA recommendations for anti-infectious trials. This category has been changed from "10-12% recommended by FDA guidelines" to "10-12% recommended by disease specific FDA guidelines" to reflect this.

P21 Lines 5-7 The authors state that there is disagreement between the different guidelines on vital issues. However the authors report no specific recommendations for authors of non-inferiority studies how to handle this, or call the different institutions behind guidelines to come up with unambiguously guidelines.
Response:
Thank you for this comment. The aim of this paper is to highlight the issues between the guidelines

which inevitably will lead to variable reporting. We agree that there should be some consensus from the various guiders, but we do not have the authority here to instruct them to come to one. We have, however, included a table of recommendations (suggested by another reviewer as well) for future publications of non-inferiority trials which we hope will be taken up into future guidelines.

P24 Lines 23-24 Besides enforcing authors to justify the choice of margin before publication by the journal, isn't this of importance before approval by a medical ethics committee is given, because it is of influence on the sample size.
Response:
Thank you for this useful comment. We agree that the choice of the margin should be determined early on before approval from an ethics committee due to changes in sample size. However, there may be cases where the non-inferiority margin changes after approval due to the availability of more recent information (recent systematic review, or meta-analysis for example). We have therefore added the following to our conclusion:

"If the choice of the non-inferiority margin changes following approval from an ethics committee, justification for the change and changes to the original sample size calculation should be explicit."

In the introduction the authors promise to give some guidance for trialist who are working with non-inferiority. There are some recommendations throughout the paper, but a clear overview at one point might be of additional value for the readers.
Response:
Thank you for your comment, which was also made by another reviewer. We have added a table of recommendations following our conclusions.

Reviewer: 3
Reviewer Name: Dr Ben Ewald
Institution and Country: University of Newcastle, NSW, Australia Competing Interests: None declared

Review-BMJ open-Non inferiority trials.
Overall
This systematic survey of non inferiority studies is a good assessment of the state of the art, and points out where analysis and reporting could be improved. It is very dry and methodological so will of limited interest to a general readership. It should be required reading for authors of non inferiority studies, but I wonder if BMJ open is the right journal.

It could be made more relevant to a general readership by including examples where the weaknesses of reporting might adversely influence clinical decisions.

The only real weakness of this study is the very limited double checking of data extraction. 8 of 168 papers were double read, leaving doubts about reproducibility.

Details
Page 9, para 2, Of course a data extraction form was used, and I assume ball point pens ? This seems excessive description.
Response:
Very good! The aim of this riveting paragraph is to inform the reader of what information we aimed to

extract and also of the results our review will report. It is dry but, in our opinion, important.

Page 9 para 3. Another cause of false non inferiority is the use of non discriminatory or unresponsive outcome measures. The quality of the trial could include the quality of the outcome measure.
Response:
Thank you for raising this important point. We feel that a potential bias in the choice of the outcome is something researchers should definitely be cautious about, although we do not believe we have any authority to judge whether specific outcome measures are non-discriminatory. We have therefore added the following to the discussion:
"It is possible that the quality of a trial may also depend on the quality of the outcome; unresponsive outcomes that miss important differences between treatments may be intentionally or unintentionally chosen to demonstrate non-inferiority. Therefore it is also important that the outcome chosen is robust."

Page 10, line 5. It is a real weakness of this review that so little checking of data extraction was done. Double checking 8 papers out of 168 is inadequate. This is recognised as a limitation but I think should be rectified before publication.
Response:
Thank you for your comment. We acknowledge this limitation in the discussion. While this is a known limitation of these reviews, we limited subjectivity by independently reviewing articles for inclusion before data was extracted from all articles until reaching a 100% agreement and a further 8 (5%) underwent independent data extraction where there was a high level of agreement. The questions were not too tricky to answer most of the time as we extracted hard data, hence very high agreement when articles were double-reviewed. The categorisation of the justification for the margin of non-inferiority, which could be subjective, was independently extracted in 100% of papers where a justification was given. We have added that justifications for the choice of the margin were independently reviewed from our additional supplement to the methods section:
"Justifications for the choice of the non-inferiority margin were reviewed by two reviewers (SR and PP)."

