Author's response to reviews

Title: Developing Search Strategies for Clinical Practice Guidelines in SUMSearch and Google Scholar and Assessing their Retrieval Performance

Authors:

Andrea Haase (a.haase@iqwig.de)
Markus Follmann (markus.follmann@iqwig.de)
Guido Skipka (guido.skipka@iqwig.de)
Hanna Kirchner (hanna.kirchner@iqwig.de)

Version: 3 Date: 7 March 2007

Author’s response to reviews: see over
Dear Ms. Puebla,

Attached please find the revised version of our manuscript “Developing Search Strategies for Clinical Practice Guidelines in SUMSearch and Google Scholar and Assessing their Retrieval Performance” (Andrea Haase, Markus Follmann, Guido Skipka and Hanna Kirchner), MS: 2143750024125566.

The point by point response for reviewers and the amended manuscript (changes highlighted with the “WORD track changes function”) are attached as WORD files.

Thank you for considering the publication of our paper in your journal.

Yours sincerely,

Andrea Haase
## Point by point response to reviewers

### Author’s note

- Two inconsistencies corrected in the Methods section on pages 4 and 5 (“depressive disorder” not “depression”; “we classified a retrieval as relevant”, not “we classified a CPG term as relevant…”).
- Inconsistencies in the formatting of table footnotes corrected.
- The reviewers requested additional information in the abstract. In order to comply with the word limit, we shortened the abstract's conclusion.

### REVIEWER’S COMMENTS

<table>
<thead>
<tr>
<th>Thomas Ganslandt – no revisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nancy Wilczynski</td>
</tr>
</tbody>
</table>

#### Minor Essential Revisions

1) Because of the way the reference standard was constructed the authors might consider referring to the diagnostic parameters as relative sensitivity, relative specificity and relative NNR as the values will change dramatically depending on the pool.  

Text added in Methods section (page 7):  

*It should be noted that these diagnostic parameters will change depending on the reference pool.*

We prefer not to use the term “relative”, as, for example, the term “relative sensitivity” is preferentially used when the sensitivity of 2 diagnostic tests is compared by a ratio.

2) On page 3 of the manuscript, top of the page, the author should change „...be used if a fast detection of CPGs...“ to „...be used if a less time consuming approach for the detection of CPGs...“

Changed accordingly on page 3.

3) On page 4 of the manuscript, 1st paragraph under the Methods section, the author should define „a limited period of time.“

Amended on page 4 (new text underlined):  

*Efficient* meant detecting as many unique and relevant CPGs as possible in a given period of time (which varies depending on the time available to the user).

4) On page 5 of the manuscript, final sentence, the authors should change the sentence slightly to „The second was one of the nine disease-specific MeSH terms mentioned earlier.“

Changed accordingly on page 6.

5) In the discussion section of the manuscript the author might consider adding that clinicians can always search the National Guidelines Clearinghouse if they are interested in locating guidelines that meet a defined quality filter (the NGC inclusion criteria) and have been published within the previous 5 years.

Added to the Background section on page 3:

*CPG-specific databases exist, such as the National Guideline Clearinghouse, which includes CPGs that meet defined inclusion criteria and have been published within the*
6) Table 1, the authors should indicate how the retrievals from Google Scholar were estimated. Also, the column heading for the 4th column should be „Relevant Retrievals CPG term + „back pain“.

Changed in Table 1 (new text underlined):
- 3rd column heading
  Retrievals Google Scholar
  (estimated number of retrievals)§

Footnote: §Defined by Google Scholar as “Results 1-10 of about…”
- 4th column heading changed according to reviewer’s suggestion.

7) Figure 1, the author should change the text slightly in the two „clouds“, the 1st to „3 CPG terms with the highest retrieval:“ and the 2nd to „3 CPG terms with the highest relevant retrieval:“ Also, there is a typo in the box headed „Google Scholar“, 1st point, „und“ should be „and“

Changed accordingly in Figure 1.

Robert Badgett

Major Compulsory Revisions

8) I had two principle difficulties. First was difficulty understanding the methods. I am sympathetic to this as when I have done similar research my methods have also evolved during the project. This evolution complicates writing the description. My specific questions are clarifying numbers, specifying final search strategies exactly (or at least an example for one of the 9 diseases), and which sections of SUMSearch were used (as its PubMed guidelines section has been offline). Details of my questions are below in the Discretionary section. My concerns are listed in discretionary because every question does not need addressing, but enough to clarify.

