## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Haphazard Reporting of Deaths in Clinical Trials – a Review of Cases of ClinicalTrials.gov Records and Matched Publications: cross-sectional study |
| **AUTHORS** | Uhlig, Katrin; Earley, Amy; Lau, Joseph |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Andrew Prayle,<br>Clinical Research Fellow, University of Nottingham.<br><br>Competing interests - I have previously published with data from ClinicalTrials.gov. |
| **REVIEW RETURNED** | 12-Sep-2012 |

| | |
|---|---|
| **THE STUDY** | This study highlights shortcoming on reporting of trial mortality results as shown through discrepancies between the number of deaths reported on the ClinicalTrials.gov website compared to the peer reviewed manuscript. This is an important aspect of clinical trials reporting – one would think it would be possible to report things as simple to define as mortality consistency across data sources about a trial.<br>The methods used appear appropriate. However, it isn't fully clear to me why the authors did not also search for a publication themselves when one wasn't given on ClinicalTrials.gov. Our group has found that the 'publications' section of the ClinicalTrials.gov record frequently does not give citation details when in fact a paper has been published. This could possibly have increased the sample size, and made the findings more robust.<br><br>Having read and re-read the methods section regarding how the authors determining the total death count, I am still not sure exactly how some records came to have a total death count assigned and some not. Perhaps an example could be given in the appendix of a record where the total death count could not be determined? |
| **RESULTS & CONCLUSIONS** | Overall, the number of trials in the main analysis (54) is low. Have the authors done everything they can to maximise the number of included trials? I think that as the authors have essentially taken a sample from a population of larger studies, some estimate of certainty should be used in the statistics, such as 95% confidence intervals.<br><br>Something odd has happened in the denominator of Table 2, Case 6, Arm 1, Publication column – it drops from 2563 to 1262 – is this a typographical error?<br><br>Are the authors planning to put the dataset into a public repository? |
| **GENERAL COMMENTS** | I think that this paper raises an important issue with the reporting of |

| | |
|---|---|
| | trials. However, the main limitations are the sample size (which may not be able to be increased further), and difficulty in deciding how the authors had decided when they could not determine a total death count. |

| | |
|---|---|
| **REVIEWER** | Goetzsche, Peter<br>The Nordic Cochrane Centre, This is the reg record he wishes to use |
| **REVIEW RETURNED** | 20-Sep-2012 |

| | |
|---|---|
| **GENERAL COMMENTS** | I think there is something interesting in this paper that we can learn from, but it needs to be written much better and in a way that allows us to understand whether the problems are major or minor without consulting the tables. It is really difficult to follow the flow of information in this paper, and the language is also difficult and sometimes inappropiate. I feel the senior author should have contributed more, as he is capable of writing far better than in this manuscript.<br><br>Abstract<br><br>Line 17: please write "whether" counts may be discordant (not "how" as<br>you don't know whether you will find any and therefore cannot say how).<br><br>I have not seen the term "death counts" before and it feels pretty odd, like the US Republications' false allegations of government "death panels", please consider using another term, e.g. number of deaths, which is how we describe this.<br><br>Please describe how many pairs you sampled right from the start.<br><br>I do not understand how there can be pairs in the database, as a pair<br>comes from an entry in the database and a published report, furthermore, if there was no information on deaths, how can then the information be discordant, particularly since some publications reported absence of death just as in the registry. This is too confusingly written.<br><br>Under Conclusion we hear about death count per arm, which was not what<br>we heard about in Results, there must be consistency.<br>—<br><br>P4, line 8: "the discrepant reporting expectations for death," confusing construct, please revise.<br><br>Line 14: we now hear about all cause death, which is commonly called<br>all cause mortality, have not heard about this before, only death counts. You need to be consistent.<br><br>Line 27-34: not clear, use much simpler language so that we can follow<br>your thinking. |

