Reviewer's report

**Title:** Benefits and harms of the human papillomavirus (HPV) vaccines: comparison of trial data from clinical study reports with corresponding trial register entries and journal publications

**Version:** 0  **Date:** 28 Aug 2018

**Reviewer:** Julie Glanville

**Reviewer's report:**

This is an interesting and useful paper and adds to the growing literature on the importance of trial information that is becoming increasingly available outside of the published journal literature.

The authors are very experienced in the field of identifying unpublished and non-journal literature, but their readers may be less experienced in the field. I have suggested some clarification points which might help such readers' understanding.

**Background line 76-77.** Some CSRs are accessible via the EMA site but not necessarily all and there are restrictions on the uses which can be made of the CSRs. Some clarifications of this lack of comprehensive coverage and an outline of current uses that are permissible for CSRs will help to show the degree of access that is available. CSRs are still not open and usable by everyone even if posted on the EMA site.

**Background lines 86-90.** In the conclusions to the paper, the authors state that '[CSRs] were the quantitatively and qualitatively superior trial documents…'. However, in describing the aims and methods of the study the means of measuring superiority are not really described in detail. It would be helpful to know how superiority was to be measured in more detail.

The second aim 'to compare the reporting of trial design aspects of the corresponding trial documents' is expressed in quite general terms, and needs to be linked more clearly to the more detailed listing of aspects in lines 101-107.

It would be helpful for readers to know the value and purpose of collecting the data in lines 101-107 of the Methods.
Methods lines 97-99. How did the authors search for the trial register entries? There are many trials registers and it would be helpful to know which ones were searched and how. Also it should be noted that some trial registers other than ClinicalTrials.gov (e.g. EU clinical trials register) do also include results summaries. How did the authors search for the primary journal publications related to the trials?

Results lines 128-131. It would be helpful to know whether the 'publication' dates of the various documents were similar - as far as that is possible to tell? In the section on trial document differences on page 8, could some of these differences just be a factor of documents being written at different timepoints.

Results line 128-131 - researchers often publish several papers around a trial, sometimes reporting outcomes outside of a primary publication or even in a conference paper. ClinicalTrial.gov records often list these multiple publications as well as the results submitted to ClinicalTrials.gov. Did the authors try to identify other related publications, where further outcomes might have been presented?

Results lines 143-144. Although the EMA reports had fewer pages, were they just condensed versions of the GSK reports, or did they omit outcomes?

Results lines 155--160 I have seen many CSRs and many do not include the protocol. Should a CSR include the protocol according to published guidance or 'accepted practice'? Protocols are often published as journal articles - did the authors search for these? Protocol information also can sometimes be included on manufacturers' websites - were these searched?

Results lines 163-164. Should the word 'all' be inserted before 'the six', since some of the issues were reported in the trial register entries in Table 2, although no records reported all six issues. Sometimes the trial methods are published in detail in journal articles separate to the 'primary' publication. Did the authors check for these? Sometimes there are supplementary documents attached to the primary publication which may contain additional information - did the authors check for these? Sometimes there may be clarifications in errata or letters relating to the primary publication. Did the authors check for these? It is not ideal when important information are 'hidden' in such documents, but it is helpful for researchers to know that this may occur.
Results lines 171-172. The example presented here might be one that contributes to the statement about 'qualitatively and quantitatively' superior - is the exclusion of immunological disorders is an important omission in this context?

Discussion - trial registers. The authors should discuss how ClinicalTrials.gov could be improved as it would not necessarily be the best use of most researchers' time to work through thousands of pages of CSR data. Surely it would be preferable to see a richer ClinicalTrials.gov? It seems that ClinicalTrials.gov could be improved, not only in requiring more timely results registration, but also in capturing the information it seems to currently miss in this case study, as highlighted by the authors' Table 2.

Discussion - trial document differences (p8) - this is such an interesting set of issues because it highlights how there is no single place one could expect to find the full accurate story of the trial and it raises questions about which of any of the documents one would 'trust'. Since there are inconsistencies even with a CSR to fall back on, where should a researcher expect to find the full data?

Discussion line 234-235 - the authors seem surprised that there were no changes in direction of 'available' results and then go on to speculate that GSK may be more transparent than other companies. It would be fair I think to put in some references here to support the assertion that other companies have been less transparent in the way that the authors suggest.

Discussion lines 239-244. A lot is crammed into these lines and I think the points being made may not be obvious to all readers. Please can the authors expand on these points to bring out the issues in a little more depth.

Discussion lines 247-248 based on the questions above, please can the authors review the statement 'Our comparison included nearly all (71/72) eligible trial documents'.
Discussion lines 255-260. The authors seem to assume that the EMA reports are missing data that are in the full CSRs but perhaps the EMA reports are condensed and the essential data are included? The authors do not report any discrepancies in these lines or number of discrepancies, so perhaps this point is not yet fully proved, or perhaps they can add some data here to clarify. It would be helpful to know if the EMA reports are less useful than a full CSR - what might have been missed.

Discussion lines 261-268. The per protocol comparison issue discussed here was not really highlighted in the Methods section as a key question, so it comes as a surprise to find it discussed in detail in the discussion. Please can the authors introduce the issue in the Methods section and then the discussion will be less of a surprise.

Discussion lines 269-271. There seem to be current limits to the number of results that ClinicalTrials.gov will capture for adverse events (https://clinicaltrials.gov/ct2/manage-recs/how-report). The authors may wish to encourage more complete recording or propose other frequency thresholds? This would build on the benefits of ClinicalTrials.gov (over journal publications) reported in lines 285-288

Discussion lines 290-292. These references might be useful as well.


Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field that should be highlighted to relevant networks
Quality of written English
Please indicate the quality of language in the manuscript:
Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I offer training courses in identifying clinical trial records.
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal