Author's response to reviews

Title: Developing Optimal Search Strategies for Detecting Clinically Sound Prognostic Studies in MEDLINE: An Analytic Survey

Authors:

   Nancy L Wilczynski (wilczyn@mcmaster.ca)
   Prof R. BRIAN Haynes (bhaynes@mcmaster.ca)

Version: 2 Date: 10 Mar 2004

PDF covering letter
March 10, 2004

BioMed Central Ltd
Manuscript Submissions

Dear Dr. Gadd:


We are resubmitting the above-noted manuscript revised based on comments from the reviewers. A point-by-point response to the reviewers’ comments identifying corresponding areas of revision in the manuscript follows.

If any additional information is required please let us know.

Yours truly,
Nancy L. Wilczynski
Responses to Reviewer Comments  
March 10, 2004

Reviewer's report

Title: Developing Optimal Search Strategies for Detecting Clinically Sound Prognostic Studies in MEDLINE

Version: 1  Date: 3 March 2004

Reviewer: Susan Wieland

Reviewer's report:

General

The question posed by the authors (what are the best search strategies for retrieving methodologically sound prognostic studies in MEDLINE) is important, and it is valuable for clinicians to be made aware of the underlying research supporting the development of such search strategies. This paper discusses the strategies that the authors used to develop the sensitive and precise search strategies currently found in PubMed Clinical Queries for prognostic studies. However, while this manuscript is readable and almost completely free of technical jargon, it does not provide enough detail about the development of the search strategies.

----------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. In the Methods section (page 6, top of page) and the Results section (page 7, bottom of page) the manuscript is imprecise about the population of studies upon which the search strategies are run. Are the searches tested against all 49,028 studies from the 161 journals, or only the 1,547 studies classified as prognosis studies? This lack of clarity makes it vague whether the research is focused on sorting high from low quality prognostic studies or on sorting high quality prognostic studies from both low quality prognostic studies and all non-prognostic studies. This needs to be clarified in the manuscript.

- the research is focused on sorting high quality prognostic studies from both low quality prognostic studies and all non-prognostic studies
- the search strategies are developed using all 49,028 studies
- in the Results section (page 8, first paragraph) the following two sentences have been added: Search strategies were developed using all 49,028 articles. Thus, the strategies were tested for their ability to retrieve articles about high quality prognostic studies from all other articles, including both low quality prognostic studies and all non-prognostic studies.
2. In the Methods section (page 6, second paragraph), the criteria used for considering a report on prognosis to be methodologically sound are quite reasonable. In the Results section (page 7, last paragraph), only 12% of all prognosis articles were considered methodologically sound. Given the small proportion of prognostic studies that meet criteria for methodological soundness, it would be useful to be given an idea of what characterize most prognostic studies. For example, what is the distribution of methodological flaws in prognostic studies -- did most studies fail to be sound due to poor analysis, or lack of followup? Is there any drawback to excluding methodologically unsound prognostic studies from a search strategy for prognosis? Some reassurance about this issue should be discussed in this manuscript.

- the exact figures for the distribution of methodologic flaws in prognostic studies was not documented
- however, the majority of studies classified as prognosis did not meet the criterion of inception cohort
- in the Result section (page 8, first paragraph) the following text has been added: Most of the studies classified as prognosis did not assemble an inception cohort and thus ‘failed’ to be categorized as methodologically sound.
- the search strategies developed help filter out methodologically unsound prognostic reports for clinicians. This is important because clinicians should make patient care decisions on research with sound research methods
- in the Methods section (page 6, second paragraph) the following sentence has been added: The focus of the strategies is to help clinicians retrieve methodologically sound study reports as patient care decisions should be based on good quality evidence.

3. Finally, the authors have written about a purely methodological search filter. They state in the Discussion section (page 9, second paragraph) that limiting searches by clinical content terms might increase precision, but that an increase in precision cannot be assumed. The authors hint in the last sentence of this paragraph that they may focus on using content specific terms in the next phases of their project. The results to be expected from combining disease-specific terms with the methodologic terms are surely of great interest to clinicians and need to be discussed further in this manuscript.

- combining disease-specific terms with the methodologic terms is currently being addressed in our research unit. The effects on precision will be determined.
- the following two sentences has been added (page 10, end of first paragraph): We are currently testing the methodologic filters by combining them with disease-specific terms in the discipline areas of mental health and infectious disease as well as the disease-specific area of tuberculosis.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. In the Discussion section (page 9, first paragraph) the authors use the term ‘false drops’. This term is jargon and should be replaced with a term that can be understood by the general medical reader.

   - we feel that it is also important to maintain links with the library science community. The term ‘false drops’ has been defined in the paper (page 9, first paragraph of Discussion) as follows: studies that meet criteria but are not retrieved by the strategy.

