

Finding RCTs in MEDLINE:

a new search filter for busy researchers

Julie Glanville¹, Carol Lefebvre², Jeremy Miles³, Janette Camosso-Stefinovic⁴

¹ Centre for Reviews and Dissemination, University of York, ² UK Cochrane Centre, Oxford ³ Dept Health Sciences, University of York

⁴ Master's Degree Student, Department of Information Science, Loughborough University (2002)

Background

Researchers and busy health care professionals want to identify reliable evidence on the effectiveness of health care interventions, especially evidence from systematic reviews of randomized controlled trials (RCTs), or in the absence of such systematic reviews, from RCTs themselves. Over the last ten years the number of RCTs, which can be easily and accurately identified in MEDLINE, has risen from 20,000 in 1993 to more than 190,000 in August 2004, an increase of 170,000, of which only 115,000 were published since 1993. Researchers and health care professionals have used search filters to identify RCTs in MEDLINE and one of the most widely used is the Cochrane Highly Sensitive Search Strategy (Cochrane HSSS). The Cochrane HSSS is also used by the Cochrane Collaboration to identify records for retagging as RCTs as part of the MEDLINE retagging exercise

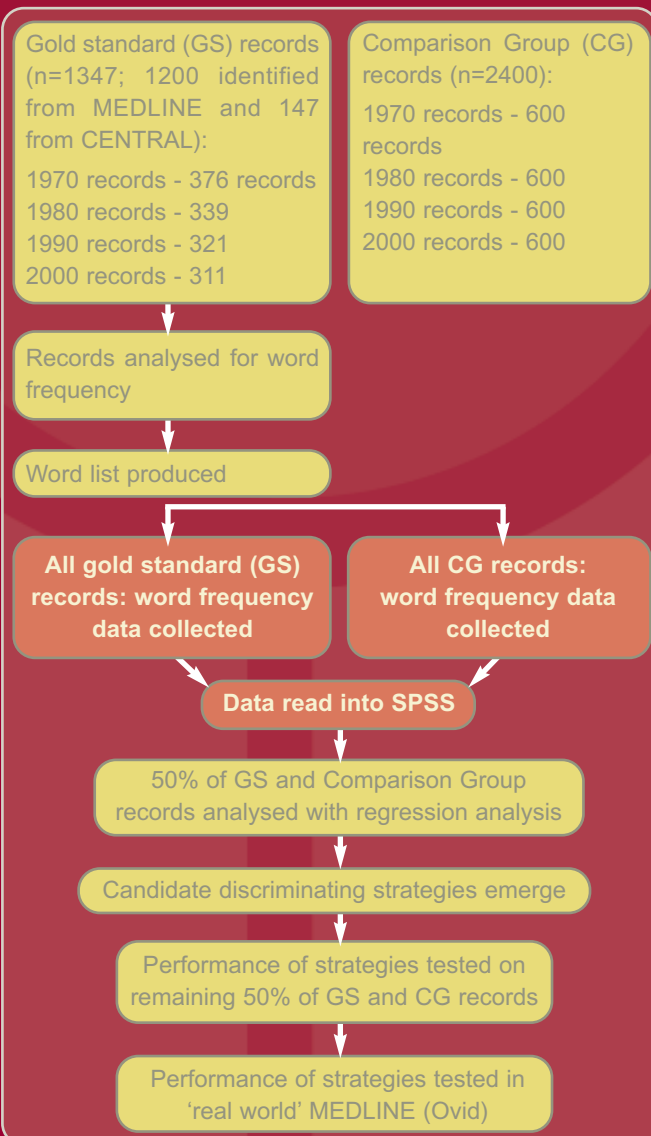
Our research questions

- Is the Cochrane HSSS effective as a tool to retrieve RCTs in MEDLINE?
- By analysis of RCT records in MEDLINE can we identify a brief set of search terms that offers busy researchers an adequately sensitive and precise strategy to identify RCTs in MEDLINE?

Methods

We adapted methods for search filter design used previously by some of the research team members (White et al, 2001)¹.

Box 1: Methods



We identified a gold standard (GS) of randomly selected records of RCTs from MEDLINE from four publication years (1970, 1980, 1990, 2000) and augmented this with a further set of RCT records (for the same years) from the Cochrane Central Register of Controlled Trials (CENTRAL) (n=1347). We compiled a randomly selected comparison set of records of non-trials (n=2400) from MEDLINE for the same four years. The Comparison Group (CG) records contained no Clinical Trial Publication Types.

