

Reviewer 1: Williams, HC

Hywel Williams,
 Professor of Dermato-Epidemiology,
 Centre of EvidenceBased Dermatology
 Possible competing interests - an advocate of clear reporting of clinical research

The Study	Yes	No
Is the research question clearly defined?	✓	
Is the overall study design appropriate and adequate to answer the research question?	✓	
Are the participants adequately described, their conditions defined, and the inclusion and exclusion criteria described?	✓	
Are the patients representative of actual patients the evidence might affect?		✓
Are the methods adequately described?	✓	
Is the main outcome measure clear?	✓	
Are the abstract/summary/key messages/limitations accurate?	✓	
Are the statistical methods described?		✓
Are they appropriate?	✓	
Is the standard of written English acceptable for publication?	✓	
Are the references up to date and relevant? (If not, please provide details of significant omissions below.)	✓	
Do any supplemental documents e.g. a CONSORT checklist, contain information that should be better reported in the manuscript, or raise questions about the work?	✓	

If you answered No to any of the above, please supply details below.

A study about studies, not patients.
 No stats as reporting is narrative/descriptive.

RESULTS AND CONCLUSION (For articles reporting research findings only)	Yes	No
Do the results answer the research question?	✓	
Are they credible?	✓	
Are they well presented?	✓	
Are the interpretation and conclusions warranted by and sufficiently derived from/focused on the data?	✓	
Are they discussed in the light of previous evidence?	✓	
Is the message clear?	✓	

If you answered No to any of the above, please supply details below.

REPORTING AND ETHICS	Yes	No
Is the article reported in line with the appropriate reporting statement or checklist (e.g. CONSORT)?		✓
Are research ethics (e.g. consent, ethical approval) addressed appropriately?		✓
Is the article free from any concerns about publication ethics (e.g. plagiarism, fabrication, redundant publication, undeclared conflicts of interest)?	✓	

If you answered No to any of the above, please supply details below or contact the editorial office.

I am not aware of any reporting guidelines on articles that report on reporting guidelines!
I don't ethical permission is needed for this secondary research.

Req Recommendation

Accept

✓ Minor Revision

Major Revision

Reject

Would you be willing to review a revision of this manuscript?

✓ Yes

No

Comments

If you have any further comments for the authors please enter them below.

Thank you for including an open comments box as the tick box approach in your main section does not apply to many of the points I want to make.

I enjoyed reading this paper as it makes a simple and straightforward point - that is that STROBE should be used for what it is meant to be used for ie reporting all the relevant bits that readers need to see in order to judge quality, relevance and clinical meaning.

The study is clearly written and provides a nice textual example of "STROBE abuse". I have some comments (all minor), that I would like the authors to reflect on:

1. In the abstract, state in the first para that STROBE was 1st published in Oct 2007 in order to give readers more context as to why you then chose August 2010 to evaluate its use (ie it is clear that they have enough time to bed down).
2. Abstract: no mention is made of the 49 articles that were not observational studies or sys reviews - either make a clear statement on why they are not considered further or say something about them rather than let them disappear mysteriously
3. Whilst I fully appreciate the distinction between reporting quality and study quality as exemplified in my editorial on the CONSORT 2010 update (Williams HC. Cars, CONSORT 2010 and Clinical Practice. *Trials* 2010 Mar 24;11(1):33), I do think you are being a bit harsh at times on the magnitude of the sin of using at least some elements of STROBE for improving observational study design or quality assessment. Granted it is completely inappropriate to use items mentioned in the title as a quality criterion, but others such as blinding are likely to be quality criteria as well as things that need to be reported. Are they such terrible side effects if the result of STROBE abuse is to improve the quality of study design? All I ask is that you revisit some of your statements and maybe soften a few as it currently reads as if anyone not using STROBE for the sole use of reporting should be shot (although I agree that using STROBE as a scoring system is a very serious offence).
4. I have not got my head around why you did a 1st search of articles containing STROBE in the title and what you did with them. You started off in the background saying that STROBE had been cited a lot, and I thought that your main purpose in this paper was to see how articles that had cited STROBE (your 2nd search) had used it. Perhaps I missed something, but please make it clearer why you searched for STROBE in the title as it seems like the wrong sort of article you needed to answer your research question.
5. In the methods on page 6, you state that one person data extracted, which is fair enough. But how many and who made that crucial largely subjective decision of whether the use of STROBE was appropriate or inappropriate - maybe it a very clear easy decision, or maybe it was a bit grey at times, in which case a single research who "had it in" for STROBE abusers might have been a bit too stringent.
6. The single most important message from the research is that over half of the systematic review made inappropriate use

of STROBE. Don't leave us hanging there with just a dry statement at the bottom on page 8 as this is the juiciest bit of your paper. What were the characteristics of those sys reviews? Were any of them Cochrane? (heaven forbid), were they more likely to be in a non-English journal, a specialist versus general journal, a specialty area, or on teams that did not include an epidemiologist? - you sort of comment on this on the top of page 12, but I think you could say more in the results section, even though it is exploratory in nature.