Please refer to the comments below in the “Discretionary Revisions” section.

9) My second difficulty is the use of the „in the title“ restriction at Scholar. This should be justified because the two major web locations for practice guidelines, the National Guidelines Clearinghouse and PubMed, do not concatenate „guideline“ into the HTML title field of their pages. This restriction would seem to reduce the sensitivity of Scholar (though should increase its specificity). Can the authors state why they did this? Even if the answer is that informal searching noted that Google was retrieving excessive citations without this limit, the reason should be stated.

Information added (page 6):

We used the “all in title” restriction, as prior informal searching in Google Scholar had produced an excessive amount of retrievals without this limit.

Discretionary Revisions

10) What was the role of the nine different diseases is
diseases (obesity, osteoporosis, rheumatoid arthritis, Parkinson disease, multiple sclerosis, alcoholism, depression, schizophrenia, and attention deficit disorder) versus the sole disease back pain? At times it appears that all of the study was done only with back pain. However, the sentence „The second term was one of the nine MeSH terms mentioned above“ and the number of guidelines found (119) makes me uncertain.

To make it clear that this study was conducted with a two-step approach, we now include the terms “preliminary study” and “main study” in the document. We have also added the following information in the Methods section on page 5 (new text underlined):

An overview of the study methodology is shown in Figure 1. We used a two-step approach: (1) the development of a GLAD search strategy (preliminary study); (2) the application of this strategy and the comparison of retrieval performance (main study).

The term “back term” was the test term used in the preliminary study as an example of a common disorder, and is otherwise not connected to the nine diseases above.

Text amended in the Methods section on page 5 (new text underlined):

We then checked these 14 terms for relevance in combination with the MeSH term “back pain” in SUMSearch (DARE, NGC, and PubMed), using the test term “back pain” as an example of a common disorder and a substitute for the specific disease terms.

<table>
<thead>
<tr>
<th>11) Finally, in the titles to Figure 2 and Table 4, after I spent much time on the paper, I realize it is clear that all the results are for the 9 diseases pools. The abstract states nine MeSH terms, but would be helpful to also state 9 diseases in the abstract.</th>
</tr>
</thead>
<tbody>
<tr>
<td>We agree that the fact that the retrievals for the nine diseases were pooled was not made clear enough. Information added in the abstract and in the Methods section.</td>
</tr>
<tr>
<td>Abstract, page 2 (new text underlined):</td>
</tr>
</tbody>
</table>

---

**Our research focussed on nine different diseases currently being evaluated by the German health authorities with regard to their suitability for inclusion in disease management programmes (obesity, osteoporosis, rheumatoid arthritis, Parkinson disease, multiple sclerosis, alcoholism, depressive disorder, schizophrenia, and attention deficit disorder).**
We used three different CPG terms and nine MeSH terms for nine selected diseases to identify the most efficient GLAD strategy for each search engine. The retrievals for the nine diseases were pooled. To compare GLAD strategies, we used a manual review of all retrievals as a reference standard.

Methods, page 7:

The retrievals for the nine diseases were pooled. The application of the GLAD search strategy therefore resulted in three retrieval pools (one for each CPG term applied) per search engine.

Due to space limitations (word limit: 350 words), we could not list the single nine diseases in the abstract.

| 12) Starting with the abstract, I suggest clarifying numbers. In the abstract, „The search yielded a total of 2830 retrievals“, I presume this is total number of citations retrieved, if so, then state how many citations were in the gold standard collection (119?).

Perhaps wording could be: „We found 119 relevant guidelines for nine diseases. SUMSearch retrieved 1843 documents including 97 (81%) of the relevant guidelines.

Abstract amended: Information on the gold standard collection included in the abstract on page 2 (new text underlined):

The search yielded a total of 2830 retrievals; 987 (34.9%) in Google Scholar and 1843 (65.1%) in SUMSearch. Altogether, we found 119 unique and relevant guidelines for nine diseases (reference standard).

Information on how these results were obtained added to the first paragraph of the Results section (please see comment on Revision No. 14).