Discretionary Revisions (which the author can choose to ignore)

1. In the Background section (page 5, second paragraph), the authors discuss how the current study builds upon previous research they have done in this area. The current research relies upon 161 journals (instead of 10 in the authors’ research from the early 1990s, used to develop the current Clinical Queries interface) to develop a gold standard, and upon a broader range of MEDLINE terms to develop the search strategies. In the Methods section (page 6, last paragraph, and page 7, second paragraph), the authors provide information upon the processes used to select the journals and the search terms, and offer to provide a list of the journals and a list of the MeSH terms and textwords upon request. However, this manuscript would be strengthened by presenting more specifics about the journals and the search terms within the text, perhaps in the form of tables highlighting key journals and search terms. This would give readers more confidence in the methods used, and allow replication of the research.

   - to illustrate the scope of the journals included in the hand search of the literature the following sentence (page 7, first paragraph) has been added: Examples of the 161 journal titles included in the hand search are Addiction, Age & Ageing, BMJ, JAMA, Lancet, New Journal of Medicine, Pediatrics, Public Health Nursing, and Stroke.

   - to illustrate the scope of the terms tested the following sentence (page 7, second paragraph) has been added: Examples of the search terms tested are ‘disease attributes’, ‘disease onset’, ‘early onset’, and ‘first diagnosis’, all as textwords; ‘recurrence’, the MeSH term, and the MeSH term ‘mortality’, exploded.

2. In the Methods section (page 7, very top of page) the authors say that inter-rater agreement for application of criteria was very high, and cite a paper from 2001. Do the authors mean to say that they used methods previously described to have high inter-rater agreement, or that the inter-rater agreement for the review of these 161 journal titles was itself 80% beyond chance?
- prior to conducting the review of the 161 journal titles research staff participated in an extensive calibration exercise which resulted in obtaining high inter-rater agreement
- it was necessary to obtain this level of agreement prior to conducting the literature review for this study
- the following clarification has been added (page 7, first paragraph): Research staff were rigorously calibrated prior to reviewing the 2000 literature and inter-rater agreement for application of all criteria exceeded 80% beyond chance [14].

3. In the Methods section (page 7, second paragraph) the authors say that the database was randomly split but they do not state the method used. Do the authors mean that the database was divided in a non-systematic (haphazard) way, or did they use a true random method?

- the database was split using a true random method. Microsoft windows’ random number generator was employed to split the database.
- the following has been added (page 7, second paragraph): The database was randomly split using Microsoft Windows’ random number generator into components of 60% and 40%.

4. In the Discussion section (page 9, last paragraph), the authors mention that the best performing strategy for sensitivity was the same as the strategy they reported in 1991, but the references are to articles in 1993 and 1994. Is it possible that the strategy was developed in 1991 but reported in 1993 and 1994, or is there a reference from 1991?

- the strategies developed using 1991 literature were published in 1993 and 1994
- the literature was hand searched in 1991 after which the databases are formed and the search strategies were developed
- no modifications to the paper

5. In the Discussion section (page 9, last paragraph), the authors do not discuss whether their 1991 results differ from the current results for specificity, precision, or accuracy. Were there any differences in these areas? If so, what is the implication of these differences?

- the following two sentences has been added (page 10, second paragraph): The operating characteristics for this strategy were similar when comparing the performance in 1991 and 2000 for sensitivity (92% vs 90% [2000]) and specificity (73% vs 80% [2000]) but were quite different for precision (11% vs 2% [2000]). This difference in precision is to be expected given the increased size and diversity of the database in 2000.
6. This research was developed using Ovid’s search engine syntax. In the Discussion section (page 10, third paragraph) and the footnotes to Table 1 and Table 2 (pages 14 and 15), the authors mention that the strategy is contained in PubMed’s Clinical Queries. However, the PubMed syntax is not presented anywhere in the manuscript. Given that PubMed is a free resource that is available to anyone with an internet connection, presenting the PubMed syntax would make the research more replicable. The authors should briefly mention why they chose to develop their strategy using Ovid syntax rather than PubMed syntax. Since Ovid and PubMed do not contain exactly the same information, it would also be interesting to know how results from searches run in Ovid might compare with searches run in PubMed.

- search strategies were developed in Ovid due to convenience and were translated into PubMed syntax by the National Library of Medicine
- the search strategies have been updated in the Clinical Queries interface of PubMed and the translations can be viewed by going to the “filters table” link
- the following sentence has been added (page 11, third paragraph): The translation from Ovid to PubMed syntax was done by staff of the National Library of Medicine, and compared for performance by the senior author (RBH).

7. In the Discussion section (page 10, third paragraph) the authors mention that the Clinical Queries interface of MEDLINE will be updated to reflect the new search strategies. Because the authors already discuss (page 9, last paragraph) that their best strategy for sensitivity is the same as that reported earlier, and they do not mention how the best strategy for specificity has changed, it is not clear from the Discussion section how Clinical Queries will be updated. The authors should clarify this point, and also discuss whether Clinical Queries will include a strategy optimizing both sensitivity and specificity, as presented in Table 2. Finally, it is not clear from looking at Clinical Queries whether the update of Clinical Queries has already happened; if the update has already happened, the authors should mention this, and if the update has not already happened but the authors have any indication of the expected date of the update, the authors should give this date.