Page14 line 56; It would be interesting to illustrate with some examples the point that justification was ambiguously worded. Maybe in the discussion.
Response:
To actively change the way justification of the margin is reported by researchers and editors of journals, we feel it would be more productive to provide well justified examples authors can access (such as Gallagher et al, 2012; 308(12):1221-6, JAMA) to enable them to produce more robust justifications rather than to specifically criticise them. Ambiguously worded justifications often came in the form of "this margin was deemed appropriate" or "the margin was clinically acceptable". To emphasize that these justifications are inappropriate we have modified that justifications were "inadequate" rather than "poor" and have added another ambiguous justification as follows:
"A statement often used in articles reviewed was 'the choice of the margin was clinically acceptable'. This statement does not contain enough information to justify the choice of the non-inferiority margin."

We have also added to the discussion that it is important to reference the estimates of the control arm from previous trials to preserve the treatment effect as follows:
"There were very few articles that referred to preserving the treatment effect based on estimates of the standard of care arm from previous trials. It is vital that authors acknowledge this to ensure the standard of care is effective. If the control were to have no effect at all in the study then finding a small difference between the standard of care and new intervention would be meaningless."

P22 line 14. Setting the MCID is a difficult judgement for clinicians, no matter how many you ask, and I suspect if a researcher asks 100 clinicians from 20 institutions (eg by a survey) the answer will be less valid than the researcher who engages in detailed discussion with a smaller number of clinicians who really come to grips with the problem. Thinking about the Dabigatran vs warfarin trial, and asking clinicians how many extra strokes are acceptable the answer is likely to be no extra strokes. Clinicians have different priorities to researchers.
This question would be well illustrated by quoting examples of a well justified and a poorly justified non-inferiority margin.
Response:
Thank you for your very helpful comments. We agree that choosing the margin is difficult for clinicians, but with the expectation a new intervention is non-inferior and therefore the new intervention would be administered to patients by all clinicians, we believe that it is important to look for assessments of the non-inferiority margin from outside the closed group of investigators who are conducting the research and to have a larger number of clinicians who really come to grips with the problem. One example in an article we reviewed conducted a survey during a symposium and this could be the way forward for making clinical assessments. We have therefore added the following: "Radford et al justify the choice of the non-inferiority margin after performing a delegate survey at a symposium. This method may be a way forward for researchers to obtain clinical assessment from a large group of clinicians. Even better would be to obtain formal assessments, using for example the Delphi method which has been used in the COMET initiative, after presenting the proposed research at a conference or symposium for clinicians to really engage with the question at hand."

As above, we feel it would be more productive to provide well justified examples authors can access.

Typos and formatting
P4 line 29 focus instead of focuses.
Response:
We have changed this to "focus".

Page4, from line 26- this paragraph would be clearer formatted as a list.
Response:
We have incorporated this suggestion.
Page 21 line 16. Compliment is used when complement was meant, but even complement is a clumsy way of saying this. Why not say that methods for non inferiority designs are yet to be optimised?
Response:
We have incorporated this suggestion.

VERSION 2 – REVIEW

REVIEWER	Robine Donken National Institute for Public Health and the Environment / VU University Medical Center, The Netherlands
REVIEW RETURNED	09-Sep-2016

GENERAL COMMENTS	P8 lines20-21 "Ignoring statistical evidence from metaanalyses or systematic reviews could have the potential for clinicians to choose an unrealistic margin." This is not only potentially a problem for clinicians but for any researcher. P19 Lines50-51 "Trials that were classed as having some 'other' conclusion about non-inferiority were excluded from the analysis." It

	is not totally clear for me what the authors mean here. Do they mean trials in which non-inferiority was not concluded?