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)	Discrepancies between ClinicalTrials.gov Recruitment Status and Actual Trial Status: a Cross-Sectional Analysis
AUTHORS	Jones, Christopher; Safferman, Michelle; Adams, Amanda; Platts-Mills, Timothy

VERSION 1 – REVIEW

REVIEWER	Roberta Scherer Johns Hopkins Bloomberg School of Public Health
REVIEW RETURNED	09-Jun-2017

GENERAL COMMENTS	The author present a study comparing the trial end date with the date that the investigators recorded the end date in ClinicalTrials.gov for a random sample of interventional phase 2-4 trials. They also searched for publications of primary results or conference abstracts for trials with a status indicating ongoing activity in the trial registry, but that had listed an anticipated study end date prior to January 2016. The rationale for conducting this study is sound in that the designation of a trial as completed in ClinicalTrials.gov may be misleading, and it is important to know the extent of the problem. 1.Maybe I'm missing something, but the number of trials retrospectively registered and the number is the table does not seem to match. Given that the authors said they looked for trials "studies registered between 01/01/2012 and 12/31/2015", it seems that the majority of trials with a start date prior to 01/01/2012 would be considered retrospectively registered (using the definition that a trial is considered to be retrospectively registered if the registration date was one month or more past the study start). However, the number of trials in Table 1 with a start date before 2012 was 287 (15+37+235) but the number retrospectively registered was only 159. There may be a simple explanation for this. Perhaps "studies registered between 01/01/2012 and 12/31/2015" mean trials that were included in the register during this time period instead of being the dates being when the trials? If so, this should be clarified. 2.It would be helpful to see if the proportion of trials registered retrospectively has changed over time, as more investigators became familiar with the process. That is, is the proportion of trial registered retrospectively decreasing? Certainly there was a bolus of trials registered in 2005- 2006 following the ICMJE requirement for registration prior to publication of a trial that may have started years before and one would expect a large proportion of these to have been retrospectively registered. Are trials more recently registered (2012-2015) less likely to have been retrospectively registered?

	It would be useful to have some 'cut-off' date for investigators doing studies that rely on 'completed' trials, or know the error rate within a certain time frame. 3.Neither the mean nor median were useful metrics to use to show the distribution of delay from between study end and the date this event was recorded in ClinicalTrials.gov. Mean is used for normal distribution and median when the distribution is skewed, but tends towards normal distribution. The data, as shown in the figure, has a definitie pattern. Only reporting the mean or median is somewhat misleading, It would be more useful to know what proportion of trials recorded the study end within a reasonable length of time (2-3 months) and could be considered 'compliant' with the requirement. One also wonders if the trials with exceptionally long delays were the trials registered shortly after the ICMJE statement was implemented. That is, were these early trials the ones with both retrospective registration and exceptionally long intervals between study start and registration? 4.When searching for publications, was the ClinicalTrials.gov site itself searched? Staff at ClinicalTrials.gov have indicated that they regularly search Medline for publications with an NCT register number in the abstract, so this would have been a way to identify publications more easily.

REVIEWER	Daniel Hartung OSU/OHSU Portland, OR USA
REVIEW RETURNED	21-Jun-2017

GENERAL COMMENTS	This is an interesting paper that summarizes discrepancies in trial status registered in clinicaltrials.gov. Findings from the study suggest that a non-trivial number of studies are mislabeled with respect to being completed and need to be updated. This is important information to know for systematic reviewers or others who are using ClinicalTrials.gov to identify completed trials in an efficient manner. It is also potential important for others trying to identify ongoing studies. While the paper is generally well written and the analyses seem sound, the authors need to improve the clarity of some of their descriptions. I also have some recommendations to improve the translation of their results. Major suggestions 1) Although frequent users are likely to understand the various ways ClinicalTrials.gov can be queried, others may not. It would be helpful to provide a basic description of the query function – specifically as it relates to Status. This will also aid readers in understanding where you abstracted certain data elements. This will help clarify the last sentence of the background which states "The objective of this study is to quantify delays observed between the end of enrollment in registered clinical trials, and the time that the registry entries are updated to reflect that enrollment has ended."