| | |
|---|---|
| | Line 52: I cannot recall you have defined what you mean by reporting<br>expectations?<br><br>There is far too much in the Article Summary, page 4, and it continues<br>on p 5, please abbreviate considerably. The first limitation on p 5 is not really a limitation, and not relevant for the Summary; it can be mentioned in the Discussion.<br><br>Several places: I do not like calling people or participants for subjects or individuals, seems a bit denigrating to me.<br><br>P6 line 14: the idea of explicitly stating even if no deaths occurred is a good one although it may seem over the top in many cases, e.g. in<br>a small migraine trial. But we so often wonder whether no deaths occurred when nothing is mentioned about this that I believe it would be better if all papers involving patients needed to say that no one died if that is the case.<br><br>Line 21: I am highly surprised that there are no regulations mandating<br>the reporting of all deaths in a trial, I would have expected there were, at least for drug trials. Are you absolutely sure about this? In your text below you say that death is a serious adverse event and that<br>it therefore must always be reported in drug trials. This appears inconsistent.<br><br>P7: again, I seriously doubt that if a drug company does not consider a death related to their drug they are not required to report all deaths to FDA.<br><br>P7: I was really amused to read that “deaths from serious adverse events would presumably be a reason for not completing a trial.” I am<br>pretty sure that dead patients cannot complete a trial! I am not really sure what the idea is that is described in lines 32-34, does it matter that deaths are only reported in the serious adverse events section, and not as a reason for discontinuation?<br><br>P1, line 36: what is “a result for death in another metric?” Metric has to do with measuring, e.g. kg, and a death is a death so I think your use of terms is not optimal.<br><br>Line 58: I have forgotten whether the 27 pairs where randomly selected<br>from all the pairs, I therefore went back to the top of P9 where I could see that you selected 75 matching publications of which 27 corresponded to database records of reported deaths. At this point, it<br>is not exactly clear to me how the study was done. It may be my fault,<br>but I am usually able to understand what I read and I am now a bit confused about the methods.<br><br>P11, lines 10-20: I wonder why concordance could not be assessed in<br>12 of the 27 pairs when they were carefully selected in order to have |

| | |
|---|---|
| | number of deaths per arm for both trial database and publication. I also miss some information about what was different, e.g. was it one or two deaths out of many or what?<br><br>Lines 25-40: I would not call it discordant when the database does not<br>say anything about deaths and the publication says that were no deaths.<br><br>P12, line 14: so, when treatment arms were missing, did you call this concordant or discordant? I suggest that you focus on the really important issues, e.g. P14, line 31 to 36, it is very serious when it is not possible to judge how many died in a trial in the two treatment arms because of ambiguity.<br><br>I also suggest the authors write something about that just a couple of<br>deaths may bias the trial considerably and if it happens in several trials, we can get it very wrong.<br><br>P15, line 23: please be careful with your choice of words, “death reporting may never be complete or simple given the challenges in ascertainment and adjudication.”<br><br>I always favour total number of deaths whatever the cause, which I suggest you also write about here as we know that assessment of cause<br>death can be biased, sometimes to a considerable degree. So do not mix<br>up total number of deaths, which you describe somewhere in your paper<br>and cause specific mortality.<br><br>P15, line 45: you say here “In summary, our study shows lack of clarity…..” Forgive me for saying this, but this is exactly the problem with your study, that the way it has been written up is not sufficiently clear (although this was not what you meant in this sentence of course).<br><br>One of your references is to the paper about better reporting of harms<br>according to CONSORT. I therefore wonder why you do not quote this<br>paper specifically, as it says under recommendation 6 that authors should always report deaths in each study group during a trial, regardless of whether death is an endpoint and regardless of whether<br>attributions to a specific cause is possible. I believe you should quote this. |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: Andrew Prayle, Clinical Research Fellow, University of Nottingham.

Competing interests - I have previously published with data from ClinicalTrials.gov.