The number “97” refers to the number of relevant guidelines retrieved with the search strategy “Guideline + (9) MeSH Terms” in SUMSearch (please see Table 4), not to the total number of relevant guidelines retrieved by SUMSearch. SUMSearch retrieved 105 relevant guidelines, of which 71 were unique and relevant: this information is now included in the text of the Results section (not just in Figure 2); please see new text in comment on Revision 14).
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>13</strong></td>
<td>This leads to NNR of... As you can see, I am still having trouble with the number because the numbers I use do not lead to the NNRS you report. Seems the NNR for SUMSearch should be 1843/97 or 19.</td>
</tr>
<tr>
<td><strong>14</strong></td>
<td>As noted above, but now in the results section, please state exactly and expand which search strategies led to the numbers in the sentence, „the search yielded a total of 2830 retrievals: 987 (34.9%) in Google Scholar and 1843 (65.1%) in SUMSearch. When I search SUMSearch today with „back pain and Guideline“ with no focus or limits, I get &lt;300 citations including all sections of the results. Likewise, please state exactly how the search was performed at Scholar. I assume the numbers you are reporting are not just for back pain, but all nine diseases pooled? If so please clearly state this in both the abstract and the body of the paper.</td>
</tr>
<tr>
<td><strong>15</strong></td>
<td>Similarly, in Table 4, why do the sums for the two search engines both equal 2830? Seems for SUMSearch the total used in the table should be 1843 and 987 for Google.</td>
</tr>
</tbody>
</table>
16) Regarding methods used to search SUMSearch:
   1. The manuscript states „the search was restricted to the category „Practice Guidelines” (NGC and PubMed)“. Did your study use the link provided by SUMSearch to guidelines at PubMed? Or did you use the original research section of SUMSearch. The PubMed practice guidelines section of SUMSearch has not been providing citations for several years. This is the only bug I have not been able to fix and I suspect it may be due to SUMSearch querying PubMed faster than allowed by PubMed. Since SUMSearch can not provide the citations from PubMed, instead it builds a link to search for guidelines at PubMed that the user can click.

   Information added in Methods section, page 6 (new text underlined):

   The search was restricted to the category “Practice Guidelines” (NGC and PubMed). Since the PubMed practice guidelines section of SUMSearch does not provide citations, we used the link provided by SUMSearch to search for guidelines at PubMed.

17). The methods state that DARE was excluded at all stages of the project, but the heading of the final column in Table 1 suggests that DARE was included.

   DARE was initially considered in the preliminary study but then excluded.

   Information added in Methods section: page 6:

   DARE had initially been included in the preliminary study, but as no guidelines were retrieved (only systematic reviews and meta-analyses), it was subsequently excluded from the main study.

   Table 1 refers to the preliminary study.

18) I could not replicate Table 1 and to me the numbers reported do not match the methods cited. When I search SUMSearch for guideline*, I get approximately 29,500 citations only if I sum the results of all sections of SUMSearch. This huge number is mainly due to including the citations from „PubMed (possible systematic reviews)” section – which does not use automated revisions. If I only look at the Guidelines Clearinghouse section and the number of citations when I follow the link built for guidelines at PubMed, I get about 15,000 citations.

   Error in footnote Table 1 corrected (new text underlined).

   †The terms ‘guideline’, ‘practice guideline’, ‘recommendation’ and ‘standard’ were entered into SUMSearch and Google Scholar with the truncation ‘*’, and into Google Scholar as singular and plural terms.