	This sentence implies there are more than one way (somewhere else in ClinicalTrials.gov) to determine if enrollment has ended. I recommend you be more carefully in what you describe as enrollment status, recruitment, etc. 2) My understanding of your analysis was that you broadly categorized trials as either “concluded” which you define as either completed or terminated and those that were “potentially ongoing”, defined as recruiting, enrolling by invitations, active not recruiting, suspended, or unknown. For trials that were concluded, where did you abstract the information about when the trial was updated? This relates to issue #1 above. It would be helpful if you generated this distinction within Table 1 with two additional columns. I would also recommend making edits to figure 1 (CONSORT) to reflect these divisions. Perhaps use two boxes (n=313, n=92) with those two titles, then clarify how many were listed as completed, terminated, ongoing, other.etc. within those boxes or sub-boxes. Other comments -Page 2, Line 44: says trials listed completed or stopped early, but I think you mean completed or terminated? Consistent terminology is important. -Page 5, Line 94: What was 500 trial sample based on?; if 500 was your target sample, why not continue to sample (randomly) until that target was reached instead of reducing the sample through exclusions. -Page 5, line 105: “We recorded the dates on which the registry entries were updated to reflect that trials had concluded.” – where is this data abstracted from? -Page 6, line 112: “For studies which were scheduled to have been completed prior to January 2016 and did not have an updated recruitment status indicating that they had concluded, we performed a comprehensive literature search to identify published evidence that the trial might in fact have been completed.” How did you determine when scheduled completion date was? In the results, you do not report the number of ongoing trials that were scheduled to be completed prior to 2016. Were all 92 ongoing studies scheduled to complete prior to January 2016? -Page 8, 169: your analysis focuses on trials with a “major discrepancy” which you define as incorrect recruitment status or delayed by more than a year between completion and recruitment status update. I would include this as row descriptor in table 1. -Page 9, page 186: “efforts should be made to confirm the enrollment status..” How does one do this? -you present some findings based on whether or not the trial was retrospectively registered, but do not really discuss the implications of this finding. You need to provide some discussion of this. Is it that retrospectively registered trials are typically more out of date than prospectively registered trials?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Roberta Scherer
Institution and Country: Johns Hopkins Bloomberg School of Public Health
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Comment: The author present a study comparing the trial end date with the date that the investigators recorded the end date in ClinicalTrials.gov for a random sample of interventional phase 2-4 trials. They also searched for publications of primary results or conference abstracts for trials with a status indicating ongoing activity in the trial registry, but that had listed an anticipated study end date prior to January 2016. The rationale for conducting this study is sound in that the designation of a trial as completed in ClinicalTrials.gov may be misleading, and it is important to know the extent of the problem.

Comment 1.Maybe I'm missing something, but the number of trials retrospectively registered and the number is the table does not seem to match. Given that the authors said they looked for trials "studies registered between 01/01/2012 and 12/31/2015", it seems that the majority of trials with a start date prior to 01/01/2012 would be considered retrospectively registered (using the definition that a trial is considered to be retrospectively registered if the registration date was one month or more past the study start). However, the number of trials in Table 1 with a start date before 2012 was 287 (15+37+235) but the number retrospectively registered was only 159. There may be a simple explanation for this. Perhaps "studies registered between 01/01/2012 and 12/31/2015" mean trials that were included in the register during this time period instead of being the dates being when the trials? If so, this should be clarified.

Response: Thank you very much for identifying this discrepancy. The initial submission contained an error: the included trials were randomly selected from studies registered between 01/01/2012 and 12/31/2012, rather than 2012-2015. These earlier dates were selected to allow a reasonable amount of time for prospectively registered trials to have been completed and published. We have corrected this error.

