Reviewer's report

Title: An Overview of the Design and Methods of an Information Retrieval Study

Version: Date: 21 March 2005

Reviewer: Kaveh Shojania

Reviewer’s report:

This group has produced a number of very useful articles and related work products for supporting effective literature searching. The work is commendable and serves an important purpose. The present work provides more methodologic details than allowed in the brief methods sections of the articles published to date.

Major Compulsory Revisions

None

Minor Essential Revisions

p.5 “Each item… were classified” should be “…was classified”

“Data” is plural, so p. 6 “Hand search data was recorded…” should be “data were recorded” and p.9 “These data was organized by…” should be “These data were…”

Discretionary Revisions

1. The extensive iterative automated testing amounts to a derivation set. Was performance for the search strategies re-evaluated on a different set of article (i.e. a validation set?) if not, the methodology runs the risk of over fitting, as in any automated analysis of a database. The authors might comment on this issue (or clarify why it is not an issue, if I have misunderstood).

2. The automated, iterative testing of the search terms is impressive. However, it remains unclear to me if there may be problems related to the requirement that all terms be combined in strings of OR operators. Surely some terms could perform quite well if they were ANDed with something or if they could have a NOT phrase attached.

For instance, suppose one has a string of the form x OR y, but it retrieves too many basic science articles might be improving by changing to x OR y BUTNOT (Physiological Processes [mh] OR Biological Phenomena, Cell Phenomena, and Immunity [mh])

I made up this example for illustrative purposes. The point is that many terms might perform better of they could be combined in different ways with respect to OR, AND, BUTNOT, but the methodology used by the authors seems to allow only the possibility of adding or subtracting to strings of ORed terms.
3. p.8 “Since the primary goal of the Clinical hedges Study was to deliver search strategies which could be used by clinicians…” Although OVID is very common among academic institutions, it is not among the total population of clinicians. It is not clear then why they used OVID instead of PUBMED as the latter would be accessible by a far greater number of clinicians. The authors’ BMJ article on review articles addresses this problem by providing translations for PubMed. Were translation performed for all of the searches and were they evaluated in the same way as described for OVID?

4. Table 2 focuses on whether or not articles were “concerned with understanding human health” but this seems quite vague. This judgment must have been operationalized more specifically. For instance, were articles that did not involve human subjects automatically excluded? Gender and equality issues are mentioned as excluded, but what about the rest of health services research? Quality measurement? Some of these articles would likely be difficult to judge as not “of interest to understanding human health”.

What next?: Accept after discretionary revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I declare that I have no competing interests