7. I realise that cyberspace is limitless, but I am not convinced that Figure 1 adds much, and I would suggest dropping it.

Reviewer 2: Little, Julian

Julian Little

Canada Research Chair in Human Genome Epidemiology

Professor and Chair

Department of Epidemiology and Community Medicine

University of Ottawa

Canada

The Study	Yes	No
Is the research question clearly defined?	✓	
Is the overall study design appropriate and adequate to answer the research question?	✓	
Are the participants adequately described, their conditions defined, and the inclusion and exclusion criteria described?	✓	
Are the patients representative of actual patients the evidence might affect?	✓	
Are the methods adequately described?		✓
Is the main outcome measure clear?	✓	
Are the abstract/summary/key messages/limitations accurate?	✓	
Are the statistical methods described?	✓	
Are they appropriate?	✓	
Is the standard of written English acceptable for publication?	✓	
Are the references up to date and relevant? (If not, please provide details of significant omissions below.)	✓	
Do any supplemental documents e.g. a CONSORT checklist, contain information that should be better reported in the manuscript, or raise questions about the work?		✓

If you answered No to any of the above, please supply details below.

- What was the rationale for choosing 100 as the number of articles to sample randomly? This sample led to 32 observational studies, and 19 systematic reviews being identified and examined. Some questions that occurred to me on basis of data in Table 1 might be more amenable to examination with larger sample size? Is this study in a sense a pilot for further work?

RESULTS AND CONCLUSION (For articles reporting research findings only)	Yes	No
Do the results answer the research question?	✓	
Are they credible?		✓
Are they well presented?	✓	
Are the interpretation and conclusions warranted by and sufficiently derived from/focused on the data?		✓
Are they discussed in the light of previous evidence?	✓	
Is the message clear?	✓	

If you answered No to any of the above, please supply details below.

See earlier point about rationale for sample size.

- The authors examined associations between number of citations and journal impact factor, and appropriate use and affiliation. Other potential relationships that might have been examined include (a) nature of journal – general medical vs. epidemiological, and language of STROBE citation (as a proxy for first language of investigators). I appreciate these features may be correlated with impact factor, but might they give more information on the problem of inappropriate use?
- On appropriate use, it looks as if this is based on the use stated to have been made in the papers reviewed. In the observational studies, did you examine whether reporting practice actually bore out the use the authors stated that they made of STROBE?
- Regarding inappropriate use of STROBE in systematic reviews, does this reflect lack of optimal tools to assess observational studies (Sanderson S et al. International Journal of Epidemiology 2007;36:666–676)?

REPORTING AND ETHICS

Yes No

Is the article reported in line with the appropriate reporting statement or checklist (e.g. CONSORT)?

✓

Are research ethics (e.g. consent, ethical approval) addressed appropriately?

✓

Is the article free from any concerns about publication ethics (e.g. plagiarism, fabrication, redundant publication, undeclared conflicts of interest)?

✓

req Recommendation

Accept

✓ Minor Revision

Major Revision

Reject

Would you be willing to review a revision of this manuscript?

✓ Yes

No

Authors Response to Decision Letter for (BMJ Open.2010.000048)**Uses and misuses of the STROBE statement: bibliographic study**

Dear Mr. Sands,

Thank you for the good news. We have read with attention the comments/suggestions provided by the reviewers. Please find below our reply to the comments of the reviewers.

Modifications to our manuscript are in colored text (red).

Thanks kindly and we look forward to hearing from you.

Best Regards,
Bruno da Costa

Reviewer(s)' Comments to Author:
Reviewer: Hywel Williams,
Professor of Dermato-Epidemiology,
Centre of EvidenceBased Dermatology

I enjoyed reading this paper as it makes a simple and straightforward point - that is that STROBE should be used for what it is meant to be used for ie reporting all the relevant bits that readers need to see in order to judge quality, relevance and clinical meaning.

Authors: Thank you very much for taking the time to review our manuscript.

The study is clearly written and provides a nice textual example of "STROBE abuse". I have some comments (all minor), that I would like the authors to reflect on:

1. In the abstract, state in the first para that STROBE was 1st published in Oct 2007 in order to give readers more context as to why you then chose August 2010 to evaluate its use (ie it is clear that they have enough time to bed down).

Authors: We have added the information to the abstract as suggested by the reviewer.

2. Abstract: no mention is made of the 49 articles that were not observational studies or sys reviews – either make a clear statement on why they are not considered further or say something about them rather than let them disappear mysteriously

Authors: We have added the information to the abstract as suggested by the reviewer.

3. Whilst I fully appreciate the distinction between reporting quality and study quality as exemplified in my editorial on the CONSORT 2010 update (Williams HC. Cars, CONSORT 2010 and Clinical Practice. *Trials* 2010 Mar 24;11(1):33), I do think you are being a bit harsh at times on the magnitude of the sin of using at least some elements of STROBE for improving observational study design or quality assessment. Granted it is completely inappropriate to use items mentioned in the title as a quality criterion, but others such as blinding are likely to be quality criteria as well as things that need to be reported. Are they such terrible side effects if the result of STROBE abuse is to improve the quality of study design? All I ask is that you revisit some of your statements and maybe soften a few as it currently reads as if anyone not using STROBE for the sole use of reporting should be shot (although I agree that using STROBE as a scoring system is a very serious offence).

