Title: Abstracts in high profile journals often fail to report harm.

Authors:

Enrique Bernal-Delgado (ebernal.iacs@aragon.es)
Elliott S Fisher (elliott.fisher@dartmouth.edu)

Version: 3 Date: 11 January 2008

Author's response to reviews: see over
Abstracts in high profile journals often fail to report harm

RESPONSE TO THE COMMENTS AND SUGGESTIONS BY THE REVIEWERS

Bernal-Delgado E MD PhD MPH¹, Fisher ES, MD MPH²

Institute for Health Sciences in Aragon (I+CS)¹, Foundation for Health Services Research¹ Center for the Evaluative Clinical Sciences, Dartmouth Medical School²

Enrique Bernal-Delgado, MD PhD MPH  (Corresponding author)
Institute for Health Sciences in Aragon (I+CS)
Zaragoza, Aragon, Spain
Address:
Avda Gómez Laguna 25,
50009 Zaragoza
Telephone: +34 976716792
Fax Number: +34 976714670
ebernal.iacs@aragon.es
REVIEWER #1

Major compulsory revisions

1. The fourth paragraph of the Results section appears to contradict the third paragraph. In the third paragraph it is stated that 75.9% of articles reported harm in the abstract where statistically significant harm was reported in the text, while only 53.5% of articles reported harm in the abstract where any harm was reported in the text. This seems to suggest that articles are more likely to report harm in the abstract if statistically significant harm is reported in the text, than if any harm is reported in the text.

However, in the fourth paragraph it is stated that the relationship between reporting harm in the text and in the abstract was stronger than that relationship between reporting statistically significant harm in the text and reporting harm in the abstract. I am no expert on statistics so I’m not saying that this result is wrong but I would like the authors to explain what these two results mean in relation to one another. I suspect that other readers of the paper may also be confused. Perhaps if more details of the results were given as well as more explanation the paper would be accessible to more people.

Reply: The apparent contradiction referred to by the reviewer lies in the fact that in the first paragraph we only report a crude non-comparative descriptive percentage and in the second one we report the magnitude of the relationship between the presence or absence of a factor (e.g. whether the article reports statistically significant adverse events in the text or not) and an effect (reporting adverse events in the abstract).

Perhaps, the way we describe results in the first paragraph may suggest to the reader —as happened with the reviewer— a kind of measure of magnitude of effect. So although there are no contradictions between data in either paragraph, we decided to improve the text by reducing confusion in the following way:

Original: However, among the articles that reported harm in the text, only 130 articles [53.5% (CI95% 47.0% to 59.9%)] reported harm in the abstract. 41 articles [75.9% (CI95%:62.4% to 86.5%)] reported harm in the abstract when harm in the text was reported as statistically significant (Additional details are shown in figure 2).

New: Among the articles that reported harm in the text, 130 articles [53.5% (CI95% 47.0% to 59.9%)] reported harm in the abstract. 41 articles [75.9% (CI95%:62.4% to 86.5%)] reported harm in the abstract when harm in the text was reported as statistically significant (Additional details are shown in figure 2).
2. In the first paragraph of the Discussion it is stated that 54% of studies that documented harm in the body of the report failed to report harm in the abstract. In the Results section, it is stated that 54% of studies that reported harm in the text DID report harm in the abstract.

Reply: We agree with the reviewer. This is a mistake we committed in the latest draft.

The corrected text reads:

We have found that 33% of the articles in our sample did not report harm in the text; additionally, 46.5% of the studies that documented harm in the body of the article failed to report these harm in the abstract.

Minor essential revisions

Reply: All of them accepted

Discretionary revisions

Reply: Accepted suggestion number 9, about the use of e.g. instead of i.e.
REVIEWER #2

Major compulsory revisions

1. Page 3, para 2: The authors refer to the “latest CONSORT Statement”.
   This is actually one of several extensions to the CONSORT Statement which cover different types of data, designs and interventions. The wording should be clarified here to refer to the CONSORT extension for reporting harms in randomized trials (http://www.consort-statement.org/?o=1044).

   **Reply:** Accepted

   New text: Furthermore, the manner in which abstracts are written tends to influence decision making⁷¹. In fact, the CONSORT extension for reporting harms in Randomized Trials includes a generic recommendation: “If the study collected data on harm and benefits, the title or abstract should so state”⁷².

2. Page 3, para 3: There are studies which have been conducted looking at the extent to which abstracts report harms, it would be good to acknowledge this additional research here thus placing this current study in context. Similarly the findings of this current study are not discussed with those of previous studies in the discussion section.

