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)	Early termination of cardiovascular trials as a consequence of poor accrual: analysis of ClinicalTrials.gov 2006-2015
AUTHORS	Baldi, Ileana; Lanera, Corrado; Berchialla, Paola; Gregori, Dario

VERSION 1 - REVIEW

REVIEWER	Sripal Bangalore New York University School of Medicine, USA
REVIEW RETURNED	14-Sep-2016

GENERAL COMMENTS	The authors assessed geographical and temporal patterns of early termination due to poor accrual in cardiovascular trials by searching the AACT database. The study found that the proportion of early termination of trials varied from Europe and America but they found a positive effect of globalization of trials in reducing the risk of early termination due to poor accrual. The study is important. The limitations are well recognized but are important. I have the following comments: 1. The authors only choose to search the ClinicalTrials.gov which is mandatory for trials originating in the US. Although non US trials, are also registered, there is no mandatory requirement for such trials. It is possible that smaller non US trials are never registered thus missing potential early termination of non US trial. This can create bias. Please comment. 2. It would be interesting to see if among the terminated trials whether the results were positive or showed no difference and categorized by early termination vs. not 3. Although the authors report rates between continents, it is not clear if these differences are statistically significant. In addition, 95% CI around these reports should be reported. 4. Figure 2 and 3. Please report p values for trend and for the difference.

REVIEWER	Neal Dickert Emory University School of Medicine
REVIEW RETURNED	07-Dec-2016

GENERAL COMMENTS	This study represents an interesting attempt to harness the ClinicalTrials.gov database to look at study performance and predictors of trial termination. Study termination is a topic of real importance for which there are insufficient data. In this respect, the authors are to be commended. In addition, while the writing is not always clear (see last sentence p. 4 as an example), the results are reasonably clearly presented. My principal concerns with the manuscript have to do with the interpretation of the findings and

	somewhat limited analysis. First, 11% of closed studies were determined to have been terminated. This really does not seem like a large number., especially in contrast to commonly cited figures of far higher rates of non-completion in oncology, for example. In this respect, the authors seem to overstate the claim that these data provide evidence of a significant problem. Second, and similarly, the authors suggest that the “heterogeiety” in termination of studies prior to completion between the Americas and Europe is substantial and meaningful, but it is really quite modest (14 vs 9.6%). Third, the authors spend some time discussing the relationship between a trial being multi-national and its likelihood of being terminated. They imply that this relationship is likely causal, but this assumption seems unwarranted. From a technical perspective, one certainly cannot infer causality from these kinds of data. More importantly, it seems most likely the case that multi-national projects just have greater investment from sponsors and are "bigger deals" so are likely to have the resources for recruitment, etc. Fourth, there is relatively limited analysis of the available data presented. For example, while the distribution of the types of studies that have been terminated are reported, the percentage of studies representing a particular type that wind up being terminated are not. It would be meaningful to know, for example, whether X% of device studies are terminated rather than simply that Y% of studies that are terminated are device studies. The data source should allow for this. Finally, there are numerous potential drivers of failed recruitment (study wasn't attractive to patients, recruitment efforts were insufficient, competing studies, inadequate budgets, etc.). To make this analysis more valuable, it would be interesting if the authors could probe more in-depth regarding what some of those drivers tend to be. This would greatly enhance the practical value of this kind of project. In the end, it is not clear precisely how much the study really expands regarding our understanding of early termination. Minor points: This is addressable, but the writing is somewhat difficult to follow. See sentence at the end of p. 4 as one example. The MeSH headings described are general. The authors might clarify that more specific headings are subsumed under these so that the search captured all CV trials.

REVIEWER	Brian Claggett Brigham and Women's Hospital USA
REVIEW RETURNED	15-Mar-2017

GENERAL COMMENTS	The statistics presented are fairly minimal. This may have been a conscious decision on the authors' behalf. However, having access to the data they have collected, it would seem to be interesting to

	build a multivariable logistic regression model to identify the characteristic(s) most associated with termination and accrual-related termination. I imagine that number of site, number of patients, number of continents, etc. will all be related to one another, so trying to disentangle which factors are the most important would be very interesting, I believe. Finally, for statements like the following: "intercontinental trials exhibited higher figures of termination . . . . as compared to intracontinental trials (13% vs 11%)", it would be worth knowing whether those two percentages are actually significantly different. I would consider these to be qualitatively similar, but if there really is justification to call one of these numbers higher than other, I am ok with using that language.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

The authors assessed geographical and temporal patterns of early termination due to poor accrual in cardiovascular trials by searching the AACT database. The study found that the proportion of early termination of trials varied from Europe and America but they found a positive effect of globalization of trials in reducing the risk of early termination due to poor accrual. The study is important. The limitations are well recognized but are important.

