

JAMIA
Letters to the Editor

Comment ■

Response to Corrao et al.: Improving Efficacy of PubMed Clinical Queries for Retrieving Scientifically Strong Studies on Treatment

NANCY L. WILCZYNSKI, MSc, R. BRIAN HAYNES, MD, PhD

■ J Am Med Inform Assoc. 2007;14:247–248. DOI 10.1197/jamia.M2297.

A recent study by Corrao et al.¹ assessed the retrieval power of the narrow (specific) search strategy for therapy available on the Clinical Queries screen in PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml>) which was developed by our research group at McMaster University.² They compared its retrieval power with a modified search string that included the Britannic English term “randomised”.

We welcome such an analysis of our work and encourage researchers to continue to investigate improved ways of searching in MEDLINE. Corrao et al. stated that the narrow therapy search strategy available on Clinical Queries, (*randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract])*), may introduce bias because it may cut off studies about therapy in which “randomized” was exclusively written as “randomised”. Thus, they modified the search string to include “randomised,” (*randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] OR randomised[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract])*). Corrao et al. hypothesized that their modified search strategy would improve recall (sensitivity) of the PubMed strategy and tested the 2 search strategies joining them to 4 terms concerning topics of broad interest: hypertension, hepatitis, diabetes, and heart failure.

Corrao et al. reported that the word “randomised” improved total citation retrieval (1% for total or indexed for MEDLINE citations), that the main gain was achieved searching not-indexed-for-MEDLINE or in-process citations, and that the greatest gain was for the topic area of hypertension. We undertook a similar investigation using hypertension as the disease content area.

We conducted a search using the PubMed strategy and Corrao’s strategy ANDed with (“premedline[sb] OR publisher[sb] OR pubmednotmedline[sb]) (as reported in the paper by Corrao et al.) AND hypertension on October 19, 2006. Eighteen citations were retrieved with the PubMed string and 27 with the Corrao string. In reviewing the 9 citations that were retrieved by the Corrao strategy but not by the PubMed strategy we found that although 8 of the 9 studies retrieved were randomized controlled trials (RCTs), 4 were not directly relevant to the topic area of hypertension leaving a net gain of 4 articles, 2 of which were greater than 5 years old.

Additionally, we conducted a second search because the term “premedline[sb]”, although reported in the paper by Corrao et al., is no longer an official search term for PubMed; “in process[sb]” is the proper PubMed term.³ A search using the PubMed strategy and Corrao’s strategy ANDed with (in process[sb] OR publisher[sb] OR pubmednotmedline[sb]) AND hypertension on October 19, 2006 yielded 47 citations with the PubMed string and 62 with the Corrao string. In reviewing the 15 citations that were retrieved by the Corrao strategy but not by the PubMed strategy we found that although 12 of the 15 studies retrieved were RCTs, 7 were not directly relevant to the topic area of hypertension leaving a net gain of 5 articles, 2 of which were greater than 5 years old.

The retrieval for the PubMed strategy ANDed with hypertension in the full MEDLINE database was just over 11,000 citations on October 19, 2006 of which about half would be RCTs. Thus, the net effect of including “randomised” in the PubMed strategy would be tiny for retrieving trials (and not likely make a difference to the overall message that one would get from the indexed literature) and it would adversely affect precision (retrieval even in the pre-indexed file includes several false positives).

To determine the effect that Corrao’s strategy would have on the indexed literature we tested it in our Clinical Hedges Database (the database used to derive and validate the PubMed therapy strategies). We found that sensitivity (93.1%), specificity (97.4%) and accuracy (97.3%) of the Corrao strategy were identical to the PubMed strategy but that precision dropped slightly from 54.4% to 53.8%.

Based on this analysis we would not recommend the addition of “randomised” to the narrow (specific) search strategy

Affiliation of the authors: Health Information Research Unit, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Correspondence and reprints: Nancy L. Wilczynski, M.Sc., Health Information Research Unit, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada; e-mail: <wilczyn@mcmaster.ca>.

Received for review: 9/29/2006; accepted for publication: 12/11/2006.

for therapy available on the Clinical Queries screen in PubMed.

References ■

1. Corrao S, Colomba D, Arnone S, Argano C, Di Chiara T, Scaglione R, Licata G. Improving efficacy of PubMed clinical queries for retrieving scientifically strong studies on treatment. *J Am Med Inform Assoc* 2006;13:485-7.
2. Haynes RB, McKibbin KA, Wilczynski NL, Walter SD, Werre SR; Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ* 2005;330:1179-82.
3. National Library of Medicine. NLM Technical Bulletin. Changes to PubMed for 2001. Available at: http://www.nlm.nih.gov/pubs/techbull/jf01/jf01_pubmed_2001.html. Accessed October 19, 2006.