   **Reply:** Although important, we consider that the suggestion by the reviewer is sufficiently accomplished with the mention of the Consort extension for reporting harms in Randomized Trials. It is certainly true that a lot of research has been conducted in order to understand the way abstracts report relevant information; unfortunately, however there are no articles analysing the specific aspects of our interest (Harm in abstracts in Phase III and IV RCTs). In fact, the CONSORT extension for reporting harms—which was the gold standard reference- bases its 1st recommendation on an empirical search using the Cochrane Library RCT Database in order to know how many times “harm” appears in title or abstract.

   Nevertheless, we will include a sentence in the first paragraph pointing out this fact. Added text:

   We have found that 33% of the articles in our sample did not report harm in the text; additionally, 46% of the studies that documented harm in the body of the article failed to report these harm in the abstract. **Unfortunately, although different approaches to the study of abstracts have been developed, there is no similar research with which compare the coherence and consistency of these results.**

3. Page 3, para 4: It is not clear how these high profile journals were identified, how reports of RCTs were identified from within these journals, and how the sample was selected. More details in the methods section clarifying these points would improve the study’s transparency and reproducibility.
Reply: With regard to the way high profile journals were identified, the text includes (last line in page 3) an explicit reference to the selection process, based on the 2003 impact factor. However we agree and accept the recommendation about how RCTs were identified and how the sample was selected.

New text:

Population and setting: A purposive sample of journals was selected based on their 2003 impact factor (figure 1). A random number table was used to get a sample of RCTs published in 2003. Box 1 shows the PubMed search strategy used for retrieving the articles. (Box 1)

<table>
<thead>
<tr>
<th>Box 1. RCT search strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 randomized controlled trial, pt</td>
</tr>
<tr>
<td>#2 limit #1 to (clinical trial, phase i or clinical trial, phase ii)</td>
</tr>
<tr>
<td>#3 #1 not #2</td>
</tr>
<tr>
<td>#4 limit #3 to yr=2003</td>
</tr>
<tr>
<td>#5 limit #4 to human</td>
</tr>
<tr>
<td>#6 limit #5 to journal article</td>
</tr>
<tr>
<td>#7 comment. pt</td>
</tr>
<tr>
<td>#8 #6 not #7</td>
</tr>
<tr>
<td>#9 #8 not letter. pt</td>
</tr>
</tbody>
</table>

4. Page 5, para 3: It would be very worthwhile to know what sort of information trialists reported in the abstracts when they reported information on harms. For example did they provide numerical data or more generic statements such as "no significant difference in harms was noted between groups". My feeling is that a number of abstracts report the later which is not useful and helpful to the reader. If abstracts did report generic statements such as these, how did you clarify these within your study?

Reply: I absolutely agree with the need to include this subtlety in that paragraph.

Text to add:

Harm was reported or quantified in 135 abstracts [37.2% (CI95%; 32.2% to 42.38%)], 10% of them informing harm as statistically significant. Among those abstracts providing information about harm, just 34% (46 out of 135) used any kind of numerical data.

5. Page 8, para 4: It is not clear whether the authors are referring to the CONSORT Statement or the CONSORT extension for reporting harms here. Both guidelines provide recommendations for reporting both efficacy and harms when reporting a RCT. The challenge is how to get journals to endorse these guidelines and for authors to comply.

Reply: We agree with the revision. Actually, reviewer #1 suggested the same. We have accepted to quote the CONSORT extension
Minor essential revisions

Reply:

All of them accepted except that which referred to the ease in which abstracts are obtained from electronic databases like EMBASE. In our opinion, to have or not to have access to a subscription is not a central point in the implications we are trying to suggest. It would be a critical aspect, indeed, if we were talking about aspects related with systematic reviews or meta-analysis.

With regard to the question number 7, 5 out of 363 articles reported harm in the abstracts although no information about harm was found in the body of the article.
REVIEWER #3

1. Methods/Population and setting. State now the random selection was made: computer generated, random numbers table. We expect trials to report this, and should do so here.

Reply: This suggestion was also proposed by the 2nd reviewer. An alternative text has been added.

2. Methods/Population and setting. “A purposive sample of leading journals was selected taking into account both their 2003 impact factor (figure 1)”. Is there something missing from this sentence, or should “both” be omitted?

Reply: It was a mistake. “Both” should be omitted.

3. Methods/Analysis.” Descriptive measures and exact confidence intervals ....” I am unsure what the authors mean by “exact”, but note that they give results to three significant figures. Given that the numbers involved are a few tens to hundreds, I think this implies spurious precision. It is common to use two significant figures.