I have the following comments:
1. The authors only choose to search the ClinicalTrials.gov which is mandatory for trials originating in the US. Although non US trials, are also registered, there is no mandatory requirement for such trials. It is possible that smaller non US trials are never registered thus missing potential early termination of non US trial. This can create bias. Please comment.

Authors' response: We agree with the reviewer. From a recent investigation (Viergever RF and Li, BMJOpen 2015) on trends in global clinical trial registration from 2005 to 2013 emerged that registered trials conducted in Northern America and Latin America and Caribbean were almost exclusively registered in ClinicalTrials.gov. Also European trials were predominantly registered in ClinicalTrials.gov rather than in the EU Clinical Trials Register. A potential selection bias can arise when considering trials conducted in Oceania and Asia since they were increasingly registered at other regional registries such as the Australian New Zealand Clinical Trials Registry and the Japan Primary Registries Network, respectively. Therefore, the extent of early termination for trials run in these two continents may be underestimated and this is probably the case of Asia with 5.6%. We have acknowledged this potential bias among study limitations.

2. It would be interesting to see if among the terminated trials whether the results were positive or showed no difference and categorized by early termination vs. not

Authors' response: It would be definitely interesting but the analysis of primary outcome data is not straightforward. Unlike the evaluation of outcome data availability (as in Williams RJ et al, PlosOne 2015), the comparison of findings with the set of hypotheses around which the study was powered (needed to assess whether the trial is “positive” or “inconclusive”) is not automatic and may require a manual review of records. Furthermore, previous studies have documented suboptimal rates of results reporting in ClinicalTrials.gov. (Chen et al, BMJ 2016). Therefore, outcome data should be searched for on ClinicalTrials.gov and also in the published literature.

Undertaking such an extensive deepening is beyond the purpose of the present study.

3. Although the authors report rates between continents, it is not clear if these differences are statistically significant. In addition, 95% CI around these reports should be reported.

Authors' response: We added inferential results in the Results section.

4. Figure 2 and 3. Please report p values for trend and for the difference.

Authors' response: In figure 4 (former figure 2) the proportion of early termination of trials recruiting from Asia (p-value < 0.001) and Europe (p-value < 0.001) is significantly different from that of trials recruiting from Americas and shows a slightly decreasing trend (p = 0.02).
In figure 5 (former figure 3) the effect of intercontinental trials is significant (p=0.001) and stable over time (p=0.21).

Reviewer: 2

This study represents an interesting attempt to harness the ClinicalTrials.gov database to look at study performance and predictors of trial termination. Study termination is a topic of real importance for which there are insufficient data. In this respect, the authors are to be commended. In addition, while the writing is not always clear (see last sentence p. 4 as an example), the results are reasonably clearly presented. My principal concerns with the manuscript have to do with the interpretation of the findings and somewhat limited analysis.

First, 11% of closed studies were determined to have been terminated. This really does not seem like a large number., especially in contrast to commonly cited figures of far higher rates of non-completion in oncology, for example. In this respect, the authors seem to overstate the claim that these data provide evidence of a significant problem.

Authors' response: Probably, greater attention has been paid to premature termination of trials in oncology as compared to other settings, but the problem is at least as relevant for cardiovascular trials.
Several studies have addressed the issue of early termination of oncology trials and their findings exhibit some variability depending on the operational definition of “termination” and the data sources used. With a definition similar to the one we adopted, Stensland and colleagues (Journal of the National Cancer Institute, 2014) estimated that 12% of phase II and phase III oncology trials, started between 2005 and 2011 and registered in ClinicalTrials.gov database, were terminated or withdrawn. If we limit our analysis to phase II and phase III cardiovascular trials, we get a similar figure (14%).

Second, and similarly, the authors suggest that the “heterogeiety” in termination of studies prior to completion between the Americas and Europe is substantial and meaningful, but it is really quite modest (14 vs 9.6%).

Authors' response: We modified the sentence in the Results section to give less emphasis to this result.

Third, the authors spend some time discussing the relationship between a trial being multi-national and its likelihood of being terminated. They imply that this relationship is likely causal, but this assumption seems unwarranted. From a technical perspective, one certainly cannot infer causality from these kinds of data. More importantly, it seems most likely the case that multi-national projects just have greater investment from sponsors and are "bigger deals" so are likely to have the resources

for recruitment, etc.