Comment 2.It would be helpful to see if the proportion of trials registered retrospectively has changed over time, as more investigators became familiar with the process. That is, is the proportion of trial registered retrospectively decreasing? Certainly there was a bolus of trials registered in 2005- 2006 following the ICMJE requirement for registration prior to publication of a trial that may have started years before and one would expect a large proportion of these to have been retrospectively registered. Are trials more recently registered (2012-2015) less likely to have been retrospectively registered? It would be useful to have some 'cut-off' date for investigators doing studies that rely on 'completed' trials, or know the error rate within a certain time frame.

Response: Thank you for this suggestion. We agree that knowing the error rate with respect to trial recruitment status within a given time frame would be potentially useful. Trials which started enrollment before 2006 were very likely to have major discrepancies in the listed recruitment status. Major discrepancies were less common among more recently initiated trials, though even among trials initiated after 2012 approximately one quarter had major discrepancies between the listed recruitment status and the actual trial status. We have updated the results section to include this information.

Comment 3. Neither the mean nor median were useful metrics to use to show the distribution of delay from between study end and the date this event was recorded in ClinicalTrials.gov. Mean is used for normal distribution and median when the distribution is skewed, but tends towards normal distribution. The data, as shown in the figure, has a definite pattern. Only reporting the mean or median is somewhat misleading. It would be more useful to know what proportion of trials recorded the study end within a reasonable length of time (2-3 months) and could be considered 'compliant' with the requirement. One also wonders if the trials with exceptionally long delays were the trials registered shortly after the ICMJE statement was implemented. That is, were these early trials the ones with both retrospective registration and exceptionally long intervals between study start and registration?

Response: Thank you for these suggestions as well. We agree that providing additional information about the distribution of our results is warranted, and have updated the results section as suggested to describe the proportion of trials which recorded the study end within 3 months. We also compared delays between prospectively and retrospectively registered trials, including trials with exceptionally long delays between starting enrollment and registration, and found that retrospectively registered trials were more likely to have substantial delays in updating their recruitment status. These results have been added to the Results, and have been addressed within the Discussion section.

Comment 4.When searching for publications, was the ClinicalTrials.gov site itself searched? Staff at ClinicalTrials.gov have indicated that they regularly search Medline for publications with an NCT register number in the abstract, so this would have been a way to identify publications more easily.

Response: We did review each relevant ClinicalTrials.gov entry as part of our search protocol, and have updated the Methods section to reflect this.

Reviewer: 2
Reviewer Name: Daniel Hartung
Institution and Country: OSU/OHSU, Portland, OR, USA
Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below

Comment: This is an interesting paper that summarizes discrepancies in trial status registered in clinicaltrials.gov. Findings from the study suggest that a non-trivial number of studies are mislabeled with respect to being completed and need to be updated. This is important information to know for systematic reviewers or others who are using ClinicalTrials.gov to identify completed trials in an efficient manner. It is also potential important for others trying to identify ongoing studies.

Response: While the paper is generally well written and the analyses seem sound, the authors need to improve the clarity of some of their descriptions. I also have some recommendations to improve the translation of their results.

Major suggestions
1) Although frequent users are likely to understand the various ways ClinicalTrials.gov can be queried, others may not. It would be helpful to provide a basic description of the query function – specifically as it relates to Status. This will also aid readers in understanding where you abstracted certain data elements.

This will help clarify the last sentence of the background which states "The objective of this study is to quantify delays observed between the end of enrollment in registered clinical trials, and the time that the registry entries are updated to reflect that enrollment has ended." This sentence implies there are

more than one way (somewhere else in ClinicalTrials.gov) to determine if enrollment has ended. I recommend you be more carefully in what you describe as enrollment status, recruitment, etc.

Response: Thank you for this suggestion. We have added the following language to our introduction in order to clarify how the query function relates to enrollment status: “Within ClinicalTrials.gov, users have the option of utilizing the “advanced search” function to restrict search results to only those trials with a particular enrollment status (ie. Not yet recruiting, recruiting, completed, etc.).“ We have also revised the entire manuscript to ensure that we are using the terms enrollment, recruitment, etc in a consistent manner.

