

Author's response to reviews

Title: Is the coverage of Google Scholar enough to be used alone for systematic reviews?

Authors:

Jean-Francois Gehanno (jean-francois.gehanno@chu-rouen.fr)
Laetitia Rollin (laetitia.rollin@chu-rouen.fr)
Stefan J Darmoni (stefan.darmoni@chu-rouen.fr)

Version: 2 **Date:** 8 November 2012

Author's response to reviews: see over

Answers to reviewer's comments

General comment

We are disappointed by the dramatic difference of appreciation of the two reviewers. While for one reviewer, the manuscript could be published nearly as it is, nothing could save the study for the other reviewer.

Reviewer's 1 comments

Henrik von Wehrden

*« There is only one minor Essential Revisions:
You wrote the word “recall” always with a large “R”, expect twice, which should be corrected, I guess. »*

This has been corrected in the manuscript.

Reviewer's 2 comments

Miguel Garcia-Perez

*« The authors of this manuscript set out to investigate whether Google Scholar (GS) could be used alone to retrieve all the relevant documents that a researcher should have in hand in order to perform a systematic review. Part of their method is correct, namely, they define a “gold standard” of references that such a search should retrieve, which is to be used as a criterion to assess the performance of GS. But from that point on, the authors’ method is essentially flawed: They then search for each of those references in GS “one by one, by searching with the title of each of the studies included in the gold standard database”. This strategy renders a retrieval rate of 100%, which the authors regard as proof that GS could indeed be used alone for a literature search: “if the authors of these 29 systematic reviews had used only GS, they would have obtain[ed] the very same results”. But the reader is left wondering how could GS be claimed to have been used “alone” in this study when the target references had previously been identified (the authors’ gold standard) and then searched for one by one by name in GS.
What the authors show is only that if one knows what studies should be identified, then one can go to GS, search for them one by one, and find out that they are indexed. But, if a researcher already knows the studies that should be included in a systematic review, why bother to also check whether those studies are indexed in GS? »*

The main remark of Reviewer 2 concerns the methodology. He considers that building a gold standard of references to assess the « performance » of GS is irrelevant.

First of all, we don't talk about « performance » as a whole, but about precision and recall, and about the coverage of a bibliographic database.

Assessing the precision was not our purpose, which is clearly stated in our manuscript, and discussed in the « discussion section ».

We assessed the coverage of GS, which is unknown, in general and more importantly for high quality studies, and therefore the possibility to obtain a Recall of 100%, or not.

Although reviewer 2 considers that the methodology we used for that purpose was irrelevant, the same kind of strategy was used by the Hedges Team (a very well known research team in the field of information retrieval methods), in an article published by Victor Montori and Brian Haynes in the BMJ in 2005. The authors hand searched for systematic reviews in 161 journals

indexed in Medline and built a gold standard of references. They then searched for these references using different search strings.

In many other studies, the authors built a gold standard of articles and then studied the coverage of the databases by searching one by one for the selected articles (Gehanno, 1998; Haafkens, 2006; Walters, 2007).

We therefore consider that the comment of Reviewer 2 is not adequate.

What we demonstrate in our study is that a Recall of 100% is theoretically possible for high quality studies in GS, which has never been demonstrated for any other bibliographic database!

We even went further on this in the discussion since we used several search strings which really led to a recall of 100% (which is therefore possible), to assess the precision for such strings giving a 100% recall. The precision of 0.1% was low and that is why our conclusion was “GS just requires some improvement in the advanced search features to improve its precision and to become the leading bibliographic database in medicine”.

Nevertheless, we realize, according to the comments of reviewer 2, that the term 100% recall could be misleading. Therefore, we added in the manuscript the fact that the coverage of GS for high quality articles is 100% and that recall can reach 100% (both being closely linked).

The reviewer says that *“At the end of the manuscript the authors give some sample cases clearly indicating that, in sharp contrast with their conclusion, GS cannot by any means be used alone: It returns thousands of documents, the vast majority of which are irrelevant » « Regarding the quotation at the end of the preceding paragraph, if a researcher had only used GS, he/she would have obtained a list in which the relevant documents for his/her systematic review would be hidden among tens of thousands of irrelevant documents. The researcher would have had to spend days on end separating the wheat from the chaff”*.

This is always the case for systematic reviews, and this is the case with every database when you use search strings that aim at optimizing Recall, even in Medline!

In the above-mentioned study by Montori et al., in Medline, their search string “Validation without CDSR”, for retrieving systematic reviews gave a Recall of 99.7, but a Precision of 1.4. Since their database included 48 258 records, this means that the string gave also “thousands of documents, the vast majority of which are irrelevant »

Another example is found in the Cochrane systematic review « Pharmacotherapy for anxiety disorders in children and adolescents » (included in our study). Using several databases, the authors retrieved 472,907 references, screened 201 abstracts and finally included 22 studies !

In the Cochrane systematic review « Erythropoietin or Darbepoetin for patients with cancer - meta-analysis based on individual patient data» (also included in our study), the authors used 11 databases, and very complex search strings. They retrieved 5,546 references, screened 447 full text publications. They also searched in several years of conference proceedings and they went through reference lists of evidence based guidelines. Overall, they included 36 trials. Using a very simple search string in GS (« (Erythropoietin or Darbepoetin) cancer »), we identified 36,630 references, with a recall of 100% for the trials included in the systematic review. This supports our statement that GS could be used alone, providing some improvements to enhance the Precision of searches.

Therefore, we disagree with the comments of reviewer 2 and we consider, like reviewer 1, that our statements are supported by our results.

References

1. Montori VM, Wilczynski NL, Morgan D, Haynes RB: Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68.
2. Gehanno JF, Paris C, Thirion B, Caillard JF: Assessment of bibliographic databases performance in information retrieval for occupational and environmental toxicology. *Occup Environ Med* 1998;55:562-566.
3. Haafkens J, Moerman C, Schuring M, van Dijk F. Searching bibliographic databases for literature on chronic disease and work participation. *Occup Med (Lond)*. 2006;56(1):39-45.
4. Walters WA. Google Scholar coverage of a multidisciplinary field. *Information Processing & Management* 2007;43(4):1121-1132.