This study highlights shortcoming on reporting of trial mortality results as shown through discrepancies between the number of deaths reported on the ClinicalTrials.gov website compared to

the peer reviewed manuscript. This is an important aspect of clinical trials reporting – one would think it would be possible to report things as simple to define as mortality consistency across data sources about a trial.
The methods used appear appropriate. However, it isn't fully clear to me why the authors did not also search for a publication themselves when one wasn't given on ClinicalTrials.gov. Our group has found that the 'publications' section of the ClinicalTrials.gov record frequently does not give citation details when in fact a paper has been published. This could possibly have increased the sample size, and made the findings more robust.

Response: We asked the ClinicalTrials.gov team to run a search for us using the available information. Also our intention was not to provide estimates of uncertainty. Rather to demonstrate a disconnect between reporting expectations and real-life practices.

Having read and re-read the methods section regarding how the authors determining the total death count, I am still not sure exactly how some records came to have a total death count assigned and some not. Perhaps an example could be given in the appendix of a record where the total death count could not be determined?

Response: Please see new Appendix 3.

Overall, the number of trials in the main analysis (54) is low. Have the authors done everything they can to maximise the number of included trials? I think that as the authors have essentially taken a sample from a population of larger studies, some estimate of certainty should be used in the statistics, such as 95% confidence intervals.

Response: Our study highlights the issues resulting from lack of unequivocal reporting guidelines for reporting of deaths in clinical trial reports. We started out with 500 records but had attrition of records due to the lack of clarity in reporting leaving a smaller number of records eligible for detailed analysis. We, therefore, did not want to provide estimates of certainty for the frequency of the problem. Yet, the discrepant findings even in these small samples demonstrate a clear disconnect between reporting expectations and reporting practices.

Something odd has happened in the denominator of Table 2, Case 6, Arm 1, Publication column – it drops from 2563 to 1262 – is this a typographical error?

Response: We have corrected this error. Thank you.

Are the authors planning to put the dataset into a public repository?

Response: We will leave this to the discretion of the editor considering the pros and cons of an online appendix identifying records by their NCT numbers.

I think that this paper raises an important issue with the reporting of trials. However, the main limitations are the sample size (which may not be able to be increased further), and difficulty in deciding how the authors had decided when they could not determine a total death count.

Reviewer: Peter C Gøtzsche
The Nordic Cochrane Centre
No competing interests

I think there is something interesting in this paper that we can learn from, but it needs to be written much better and in a way that allows us to understand whether the problems are major or minor

without consulting the tables. It is really difficult to follow the flow of information in this paper, and the language is also difficult and sometimes inappropriate. I feel the senior author should have contributed more, as he is capable of writing far better than in this manuscript.

Abstract

Line 17: please write "whether" counts may be discordant (not "how" as you don't know whether you will find any and therefore cannot say how).

Response: We have revised the Abstract accordingly.

I have not seen the term "death counts" before and it feels pretty odd, like the US Republications' false allegations of government "death panels", please consider using another term, e.g. number of deaths, which is how we describe this.

Response: Thank you. We have changed "death counts" to "number of deaths" or "death numbers" throughout.

Please describe how many pairs you sampled right from the start.

Response: We have made the suggested edits to the Results section of the Abstract.

I do not understand how there can be pairs in the database, as a pair comes from an entry in the database and a published report, furthermore, if there was no information on deaths, how can then the information be discordant, particularly since some publications reported absence of death just as in the registry. This is too confusingly written.

Response: We specified that 'without information on deaths' means "without any information on number of deaths", i.e. death or mortality were not mentioned in the CT.gov trial record. There was discordance when the publication mentioned that there were no deaths (number of deaths = 0), or that there was some death(s). We included a definition of discordance in the methods. See additional line in the last paragraph on page 9.

Under Conclusion we hear about death count per arm, which was not what we heard about in Results, there must be consistency.

Response: Thank you but we have mentioned "total deaths per arm" in the Results section.

P4, line 8: "the discrepant reporting expectations for death," confusing construct, please revise.

Response: Please see the revisions to the Article focus.

Line 14: we now hear about all cause death, which is commonly called all cause mortality, have not heard about this before, only death counts. You need to be consistent.

Response: we have revised to simply "deaths".

Line 27-34: not clear, use much simpler language so that we can follow your thinking.

Response: Please see revisions to this first bullet.

Line 52: I cannot recall you have defined what you mean by reporting expectations?

Response: The reporting expectations are those by the FDA, ClinicalTrials.gov, and CONSORT.

There is far too much in the Article Summary, page 4, and it continues on p 5, please abbreviate considerably. The first limitation on p 5 is not really a limitation, and not relevant for the Summary; it can be mentioned in the Discussion.

Response: We deleted the first limitation on page 5.

Several places: I do not like calling people or participants for subjects or individuals, seems a bit denigrating to me.

Response: We revised accordingly.

P6 line 14: the idea of explicitly stating even if no deaths occurred is a good one although it may seem over the top in many cases, e.g. in a small migraine trial. But we so often wonder whether no deaths occurred when nothing is mentioned about this that I believe it would be better if all papers involving patients needed to say that no one died if that is the case.

Response: Thank you. We agree.

Line 21: I am highly surprised that there are no regulations mandating the reporting of all deaths in a trial, I would have expected there were, at least for drug trials. Are you absolutely sure about this? In your text below you say that death is a serious adverse event and that it therefore must always be reported in drug trials. This appears inconsistent.

P7: again, I seriously doubt that if a drug company does not consider a death related to their drug they are not required to report all deaths to FDA.

Response (to the 2 comments above): We appreciate your query. We contacted the FDA and obtained additional information on regulations which we have now incorporated into the third paragraph of the introduction.

P7: I was really amused to read that “deaths from serious adverse events would presumably be a reason for not completing a trial.” I am pretty sure that dead patients cannot complete a trial! I am not really sure what the idea is that is described in lines 32-34, does it matter that deaths are only reported in the serious adverse events section, and not as a reason for discontinuation?

Response: A death is not always a reason for non-completion, (e.g. when death is an outcome, or in survival analyses where deaths may be censored). Regardless, we believe a death should at minimum be accounted for in the participant flow, if not also in results or adverse events.

P1, line 36: what is “a result for death in another metric?” Metric has to do with measuring, e.g. kg, and a death is a death so I think your use of terms is not optimal.

Response: See additional examples in the second paragraph of the Results.

Line 58: I have forgotten whether the 27 pairs where randomly selected from all the pairs, I therefore went back to the top of P9 where I could see that you selected 75 matching publications of which 27 corresponded to database records of reported deaths. At this point, it is not exactly clear to me how the study was done. It may be my fault, but I am usually able to understand what I read and I am now a bit confused about the methods.

Response: Please see revisions to the third paragraph of the Methods.

P11, lines 10-20: I wonder why concordance could not be assessed in 12 of the 27 pairs when they were carefully selected in order to have number of deaths per arm for both trial database and publication.

Response: The 27 records were selected because they contained some number for deaths in at least one data module. However, in 12 pairs we could not compare the counts, because either the record or the publication did not allow unambiguous determination of the total number of deaths per arm.

I also miss some information about what was different, e.g. was it one or two deaths out of many or what?

Response: Please see table 2.

Lines 25-40: I would not call it discordant when the database does not say anything about deaths and the publication says that were no deaths.

Response: We chose to count this as discordant given the difference in certainty about death counts between absence of any information and confirmation of number of deaths equal to zero.

P12, line 14: so, when treatment arms were missing, did you call this concordant or discordant? I suggest that you focus on the really important issues, e.g. P14, line 31 to 36, it is very serious when it is not possible to judge how many died in a trial in the two treatment arms because of ambiguity.

Response: We agree that this is an important finding. We stated this as the first sentence of the second paragraph in the discussion.

I also suggest the authors write something about that just a couple of deaths may bias the trial considerably and if it happens in several trials, we can get it very wrong.

Response: See addition to the sixth paragraph in the Discussion.

A good case in point is not about death but about the three missing cases of myocardial infarction in the infamous Vigor trial published in NEJM in 2000. These myocardial infarctions were deleted very shortly before final submission of the manuscript and it changed the result from significant harm to no significant difference. That fraud cost the lives of thousands of patients.

Response: Thank you. We cited the editorial by Dr Dazen on this case study. See ref 16.

P15, line 23: please be careful with your choice of words, “death reporting may never be complete or simple given the challenges in ascertainment and adjudication.”

Response: We deleted this sentence.

I always favour total number of deaths whatever the cause, which I suggest you also write about here as we know that assessment of cause death can be biased, sometimes to a considerable degree. So do not mix up total number of deaths, which you describe somewhere in your paper and cause specific mortality.

Response: We agree. This is stated in the new sentence added to the sixth paragraph of the

Discussion.

P15, line 45: you say here “In summary, our study shows lack of clarity…..” Forgive me for saying this, but this is exactly the problem with your study, that the way it has been written up is not sufficiently clear (although this was not what you meant in this sentence of course).

Response: We hope it is clearer now. Thank you for your comments.

One of your references is to the paper about better reporting of harms according to CONSORT. I therefore wonder why you do not quote this paper specifically, as it says under recommendation 6 that authors should always report deaths in each study group during a trial, regardless of whether death is an endpoint and regardless of whether attributions to a specific cause is possible. I believe you should quote this.

Response: Thank you for this comment. We have included this reference in the introduction. Please see revision.

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Prayle, Andrew<br>University of Nottingham, Division of Child Health |
| **REVIEW RETURNED** | 09-Nov-2012 |

| | |
|---|---|
| **RESULTS & CONCLUSIONS** | The sample sizes remain small. I take the point well that attrition of studies reduced the sample size, but I am pretty certain that papers exist for more than 27 out of 123 records (figure 1 right hand arm). It’s a shame that a manual search in addition to the one done by ClinicalTrials.gov has not been done. This would have easily added value to the conclusions.<br>The authors essentially want to make the point that some records have discordant reporting of death when comparing ClinicalTrials.gov and linked publications. I still think that a measure of how common this is would be helpful. However, I do think that the sample size is very small to make robust conclusions. 27 records which did not list any deaths on ClinicalTrials.gov were sampled, and of these 5 of the papers give evidence of mortality within the trial. A 95% confidence interval for this proportion is approximately 6 – 32%.<br>If I was reading a ClinicalTrials.gov record, and it did not list any deaths, and was told that in this scenario 6% of records deaths actually occurred in the trial but are not reported in the summary results, I would be far less worried about unreported deaths than if I was told that it was 32%. Whilst we aspire to have complete integrity of data, in practice this is not likely to be achieved, and some estimate of how often deaths are missing from the ClinicalTrials.gov record would be useful. Similar arguments can be made for the scenario of a paper not listing deaths, but a ClinicalTrial.gov record recording deaths.<br>The authors don’t want to release the data. I could identify the trials from the data in the tables by restricting the search on ClinicalTrials.gov to studies with results and using key words, and checking for the numbers reported in the results (e.g. to identify the last trial in table 2 search for the keyword “penile” and restrict the search to trials with result, only 6 trials are retrieved and the first one in the list is currently the trial in table 2). In practice, table 2 tells you all you need to identify these trials anyway, it just gives you more work to do it. |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed one of my points by including appendix 3, and have re-written much of the methods. I think that this is now somewhat clearer. The casual reader will still probably give up if this paper is not in their area of interest.<br>Introduction: p6 line 49 the rules for which studies are required to upload results is actually more complex than this.<br>Methods: p9 line 49 – 56. I don't think that there is a requirement in ClinicalTrials.gov (and I have re-read the reporting guidance on this) to state "no deaths" when no deaths have occurred, but I can see why this may then be reported in the paper as no deaths. Therefore, if a ClinicalTrials.gov record doesn't list any deaths, and the paper says no deaths, I think that this in actual fact is concordant rather than discordant. In any case, none of the discordant cases in table 2 fall into this category.<br><br>In summary, only one of my points has been addressed by the authors. The authors make cogent arguments as to why they don't need to address the points of small sample size, and confidence intervals. I will let the editors decide if addressing these points would improve the paper. |

## VERSION 2 – AUTHOR RESPONSE

Reviewer: Andrew Prayle
COI: I have previously published data from ClinicalTrials.gov.