2) My understanding of your analysis was that you broadly categorized trials as either “concluded” which you define as either completed or terminated and those that were “potentially ongoing”, defined as recruiting, enrolling by invitations, active not recruiting, suspended, or unknown. For trials that were concluded, where did you abstract the information about when the trial was updated? This relates to issue #1 above. It would be helpful if you generated this distinction within Table 1 with two additional columns. I would also recommend making edits to figure 1 (CONSORT) to reflect these divisions. Perhaps use two boxes (n=313, n=92) with those two titles, then clarify how many were listed as completed, terminated, ongoing, other.etc. within those boxes or sub-boxes.

Response: Thank you for identifying this issue. We have edited the Methods section to clarify that updates to the recruitment status were determined through a review of the “History of Changes” section within ClinicalTrials.gov. We have also added the requested information to both Table 1 and Figure 1.

Other comments

-Page 2, Line 44: says trials listed completed or stopped early, but I think you mean completed or terminated? Consistent terminology is important.

Response: We have corrected this, thank you.

-Page 5, Line 94: What was 500 trial sample based on?; if 500 was your target sample, why not continue to sample (randomly) until that target was reached instead of reducing the sample through exclusions.

Response: Because this work was unfunded, the sample size was largely limited by the time the investigators were able to devote to the project. Based on a small pilot phase, we estimated that an initial sample of 500 registry entries would result in the inclusion of about 400 trials, which was both manageable given the resources we had available, and large enough to provide relatively stable estimates.

-Page 5, line 105: “We recorded the dates on which the registry entries were updated to reflect that trials had concluded.” – where is this data abstracted from?

Response: This information was abstracted from the ClinicalTrials.gov History of Changes section within each registry entry. We have updated the section in question to clarify this.

-Page 6, line 112: “For studies which were scheduled to have been completed prior to January 2016 and did not have an updated recruitment status indicating that they had concluded, we performed a comprehensive literature search to identify published evidence that the trial might in fact have been completed.”

How did you determine when scheduled completion date was? In the results, you do not report the number of ongoing trials that were scheduled to be completed prior to 2016. Were all 92 ongoing studies scheduled to complete prior to January 2016?

Response: Thank you very much for identifying these issues. The scheduled completion date was abstracted from the "Estimated Primary Completion Date" field within ClinicalTrials.gov. We have revised the Methods section to clarify this. We excluded potentially ongoing trials with an estimated primary completion date after 1/1/16, as even if these trials had been completed we thought that it would be likely that publications would not yet be available, and that therefore we would bias our sample by including very recently concluded studies. Therefore all 92 of the included ongoing studies were scheduled to have a primary completion date before 2016. This information was available in Figure 1, but should have been more clearly discussed within the Methods section. We have updated the Methods to explicitly address this.

-Page 8, 169: your analysis focuses on trials with a "major discrepancy" which you define as incorrect recruitment status or delayed by more than a year between completion and recruitment status update. I would include this as row descriptor in table 1.

Response: Thank your for this suggestion; we have included this information in Table 1.

-Page 9, page 186: "efforts should be made to confirm the enrollment status.." How does one do this?

Response: We agree that the Discussion should expand on this point, and have amended it accordingly.

-you present some findings based on whether or not the trial was retrospectively registered, but do not really discuss the implications of this finding. You need to provide some discussion of this. Is it that retrospectively registered trials are typically more out of date than prospectively registered trials?

Response: Thank you for this suggestion. We have added a brief discussion of this issue to our Discussion section.

VERSION 2 – REVIEW

REVIEWER	Roberta Scherer Johns Hopkins Bloomberg School of Public Health United States of America
REVIEW RETURNED	30-Aug-2017

GENERAL COMMENTS	A previous comments have been adequately addressed. There are some within the clinical trial community who object to the use of 'subject' to describe study participants. While it is not necessary to revise the manuscript, it should be kept in mind.