   In the preliminary study, the terms guideline/guidelines/guideline* were entered into SUMSearch. The results for all SUMSearch categories in the retrieval list (e.g. Practice Guidelines NGC etc.) were summed up. The number in Table 1 (340105) is this total number.
<table>
<thead>
<tr>
<th>General comment: due to the changes in both search engines over the past year, the results of this search will not be exactly replicable. Information added in the Methods section (page 5): The study was conducted in October 2005. and in the Discussion (page 9): It should also be noted that since conducting our search, changes have been made to both SUMSearch and Google Scholar. Therefore, the replicability of results may be affected.</th>
</tr>
</thead>
<tbody>
<tr>
<td>19) Regarding the methods used to search Scholar: 1. What was your search method exactly? Did you use the advanced interface and select „in the title“. Then which search box did you put the terms in? When I search Scholar with „allintitle: back pain OR guidelines guideline“ I get 53 results which may be the same as yours with the passage of time and more citations to accrue. Still clarifying whether you used the „with all of the words“ versus „with at least one of the words“ box would help. Did all GLAD terms go into one of these boxes, or did you split and put the disease terms in the „with all of the words“ and the guidelines terms in the „with at least one of the words“. Information added on page 6: We used the advanced interface and selected “in the title“. We entered the combination of each CPG term (single or plural) and each MeSH term into the search box “with all of the words”.</td>
</tr>
<tr>
<td>20) Less important comments: Did the results vary by disease? „Attention Deficit Disorder“ is the most difficult term as is canonical MeSH term is not a phrase used commonly. Did Google have the most trouble with this disease since the study used MeSH terms? The results varied by disease. We have attached a table for information purposes (please see below). However, we think that the discussion of the differences between single results would go beyond the scope of this paper and have therefore not addressed this issue in the manuscript. With regard to the MeSH term “Attention deficit disorder”: Google Scholar did in fact have the most trouble with this disease: it was the only disease for which Google Scholar did not detect a unique and relevant guideline (please see attached table).</td>
</tr>
<tr>
<td>Disease/ MeSH term</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>Obesity</td>
</tr>
<tr>
<td>Osteoporosis</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Parkinson disease</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>Alcoholism</td>
</tr>
<tr>
<td>Depressive disorder</td>
</tr>
<tr>
<td>Schizophrenia</td>
</tr>
<tr>
<td>Attention deficit disorder</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td><strong>Total: All unique and relevant CPGs</strong></td>
</tr>
</tbody>
</table>
| 21) „Meta-analyses and systematic reviews were excluded“. Please clarify – did you include practice guidelines that are based on systematic review of the literature (such as the USPSTF)? | Information added in Methods section (pages 5/6):

*Publications identified as meta-analyses or systematic reviews were excluded. The inclusion of a CPG was independent of the methodology of the guideline development (e.g. CPGs based on systematic reviews were not excluded).* |
|---|---|
| 22) Please clarify (this may be in the manuscript and I missed it) how many of the search results for both search engines that you reviewed for possible guidelines. Did you look at all of the results, or just the first 50, or first 100, etc? | Clarification in Abstract (page 2) and on page 7 (new text underlined):

*To compare GLAD strategies, we used a manual review of all retrievals as a reference standard.*

*After collecting the raw retrievals in SUMSearch and Google Scholar, we defined a reference standard by manually reviewing all retrievals (links) from the retrieval list of each search engine.* |
| 23) Although the authors and myself have described SUMSearch as a meta-search engine, emerging terminology might suggest ‘federated search’ as more appropriate ([http://en.wikipedia.org/wiki/Federated_search](http://en.wikipedia.org/wiki/Federated_search)) as SUMSearch searches databases rather than search engines. However, I admit the meta-search versus federated search distinction is murky. | We agree that the terminology used in the literature to describe search processes is inconsistent. However, we do not think that a further discussion of the terms “federated search” vs. “meta-search” beyond the information presented in the Background section is relevant for the focus of the paper. |
| 24) The description states SUMSearch searches the Merck Manual. The Merck search was removed several years ago. | Information removed from page 3.

This was not clear to us from the information provided in “SUMSearch - More details” [http://sumsearch.uthscsa.edu/MoreDetail.htm](http://sumsearch.uthscsa.edu/MoreDetail.htm) (accessed on 1 March 2007):

*“SUMSearch always searches the following resources: 1. Textbook. The default textbook to search is the Merck Manual. This only currently the only free access medical textbook on the Internet. It is currently offline until summer 1999 while a new edition is prepared.”* |
| 25) A major difference between SUMSearch and the various flavors of Google is that SUMSearch performs Boolean searching with automated revisions of searches; whereas, Google versus uses relevancy with PageRank™. This makes an interesting | We have emphasised this contrast in the Background section in page 4 (new text underlined):

*In contrast to SUMSearch, Google Scholar presents the search results in a ranked list...* |
<table>
<thead>
<tr>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>26) I would very much like to mix the methodologies of SUMSearch and Google into one portal. However, the closing of Google’s SOAP API prevents this (<a href="http://code.google.com/apis/soapsearch/">http://code.google.com/apis/soapsearch/</a>).</td>
</tr>
</tbody>
</table>
| Such a mix of methodologies would be a significant step in search engine development. We addressed this future possibility in the Conclusion on page 12: 

*A further vision for the combination of the advantages of both search principles – the federated search of Google Scholar and the metasearch of SUMSearch – could lead to the development of a ‘Medical Internet Portal’.* |