Reply: Due to the fact that we have chosen a sample of abstracts confidence intervals estimation –whatever the method- is appropriate, not just to inference purposes but to add information about the precision of the measure. Actually, p values are not useful to determine precision. This is particularly relevant when continuity correction for a binomial distribution –as it is assumed in our data- might not be assured. Then Exact Confidence Interval method improves the estimation of limits. A more familiar statistic for categorical outcomes is Fisher’s exact test. STATA automatically provides these estimations. More may be reached at: (http://www.ats.ucla.edu/stat/Stata/)

4. Results/Table 1. There is inconsistency between the text and the table resulting, I think, from the text reporting funding form industry compared to public institutions (cPR 1.29, CI not given) and the table comparing funding from public institutions compared to industry (cPR 0.77, 95% CI 0.60 to 0.97). It would be sensible to report the same thing in both places, for consistency and clarity.

Reply: There is no inconsistency between the figures as suggested by the reviewer. cPR equals 1.29 is the inverse value of 0.77. However, we agree that this could be confusing for readers so we have changed the sentences thus:

New text: On the other hand, where funding sources were concerned, harm in the abstract was less likely to be reported when public institutions-as opposed to companies- funded the RCTs (cPR=0.77).
5. Discussion. I calculate that 33% of articles did not report harm (120/363), and 47% of studies reporting harm in the text did not report it in the abstract (113/263).

Reply: The reviewer is right, although the denominator of the second figure is not 263 but 243. In any case, there was a mistake pointed out by the 1st reviewer, as well. The text will be corrected in the above mentioned way.

New text: We have found that 33% of the articles in our sample did not report harm in the text; moreover, 46.5% of the studies that documented harm in the body of the article failed to report these harm in the abstract.

6. Discussion. I do not think that mortality is a valid proxy for clinical significance in this context, so cannot infer from this study “that authors report harm in abstracts driven by statistics more than by clinical significance of harm”. Thankfully few patients die in clinical trials.

Reply: Even though we agree with the reviewer in essence, we do think that mortality is a proxy of “clinical significance of harm”. 1) Firstly, in terms of face validity, it is obvious that mortality is the worse outcome expected by a clinician. 2) Secondly, in terms of content validity, we are studying whatever kind of adverse consequence of a therapy or technique (see our definition of harm); mortality alone or within a composite measure, results in the most relevant harm to be studied. 3) Finally, in terms of empirical validity, even though it is less important, mortality is reported by authors in 40% of articles who report harm in the body of the article.

In any case, even though we have recognised the limits of using a proxy, we have not made, even suggested, any inference from the result. Results from our point of view may suggest but not imply, that the probability of harm to be reported is driven, in our sample, by statistics more than by clinical significance.

In order to be more accurate with the expression, we will change text with this new expression.

However, “clinical significance of harm” did not remain in logistic models when statistically significant effects were considered. This may suggest that, in our sample, reporting harm in abstracts is driven more by statistics than by clinical significance of harm.

7. Discussion. In Table 1 for the comparison of industry and public institution funding there is borderline statistical significance at the 5% level. There should be some comment on this. Given that there were only 226 articles in the comparison, and my guess (data not given) is that the majority were from industry, I would not put too much weight on this result. Similarly, the comparison of sample size is only just statistically significant at the 5% level. Perhaps it would be more accurate to
say that this study did not demonstrate a robust difference in likelihood of reporting of adverse events in the abstract according to these two criteria.

**Reply:** Although the hypothesis suggested by the reviewer is consistent with previous knowledge, the majority of missing data was expected from industry evidence in our sample is particularly robust. In fact, we reanalysed data, considering that all missing data belonged to the industry funded RCTs. Even though, this was a probably impossible scenario, results remained similar: industry was more likely to report harm in abstract and the relationship was statistically significant \[cPR=0.78 \text{(CI95\% 0.62 to 0.99)}\].

8. Implications. Para 1. I hope, although I may be wrong, that clinicians and policymakers do not base their judgements on abstracts alone. An important implication of this study is that if adverse events are not flagged up in abstracts, those who are looking for information about them will not know to look further, to read the whole article, to find out about them – even if it is only to find reassurance that there were no differences between test treatment and comparator. Since most trials are not powered to detect differences in rates of adverse events, it is important that it is easy to locate studies with adverse event data so that they can be reviewed, and where appropriate combined with other similar studies, to determine rates with greater confidence. Studies that do not mention adverse events in the abstract may be overlooked in such reviews, except by the most diligent of reviewers, with consequent “loss” of data.

**Reply:** We agree with the reviewer. This is one of the reasons argued by CONSORT extension for reporting harm in abstracts, and this is the reason we decided to point out in the 1st paragraph of the implications chapter, that our article reinforced the policies suggested by CONSORT. They explicitly referred to the need to report harm in the abstract in order to facilitate the appropriate database indexing and information retrieval.

We will add a text pointing out this particular element:
Both elements combined with the fact that a third of the articles in our sample did not report harm and more than a half of the studies that documented harm in the body of the article failed to report this harm in the abstract support strongly the reporting policies suggested by the CONSORT extension; particularly, in those aspect related with appropriate database indexing and information retrieval12.