- since we submitted the paper, the NLM has updated the Clinical Queries interface on PubMed to include our new strategies
- the National Library of Medicine is waiting for this work to be published prior to making an official announcement
- once the work has been published links to the new publications will be provided on the Clinical Queries screen
- the paper has been modified as follows (page 11, third paragraph): The National Library of Medicine (NLM) has updated the Clinical Queries interface of MEDLINE to reflect our new strategies for maximizing sensitivity and maximizing specificity (http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html). The translation from Ovid to PubMed syntax was done by staff of the National Library of Medicine, and compared for performance by the senior author (RBH). SKOLAR
has also implemented our high specificity strategies (personal communication, Howard Strasburg, Ovid Technologies Inc) and both sensitive and specific strategies will be implemented for the main Ovid MEDLINE search engine (Timothy Roberts, Ovid Technologies Inc, personal communication).

8. The first column of Table 1 (page 14) is laid out in a confusing manner so that it appears that there are three terms in the first column. If it is possible, the term should be contained on a single line.

   - Table 1 has been reformatted to show the search term on a single line

9. The text does not explain why the best single search term presented in Table 1 (page 14) is of interest. When would a searcher want to use the single search term rather than the various combinations of terms? Please discuss this further in the text.

   - in the first paragraph of the Results section (page 8) the following two sentences have been added: A clinical end-user of MEDLINE may find that searching with this single term is worthwhile when using interfaces that do not store the more complex search strategy. This single term is easy to remember and will provide the best retrieval compared with any other single methodologic search term.

Which journal?: Appropriate or potentially appropriate for BMC Medicine: an article of outstanding merit and interest in its field

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests: None
Reviewer's report

Title: Developing Optimal Search Strategies for Detecting Clinically Sound Prognostic Studies in MEDLINE

Version: 1 Date: 22 February 2004

Reviewer: Petteri Sjögren

Reviewer's report:

General
This paper addresses an important issue of developing effective search strategies for locating clinically relevant studies in MEDLINE for clinical studies. I consider this article appropriate for publication in BMC Medicine after a minor revision. The question posed by the authors is new and well defined. The authors offer lists of all the search terms and the included papers on requests which is acceptable in this case due to a large number of search terms and articles included. The results seem to be sound and validated. The manuscript adheres to relevant standards for reporting and data deposition. However, some Minor Discretionary Revisions are suggested below. The discussion section would be improved by discussing findings from previous comparable studies from other groups (see Minor Discretionary Revisions). The conclusion is sound and supported by the findings. The title and abstract are adequate and the manuscript is well written.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. In the results section as well as in Table 1 and 2, it is stated that the differences were not statistically significant. However, the statistical methods are not described in the methods section, nor are the p-values, or the p-levels below which the differences are considered statistically significant. These issues need to be clarified in the final manuscript.

- the statistical test used to determine statistical significance was the iterative method of Miettinen and Nurminen for two independent binomial proportions
- providing the 95% label for the CI indicates that the p-level for determining statistical significance was at the 0.05 level, 2-tailed
- reporting the 95% CIs is superior to reporting the p-values as it provides more information. If the CI crosses 0 the difference is not statistically significant.
- the following has been added to the footer of Tables 1 and 2: †Diff = Difference, comparing the development and validation data sets using the iterative method of Miettinen and Nurminen for two independent binomial proportions.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. On page 6, please check if it would be appropriate to change “Science Citation Index Impact Factors” to “Journal Citation Reports® (Institute for Scientific Information)”, as the correct source?

- Science Citation Index Impact Factors has been changed to Science Citation Index Impact Factors provided by the Institute for Scientific Information (page 6, last paragraph)

2. In Table 2: A “Difference” value is missing in the box: “Best Optimization of Sensitivity & Specificity” / “Precision (%) Development Validation”. The value “-0.1” seems to be is missing before the CI values?

- -0.1 has been added to Table 2

Discretionary Revisions (which the author can choose to ignore)

1. In the methods section on page 6, lines 5-8: “Sensitivity and specificity are not affected by the proportion of high quality … “ could be moved to the discussion and expanded in a simple way to make this matter clearer for readers who are not quite so familiar with database search strategies.

- the sentence in question has been moved to the discussion section of the paper (page 9, second paragraph of the Discussion)

2. In the results section on page 8, the last three lines: “In most instances the results were trivially different …”. This sentence should be moved to the discussion or rewritten in an objective way (the term “trivially” seems too subjective).

- the sentence in question (page 9, first paragraph) has been changed to: In most instances the differences in results when comparing the performance in the development and validation databases were nonexistent or very small.

3. The discussion section would be improved by a discussion around earlier findings from previous comparable studies from other research groups. As written in the introduction “Little work has been done in the areas of prognosis”. Are there any relevant studies from other groups?.

- to our knowledge, in the area of prognosis, no other empirical study has been done except our own previous study

- the sentence in the Background section of the paper (page 5, first paragraph) has been changed to: Little work has been done in the area of prognosis and to our knowledge our previous study (11, 12) was the only one in which search strategies for prognosis were empirically tested.
Which journal?: Appropriate or potentially appropriate for BMC Medicine: an article of importance in its field

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests: None