The sample sizes remain small. I take the point well that attrition of studies reduced the sample size, but I am pretty certain that papers exist for more than 27 out of 123 records (figure 1 right hand arm). It's a shame that a manual search in addition to the one done by ClinicalTrials.gov has not been done. This would have easily added value to the conclusions.
The authors essentially want to make the point that some records have discordant reporting of death when comparing ClinicalTrials.gov and linked publications. I still think that a measure of how common this is would be helpful. However, I do think that the sample size is very small to make robust conclusions. 27 records which did not list any deaths on ClinicalTrials.gov were sampled, and of these 5 of the papers give evidence of mortality within the trial. A 95% confidence interval for this proportion is approximately 6 – 32%.
If I was reading a ClinicalTrials.gov record, and it did not list any deaths, and was told that in this scenario 6% of records deaths actually occurred in the trial but are not reported in the summary results, I would be far less worried about unreported deaths than if I was told that it was 32%. Whilst we aspire to have complete integrity of data, in practice this is not likely to be achieved, and some estimate of how often deaths are missing from the ClinicalTrials.gov record would be useful. Similar arguments can be made for the scenario of a paper not listing deaths, but a ClinicalTrial.gov record recording deaths.

Response: Please see above response to comments from the managing editor. We emphasized that the cases with discordant numbers are illustrative rather than generalisable. However, even if the percentage of discordant pairs was low, this may include cases with small discrepancies or large discrepancies (see Table 1. Cases 4 and 9).

The authors don't want to release the data. I could identify the trials from the data in the tables by restricting the search on ClinicalTrials.gov to studies with results and using key words, and checking

for the numbers reported in the results (e.g. to identify the last trial in table 2 search for the keyword “penile” and restrict the search to trials with result, only 6 trials are retrieved and the first one in the list is currently the trial in table 2). In practice, table 2 tells you all you need to identify these trials anyway, it just gives you more work to do it.

Response: The PMIDs and the NCT numbers are now included in Tables 1 and 2.

The authors have addressed one of my points by including appendix 3, and have re-written much of the methods. I think that this is now somewhat clearer. The casual reader will still probably give up if this paper is not in their area of interest.

Response: Thank you for your comments which helped improve the clarity of the manuscript. We hope our manuscript will stimulate action to improve reporting of deaths in clinical trials going forward.

Introduction: p6 line 49 the rules for which studies are required to upload results is actually more complex than this.

Response: The rules for which studies are required to report results to clinical trials.gov were expanded.

Methods: p9 line 49 – 56. I don’t think that there is a requirement in ClinicalTrials.gov (and I have re-read the reporting guidance on this) to state “no deaths” when no deaths have occurred, but I can see why this may then be reported in the paper as no deaths. Therefore, if a ClinicalTrials.gov record doesn’t list any deaths, and the paper says no deaths, I think that this in actual fact is concordant rather than discordant. In any case, none of the discordant cases in table 2 fall into this category.

Response: As the reviewer states, at the present, it is not a requirement for results reporting in ClinicalTrials.gov to explicitly state the absence of deaths. If neither ClinicalTrials.gov nor publication contained information on deaths, we considered them concordant in terms of being ambiguous. However, we would have considered a case to be discordant if the record contained no information on deaths, but the publication stated that there were no deaths. As the reviewer points out, none of the pairs included in our table fall into this category. Still, we believe our decision to not imply absence of deaths if there was no direct information on deaths, is supported by the finding of pairs where the record does not contain information on deaths, but the publication does report deaths. We hope that our manuscript will stimulate better reporting of deaths to avoid this ambiguity.

In summary, only one of my points has been addressed by the authors. The authors make cogent arguments as to why they don’t need to address the points of small sample size, and confidence intervals. I will let the editors decide if addressing these points would improve the paper.

Response: We very much appreciate your thorough and thoughtful review.