Authors' response: We agree with the reviewer and we modified the text accordingly. To clarify the concept of globalization of cardiovascular clinical trials, we cited the example of heart failure trials (O'Connor, JACC Heart Fail, 2015) which have evolved into "mega trials" requiring an increasingly large number of patients, thus necessitating global enrolment. Such "mega trials" are exactly the "bigger deals" that the reviewer refers to.

Fourth, there is relatively limited analysis of the available data presented. For example, while the distribution of the types of studies that have been terminated are reported, the percentage of studies representing a particular type that wind up being terminated are not. It would be meaningful to know, for example, whether X% of device studies are terminated rather than simply that Y% of studies that are terminated are device studies. The data source should allow for this.

Authors' response: We agree with the reviewer and we would like to specify that the percentages shown in Table 1 are exactly the row percentages which the reviewer refers to. For example, 199 out of 1679 device trials are terminated (11.9% as it appears in Table 1).

Finally, there are numerous potential drivers of failed recruitment (study wasn't attractive to patients, recruitment efforts were insufficient, competing studies, inadequate budgets, etc.). To make this analysis more valuable, it would be interesting if the authors could probe more in-depth regarding what some of those drivers tend to be. This would greatly enhance the practical value of this kind of project.

Authors' response: The median number of characters used to fill in the reason for termination field is 42 (interquantile range=43). As an example, common statements for termination due to poor accrual are "Because the inclusion rate was lower than expected" (50 characters) or "This study was terminated early due to poor recruitment" (55 characters). Clearly, this explains why we made only a general description of drivers of early termination, distinguishing between poor accrual, lack of funding and sponsor decision and appropriate reasons for termination based on internal or external evidence of futility or lack of safety.
To better exploit all the available information, we provided a graphical visualization of text mining results through a network of relationships between (pre-processed) words appearing in the reason for termination text field.

In the end, it is not clear precisely how much the study really expands regarding our understanding of early termination.

Authors' response: We deepened the analysis by including new descriptive and inferential results and modified the text accordingly.

Minor points:
This is addressable, but the writing is somewhat difficult to follow. See sentence at the end of p. 4 as one example.
We performed English editing.

The MeSH headings described are general. The authors might clarify that more specific headings are subsumed under these so that the search captured all CV trials.
We modified the description of MeSH headings accordingly.

Reviewer: 3
Reviewer Name
Brian Claggett
Institution and Country

Brigham and Women's Hospital USA

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below

The statistics presented are fairly minimal. This may have been a conscious decision on the authors' behalf. However, having access to the data they have collected, it would seem to be interesting to build a multivariable logistic regression model to identify the characteristic(s) most associated with termination and accrual-related termination.
I imagine that number of site, number of patients, number of continents, etc. will all be related to one another, so trying to disentangle which factors are the most important would be very interesting, I believe.
We are grateful to the reviewer for this central comment. We are absolutely aware of the nature of the data source and of the perils of making statistical inference on such big data (i.e., with a large enough sample, almost any effect can cause an impressively small p-value and may be over-interpreted). This explains why we intentionally performed a descriptive analysis, omitting tests and p-values which could be wrongly interpreted as a “seal of validation” of (unsubstantiated) causality claims.
Though, we understand that this could make readers uncomfortable and we decided to comply with reviewers’ requests by adding some inferential results. In particular, we added a conditional inference tree model to detect associations between available study characteristics and early termination.

Finally, for statements like the following: "intercontinental trials exhibited higher figures of termination . . . . as compared to intracontinental trials (13% vs 11%)", it would be worth knowing whether those two percentages are actually significantly different. I would consider these to be qualitatively similar, but if there really is justification to call one of these numbers higher than other, I am ok with using that language.
As explained in the previous point, we added inferential results (including p-values) and modified the text accordingly.

VERSION 2 – REVIEW

REVIEWER	Sripal Bangalore New York University School of Medicine, NY, NY, USA
REVIEW RETURNED	30-Apr-2017

GENERAL COMMENTS	The authors have adequately answered all my concerns.