**Minor essential revisions**

**Reply:** All of them accepted
REVIEWER #4

1. Sample selected, for example, why didn’t the authors hand search the nominated journals? Hand searching is a more precise way of identifying reports of randomized trials compared to electronic searching.

Reply: We certainly agree with the reviewer that hand-searching is a better way to identify reports of RCTs. However, our aim was not to carry out a systematic review or a meta-analysis which requires having a comprehensive searching strategy in order to reach, if possible, both internal and external validity. On the contrary, we aimed to have a sample of RCTs from a set of purposively selected journals in order to test a concrete hypothesis. The criticism of the reviewer would affect, therefore, the external validity of the results, but never the internal one.

2. The reader knows nothing about the clinical domains selected. Some information is needed.

Reply: We think that, in spite of the interesting nature of the proposal suggested by the reviewer, we did not aim to know the influence of clinical domains in reporting harm in the abstract. In any case, figure 1 details the journals selected so that the reader has some available sensitive information about the studied domains: cancer, cardiovascular diseases, surgery –general and thoracic-, gastroenterology and rheumatic diseases.

3. The authors state that a blinded structured review was part of the methodology yet there is no description of how blinding of the articles was achieved. This needs to be reported.

Reply: We certainly agree with the importance of reporting how the reviewers were blinded. In fact, this issue is partially explained in page 4 of our manuscript but probably a better description is needed and we will add some additional information.

New text

Data abstraction: An instrument was developed ad hoc to retrieve key information from each article. Its construct and face validity was assessed by two independent researchers, blinded to the study hypothesis and the RCTs authors and journals.

Once the tool was designed, information from each article was obtained following a three steps method. In the first step, one of us (EBD) selected the articles regarding the inclusion criteria and entered the information about topic, treatment groups, and number of treatment arms. Secondly, two trained junior researchers, blinded to the hypothesis of the study received an electronic copy of each article –any single reference to the authors and to the journal was masked electronically – and
retrieved all the remaining information using the developed tool. Finally, disagreement was resolved by a third blinded observer using consensus when necessary.

4. The reader knows nothing about the type of randomized trial selected. For example, were all randomized trial designs included or were there certain eligibility criteria used (e.g., two group parallel)?

Reply: Figure 1 and the search strategy (included in the text after reviewer #2 suggestions) show the number and reasons for excluding RCTs. Thus, phase I and phase II RCTs and factorial designs were excluded. Crossover studies, two group parallel etc. were included.

5. The reader knows almost nothing about the eligibility criterion as it relates to the interventions reported in the sample: were all interventions considered equivalent (anticipated harms) and whether the trials were short term or longer term ones with follow-up. This is important when trying to ascertain information about (relevant) reporting of harm. For short term trials, perhaps with less serious anticipated harms, such as schizophrenia, where a pharmaceutical might be given for six to eight weeks possibly with no follow-up, the expectation of reporting harms might be considerably less serious than for example, a cancer intervention given for three months with 12 months of follow-up. This report does not appear to differentiate such differences.

Reply: With regard to the hypothesis presented by the reviewer, he is assuming that the follow-up period must be considered a factor related with the probability of reporting harm. However, published evidence refutes this hypothesis, since Ioannidis and Lau (JAMA 2001; 285: 437-43), in their article Completeness of Safety Reporting in Randomized Trials showed that follow-up period has no effect when adjusted for several other predictors.

When we designed our causal model, we took into account those factors related to the way harm is reported, and we rejected this one because of the results of this study.

Minor Essential Revisions

1. In the abstract (Methods) the authors use of the term “leading” is imprecise. Leading in who’s eyes. The authors should use “impact factor” as their descriptor replacing “leading”.

Reply: We agree with the reviewer in seeking a more appropriate description for a scientific paper. We will change the text following his recommendation

New text:
(...) published in high impact factor medical journals in 2003.

2. One page 4 (1st line) the authors indicate that 765 reports of phase III and IV RCTs were included. Aren’t phase IV more generally open label and not randomized?

Reply: We simply selected Randomized Trials for the study. Actually, we rejected (figure 1) 30 studies which were indexed as RCTs that were not.

3. While the authors have selected what they call “high profile literature” these results might not be applicable to where most trials are reported. Chan and Altman indicate that the majority of trials are reported in small specialty journal (Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005; 365:1159-1162). The authors need to more fully address this point in any revision.

Reply: We certainly agree with the reviewer. However, this just affects the generalization of our results to the whole universe of RCTs which was not our aim. We just were interested in what we called “high profile” journals.

In fact, with our results, we just might infer –as we did in our manuscript- what happens -with a certain grade of confidence- in the universe from which we got the sample we have analysed.