We identified words occurring in at least 1% of the GS records using WordStat and SimStat content analysis and statistical software. We analysed separately:

- Words in title and/or abstract
- Phrases in title and/or abstract
- MeSH headings
- Publication Types
- Check tags

We collected information on the occurrence of those terms in both the GS and the CG records and loaded the data into SPSS. We analysed 50% of the records (GS plus CG) using logistic regression to identify the search terms that best discriminated RCTs from non-trials.

We tested the performance of these discriminating terms on the remaining 50% of the GS and CG records, and then tested the performance of several strategies (including the Cochrane Highly Sensitive Search Strategy) in terms of their sensitivity in identifying RCTs already indexed as such and un-indexed RCTs, and their precision in 'real world' MEDLINE (Ovid). We also searched MEDLINE using various strategies to identify records of RCTs in breast cancer published in 2002 (not at the time assessed by the Cochrane retagging exercise) and then 'handsearched' the records to see which were (un-indexed) records of controlled trials.

Results

The most discriminating search terms to identify records of RCTs from non-trial records were two Publication Types in combination:

- RANDOMIZED CONTROLLED TRIAL.PT. OR CLINICAL TRIAL.PT.

We handsearched 12,255 records produced by a combination of the candidate strategies referred to above to see if the strategies identified untagged/un-indexed RCTs/CCTs in breast cancer published in 1960, 1970, 1980 and 2000 (these years had already been assessed by the Cochrane retagging exercise). As expected we found few untagged RCTs/CCTs (4 RCTs and 50 CCTs) giving an overall poor precision rate (0.44%).

The strategies were also tested to see how well they could identify potential RCTs/CCTs in breast cancer in MEDLINE, published in 2002. 2002 was, at the time of our research, untagged as part of the Cochrane re-tagging project. Using the new strategies and the Cochrane HSSS we identified 53 extra reports of RCTs in breast cancer for 2002, and our best strategy had a precision of 2.54%.

Acknowledgments

We are grateful for the assistance of Ann Eisinga and Teresa Clarke.

Disclaimer: The views expressed represent those of the authors and are not necessarily the views or the official policy of the UK Cochrane Centre or the Cochrane Collaboration.

Discussion

Using both Publication Types (RCT OR CT) seems counter-intuitive because all MEDLINE records indexed with RANDOMIZED CONTROLLED TRIAL as a Publication Type should also be indexed with CLINICAL TRIAL as a Publication Type. However, 59 of the gold standard records downloaded from Ovid and analysed in 2002/3 had RANDOMIZED CONTROLLED TRIAL (Publication Type) but no matching CLINICAL TRIAL (Publication Type). This inconsistency has since been corrected and when we searched Ovid in May 2004 we found no such inconsistencies. If the records were reanalysed today we would expect that our analysis would indicate that both search terms were NOT necessary.

Adding in other search terms that the regression analysis suggested would be discriminating in terms of identifying RCTs, in addition to the 2 Publication Types above, produced very low numbers of extra untagged/un-indexed trials in years where Cochrane retagging had taken place.

In the test searches in MEDLINE 2002 records (prior to the Cochrane retagging exercise), there is an indication that the PT search is sensitive but that sensitivity can be improved by the addition of search terms suggested by the regression analysis. This strategy is being prepared for publication.

Limitations

- There are still elements of subjectivity in the methods, which we would like to reduce.
- SPSS analysis presents terms with weightings attached and so far we are unable to implement weightings attached to search terms in real world interfaces.
- The generalizability of the results to other health care areas is currently being tested as part of this project.

Key Messages

The potential good news for busy searchers:

- A subject search of the Cochrane Central Register of Controlled Trials (CENTRAL) plus a top up search using RANDOMIZED CONTROLLED TRIAL.PT. in MEDLINE is very sensitive in terms of finding RCTs

The potential good news for researchers pressed for time:

- A search of CENTRAL plus use of CLINICAL TRIAL.PT. in MEDLINE will give a highly sensitive search with manageable precision
- NB don't forget to remove animal studies in MEDLINE: (animals/ not (human/ and animals/))

.PT. denotes a Publication Type search in Ovid

Next steps

- Further testing in other subject areas to see if the suggestions are robust (in progress)
- EMBASE RCT filter derivation using the same approach
- Development of regression analysis software that will allow user specification of preferences for levels of sensitivity and precision

References

White, V., et al. (2001). "A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy." *Journal of Information Science* 27(6): 357-370.

For further details contact Julie Glanville jmgl@york.ac.uk