The editors' comments are listed below:

1. Editors felt that your paper covered an important topic, with an interesting central point and call for more standardised and precised language when reporting harms from trials.

Thank you.

2. However, we felt that some aspects of the analysis were weak, particularly around defining certain terms (e.g. what is downplaying, what is acceptable risk) and how consensus was reached among the author group. Usually we'd expect such papers to have some standardisation of terms followed by blinded comparisons between different researchers looking at the same papers. We agree with reviewers points that the methods and data should be more transparently described.

Thank you for your comments. The reason we didn't have a standard and detailed Methods section was we thought it would diminish the spirit of "Analysis" type articles at the BMJ and look more like an original research paper. In the revised manuscript, we have added further details on how the trials were selected, data were extracted and the correctness of data were ensured.

3. We felt that some of the language was overly emotive and could be toned down (e.g. preposterous, absurd)

Thank you for this helpful comment. Looking back, we also felt the same and have tried to tone down our manuscript. We hope you will also find the revised manuscript appropriately toned.

4. The strategies to improve the situation could be developed further. What threshold for harm would be considered acceptable or who has the authority to make the call? How would defininig this with patients work in pratice?

We don't think we should come up with a threshold for considering any adverse effects as acceptable or tolerable. We argue that only the patients can decide whether any AE is acceptable. So, there cannot be any agreed upon threshold. As suggested by CONSORT, use of such vague terms should be avoided altogether. To make our arguments clear, we have added this paragraph in the Summary and Recommendations section:

*There is no threshold for harms that can be considered acceptable or tolerable. We health care providers cannot label any adverse effect as manageable or any safety profile as favorable. It is only up to the patients to decide whether any toxicities they suffered were acceptable. Labeling adverse effects from a drug as "acceptable" would only be acceptable if the patients who were enrolled were asked whether the toxicities were acceptable. If such data have not been collected, the use of vague terms such as "the*

*treatment was safe" should be avoided. The physicians and patients would decide whether the side effects from any treatment are tolerable or not on a case by case basis.*

5. You allege that triallists and authors knowingly mislead. Are you able to substantiate this?

We don't mean to imply that triallists and authors mislead the readers "knowingly". Downplaying of harms is common in trial reports, but whether this is intentional or an established culture is unknown. We have toned down the manuscript to avoid being interpreted as accusing the authors of "knowingly" misleading the readers. Our main objective with this paper is to bring light to this problem rather than blaming anyone.

6. We think that any resubmitted paper is likely to require a substantial amount of revision, with likely legal and statistical review from our end once it is returned. Given the likely extent of revision, and the fact that we are unable to make any promises of publication, you may wish to submit this paper elsewhere.

Thank you for this suggestion, but we believe the BMJ is the best venue to give voice to these concerns and given the time it has already taken at BMJ, a resubmission here would probably give us a faster chance of publication. Also, with regards to statistical review, we would like to emphasize that the meta-analysis is not the most important part of our paper and shouldn't take focus off the issue of reporting of harms. We acknowledge in the paper that this is likely an apple and orange meta-analysis as we pool adverse effects across different trials and drugs; our intention in doing so is not to provide an absolute number and claim the risk is higher by a given percent but that in general, the claims of lower or manageable risks in these trials are unsubstantiated.

Reviewer(s)' Comments to Author:

Reviewer: 1

Recommendation:

Comments:
This paper makes some important points in relation to the reporting of chemotherapy trials i.e. That the way that toxicity results are presented in high ranking journals does not follow CONSORT guidelines through the use of inappropriate vague terms and through arbitrarily suggesting toxicity is " acceptable" and " manageable " with no patient reported outcome as evidence and as a result , reports may give an inappropriately optimistic view of risk benefit ratio. The authors point out that this is particularly important given the high cost and relatively small clinical benefit of some new medicines . In order to make this case, They present data on adverse effects reporting together with associated " vague descriptions" from trials reported in 2016. This is an interesting approach to presenting data and I think will stimulate constructive debate.

Thank you. Indeed, our objective is to stimulate constructive debate.

I have only reviewed a few of the studies reported and I assume there is a brief statistical review for this paper.
I'm not completely convinced by some of the assumptions e.g.adverse effects are "falsely downplayed " if the incidence of adverse effects is more in the control arm than the treatment arm.
The authors , perhaps, fall into the same trap they have identified by using vague terms "downplay " and " falsely downplay " and if these are used They should be clearly defined.

Thank you for these important comments. We see the irony. We had defined apriori what would constitute as “downplaying” of harms and have mentioned that in the revised manuscript. In light of your and reviewer 2’s comment no.4, we have removed the mention of “false” or “true” with respect to downplaying, and focused instead on whether the toxicities were increased or decreased on such trials that used vague downplaying terms.

Currently the use of inflammatory language reduces the impact of this interesting report and I would suggest a review,restructure and shortening focussing on the key findings .

Thank you. We have revised the manuscript accordingly and hope that you find the revised manuscript more balanced.