REVIEWER	Neal Dickert Emory University School of Medicine
REVIEW RETURNED	02-May-2017

GENERAL COMMENTS	This piece is a helpful contribution in quantifying the extent to which premature termination of clinical trials of cardiovascular disease occurs. As articulated below, the statements about the impact of intercontinental status on the primary outcome do still seem a bit

	overstated. One variable which would be particularly helpful is “planned sample size.” Is this not available? This seems particularly relevant to the inter-continental/intra-continental difference. If intercontinental studies are systematically different in terms of larger planned sample sizes (and likely larger budgets, etc.), then comparing completion rates between intra and intercontinental studies may be comparing very different kinds of studies. If a model controlling for number of sites and planned enrollment were presented, this would be more persuasive in that it would more truly isolate the effect of globalization. As articulated by another reviewer previously, the finding of a 13 vs 11 % difference in intra-continental versus intercontinental trials really does not seem “clinically” meaningful (at least not enough to highlight as one of the key findings), especially in the absence of more detailed information about potential qualitative differences in these trials. Additionally, the effect of globalization is heterogenous; Americas-Europe trials had relatively high termination, for example. In short, I do think this finding remains a bit overstated. Specifically, the implication that these data suggest that globalization is a potential solution to this problem doesn’t seem truly justified by the analysis provided. I defer to statistical reviewers, but the logic of choosing a conditional decision tree versus regression models is not clear to me, and it appears to substantially lessen the degree of covariate adjustment that can be performed. Very minor detail- p. 7, last few lines- This is hard to follow. Is it the case that the statement that “trials run in America and multinational trials run in Americas…showed different rates…” means that the type of trial had an effect on rate of early termination within these locations? This could be written more directly.

REVIEWER	Brian Claggett Brigham and Women's Hospital USA
REVIEW RETURNED	13-Apr-2017

GENERAL COMMENTS	Authors have addressed my comments.

VERSION 2 – AUTHOR RESPONSE

1. One variable which would be particularly helpful is “planned sample size.” Is this not available? This seems particularly relevant to the inter-continental/intra-continental difference. If intercontinental studies are systematically different in terms of larger planned sample sizes (and likely larger budgets, etc.), then comparing completion rates between intra and intercontinental studies may be comparing very different kinds of studies. If a model controlling for number of sites and planned enrollment were presented, this would be more persuasive in that it would more truly isolate the effect of globalization.
We definitely agree with the reviewer on the importance of conveying a correct message. The primary endpoint of our investigation is early termination due to poor accrual, as also stated in the title from the very beginning. We have already acknowledged that intercontinental trials likely are megatrials (In the Discussion “Clearly, intercontinental trials are often resource-intensive large-scale

randomised controlled clinical trials, recruiting thousands of patients from large numbers of trial sites (i.e., megatrials)") but it is reasonable that within trials terminated early, megatrials are less affected by poor accrual as the reason for termination and other organizational and financial issues challenge their completion.
As explained in the AACT data dictionary, the variable "Enrollment" contains the estimated total number of participants to be enrolled (for ongoing trials) or the actual total number of participants that are enrolled in the clinical study (for trials no longer recruiting).

2. As articulated by another reviewer previously, the finding of a 13 vs 11 % difference in intra-continental versus intercontinental trials really does not seem "clinically" meaningful (at least not enough to highlight as one of the key findings), especially in the absence of more detailed information about potential qualitative differences in these trials.
We have already addressed the point raised by the other reviewer by modifying the text in the following way "Intercontinental trials exhibited comparable figures of termination and lower figures of unsuccessful accrual as the reason for their early stopping, as compared to intracontinental trials (13% vs. 11%, p-value = 0.24, termination, of whom 28% vs. 44%, p-value = 0.002, due to poor accrual, respectively)".

3. Additionally, the effect of globalization is heterogenous; Americas-Europe trials had relatively high termination, for example. In short, I do think this finding remains a bit overstated. Specifically, the implication that these data suggest that globalization is a potential solution to this problem doesn't seem truly justified by the analysis provided.
We agree with the reviewer but we focused our attention on the difference in early termination due to poor accrual. In this case, 46.3% for trials run in America and decreased to 34.7% for trials run in bilateral collaboration with Europe.

4. I defer to statistical reviewers, but the logic of choosing a conditional decision tree versus regression models is not clear to me, and it appears to substantially lessen the degree of covariate adjustment that can be performed.
In the authors' experience with data modelling, conditional decision trees are better suited to the analysis of big data.

5. Very minor detail- p. 7, last few lines- This is hard to follow. Is it the case that the statement that "trials run in America and multinational trials run in Americas…showed different rates…" means that the type of trial had an effect on rate of early termination within these locations? This could be written more directly.
We modified the text accordingly.