Authors: We agree with the reviewer that we are rather harsh at times when referring to the inappropriate use of the STROBE. However, we did so deliberately, in order to send out the message that the intention of STROBE authors was not to dictate methodology of observational studies, or to stifle creativity. We included a remark on the previous version of our manuscript on how STROBE could be helpful in certain situations when designing a study. We have now expanded this paragraph (pag. 10) to clarify this further.

4. I have not got my head around why you did a 1st search of articles containing STROBE in the title and what you did with them. You started off in the background saying that STROBE had been cited a lot, and I thought that your main purpose in this paper was to see how articles that had cited STROBE (your 2nd search) had used it. Perhaps I missed something, but please make it clearer why you searched for STROBE in the title as it seems like the wrong sort of article you needed to answer your research question.

Authors: Thank you for pointing that out. We have provided further explanation of this search strategy in the manuscript under "literature search". In order to identify articles which cited STROBE, we first needed to identify STROBE publications; ergo the search strategy was built to identify STROBE publications (first step). In the second step, we used the tool "create citation report" available in the WoK database. This tool enables the search for articles which cited the articles yielded for the first search.

5. In the methods on page 6, you state that one person data extracted, which is fair enough. But how many and who made that crucial largely subjective decision of whether the use of STROBE was appropriate or inappropriate – maybe it a very clear easy decision, or maybe it was a bit grey at times, in which case a single research who "had it in" for STROBE abusers might have been a bit too stringent.

Authors: Prior to data extraction, the definition of appropriate use of STROBE was extensively discussed within our group. Discussions were led by a senior epidemiologist (ME). There was one single data extractor for all articles included (BRdC). Whenever the data extractor had a doubt about the classification of an article, he discussed it with other authors. In case of disagreements, final decision was made by the senior epidemiologist (ME). We have further explained this in our manuscript under "data collection".

6. The single most important message from the research is that over half of the systematic review made inappropriate use of STROBE. Don't leave us hanging there with just a dry statement at the bottom on page 8 as this is the juiciest bit of your paper. What were the characteristics of those sys reviews? Were any of them Cochrane? (heaven forbid), were they more likely to be in a non-English journal, a specialist versus general journal, a specialty area, or on teams that did not include an epidemiologist? - you sort of comment on this on the top of page 12, but I think you could say more in the results section, even though it is exploratory in nature.

Authors: Thank you for this suggestion. None of these reviews were Cochrane reviews. We have added additional information to the last paragraph of the methods and results section, and to paragraph 8 of the discussion section (pag. 13).

7. I realise that cyberspace is limitless, but I am not convinced that Figure 1 adds much, and I would suggest dropping it.

Authors: We have removed figure 1 from the manuscript as suggested by the reviewer.

Reviewer: Julian Little
Canada Research Chair in Human Genome Epidemiology
Professor and Chair
Department of Epidemiology and Community Medicine
University of Ottawa
Canada

Authors: We would like to thank the reviewer for taking the time to assess our paper.

- What was the rationale for choosing 100 as the number of articles to sample randomly? This sample led to 32 observational studies, and 19 systematic reviews being identified and examined. Some questions that occurred to me on basis of data in Table 1 might be more amenable to examination with larger sample size? Is this study in a sense a pilot for further work?

Authors: The decision to randomly sample 100 articles for detailed assessment was pragmatic as we did not have the necessary resources to investigate all studies which have cited STROBE. The primary intention of this study was to inform the STROBE group about its use, and results were presented and discussed during the STROBE revision meeting. We do not plan at this moment to examine a larger sample of articles.

- The authors examined associations between number of citations and journal impact factor, and appropriate use and affiliation. Other potential relationships that might have been examined include (a) nature of journal – general medical vs. epidemiological, and language of STROBE citation (as a proxy for first language of investigators). I appreciate these features may be correlated with impact factor, but might they give more information on the problem of inappropriate use?

Authors: A similar request was made by reviewer 1 (comment #6). We have now added some additional information to the manuscript.

- On appropriate use, it looks as if this is based on the use stated to have been made in the papers reviewed. In the observational studies, did you examine whether reporting practice actually bore out the use the authors stated that they made of STROBE?

Authors: Indeed, we have only checked whether authors claimed to have used the STROBE. We did not assess further the use of STROBE by the authors as this was beyond the scope of our review.

- Regarding inappropriate use of STROBE in systematic reviews, does this reflect lack of optimal tools to assess observational studies (Sanderson S et al. International Journal of Epidemiology 2007;36:666–676)?

Authors: Thank you very much for this comment. We agree with the reviewer that one of the reasons researchers use STROBE as a quality assessment tool may be the lack of validated tools to do so. We have added this to our discussion section (pag. 11) and cited the study mentioned by the reviewer.