Reviewer: 2

Recommendation:

Comments:
This paper assessed whether the appropriateness of using favorable toxicity terms in the reporting of cancer drug randomized controlled trials. The authors conclude that a large percentage of such trials (>40%) inappropriately used favorable terms or “downplayed” the harms.

The authors’ question has merit from both a scientific and clinical perspective. However, some sections of the paper are ‘editorial’ in nature and would benefit from a thorough display of all the data. Specific comments are provided below.

Thank you. The reviewer might have found some sections of our paper “editorial or viewpoint” like because we submitted this manuscript as Analysis article rather than original research.
1. The paper focuses on trials that downplayed the harms. But what about the trials that did not “downplay” the harms? How good was the reporting of their toxicity information? This information is key to convince readers that there is a difference between trials that “downplay” harms vs those that did not.

We didn’t collect data on trials that didn’t use the downplaying terms because our objective was not to show that trials that downplay the harms are different from those that did not, or that trials that downplay harms have more toxicities than others. It is not a comparison with other trials. We want to bring attention to the simple fact that many trials use such “downplaying” terms in their reports, that the use of such general terms might mislead the readers, is against the recommendation by CONSORT

and most of them actually don't decrease toxicities versus control contrary to what such terms (acceptable, tolerable, favorable etc.) might imply. Similar to other studies that looked into spin into trial reports (e.g. https://www.ncbi.nlm.nih.gov/pubmed/20501928 ) , we wanted to investigate the use of downplaying terms in the reporting of toxicities. We also believe that even if the toxicities were lower than the other trials or even lower than control arms in the same trials, such general terms cannot be used (eg. If fatal adverse event for a drug is 3% versus 5% for the control arm, the drug cannot still be said to be safe or well tolerated.)

2. The authors provide a list of terms that helped them identify trials that "downplayed harms". How was this list derived/constructed? Was it previously validated? As the entire analysis hinges on this list, it is crucial that additional information on its development be presented in detail.

We decided this list apriori based on our experience with reading cancer drug trial reports over the years. However, we also accepted to include any new similar terms that could imply downplaying of harms, if discovered during the study selection process. We didn't find any beyond that mentioned in the list. We have mentioned this in the revised manuscript.

3. Who extracted the data and made the determination of "downplaying harms"? If more than one author, how were discrepancies handled?

We have added a description of this in the revised manuscript.

4. The title of the section "Was the downplaying of harms true or false?" is a bit odd. The term "downplaying" already offers a judgment on what was reported. In that case, it should not matter if it is true or false? This section also suggests that 23% of trials that "downplayed" the harms were "true"? How do authors reconciliate these apparent divergent ideas?

Thank you for this important comment. We agree and have removed any mention of true or false. We have changed the title to "Were toxicities lower in the trials that downplayed the harms?"

5. The section on "What was the incidence of toxicities in trials that downplayed harms?" provides overall toxicities reported in the experimental vs control arms. However, these are presented as aggregates, i.e. combining the experimental arms (and control arms) of all trials. As presented, the data are less alarming – for example 50.6% toxicity with experimental drug vs 43.7% toxicity with the control drug. It would be more relevant to provide trial-specific data, as it is likely that some trials have larger differences.

We have now supplied all those data as supplementary files. As the reviewer suggested, certainly some trials have larger and some have smaller differences. However, we would like to highlight that focusing on exact percentages or specific trials might take the attention away from the agenda we want to highlight; which is not that the AEs are higher in these trials but that the use of such terms is not valid in any context, irrespective of the incidences unless a drug is being trialed for the very reason to prove it's safety against an active comparator (in a non inferiority trial for instance)

6. Inappropriate reporting of safety information is certainly a valid issue. However, this does not take into account the clinical experience of physicians with a given drug. Physician preference is also influenced by their experience with a given drug, and not solely based on how the safety information was initially reported. The discussion would merit some thoughts on this important point.

This is an excellent point, many thanks. We have added this discussion to our revised manuscript, in the paragraph preceding the title "Any measures to control such reporting practices?" Trial reports are more important in oncology because many new drugs are being approved at a rapid rate, and at least in the beginning, most of us had to rely on safety reporting in trial publications before getting "used to" using the drug. We have also added that safety information, and terms used to describe safety information, should take into account both physician and patient experiences. This is not possible when the trial is published, and hence it might be a better idea to avoid such general terms altogether in the spirit of CONSORT.

7. A major limitation of trials is that they are primarily designed to assess efficacy and not safety. While toxicity information collected during a trial is important, it is important to recognize that often, no firm conclusions can be derived based on the sample size of these trials. Given this context, at which point can one claim that the experimental drug is safe or unsafe?

Again an excellent suggestion, thank you. We have added to the discussion that trial reports aren't final and safety information keeps getting updated as more real world experience and data are available, and for that reason, it is all the more important not to label any safety signals as safe or tolerable etc. at the time of trial publication.