Evidence Based Library and Information Practice

Evidence Summary

Statistical Measures Alone Cannot Determine Which Database (BNI, CINAHL, MEDLINE, or EMBASE) Is the Most Useful for Searching Undergraduate Nursing Topics

A Review of:
Stokes, P., Foster, A., & Urquhart, C. (2009). Beyond relevance and recall: Testing new user-centred measures of database performance. *Health Information and Libraries Journal, 26*(3), 220-231.

Reviewed by:
Giovanna Badia
Librarian, Royal Victoria Hospital Medical Library
McGill University Health Centre
Montreal, Quebec, Canada
Email: giovanna.badia@mail.mcgill.ca

Abstract

Objective – The research project sought to determine which of four databases was the most useful for searching undergraduate nursing topics.

Design – Comparative database evaluation.

Setting – Nursing and midwifery students at Homerton School of Health Studies (now part of Anglia Ruskin University), Cambridge, United Kingdom, in 2005-2006.

Subjects – The subjects were four databases: British Nursing Index (BNI), CINAHL, MEDLINE, and EMBASE).

Methods – This was a comparative study using title searches to compare BNI (British Nursing Index), CINAHL, MEDLINE and EMBASE.

According to the authors, this is the first study to compare BNI with other databases. BNI is a database produced by British libraries that indexes the nursing and midwifery literature. It covers over 240 British journals, and includes references to articles from health sciences journals that are relevant to nurses and midwives (British Nursing Index, n.d.).

The researchers performed keyword searches in the title field of the four databases for the dissertation topics of nine nursing and midwifery students enrolled in undergraduate dissertation modules. The list of titles of

journals articles on their topics were given to the students and they were asked to judge the relevancy of the citations. The title searches were evaluated in each of the databases using the following criteria:

- precision (the number of relevant results obtained in the database for a search topic, divided by the total number of results obtained in the database search);
- recall (the number of relevant results obtained in the database for a search topic, divided by the total number of relevant results obtained on that topic from all four database searches);
- novelty (the number of relevant results that were unique in the database search, which was calculated as a percentage of the total number of relevant results found in the database);
- originality (the number of unique relevant results obtained in the database for a search topic, which was calculated as a percentage of the total number of unique results found in all four database searches);
- availability (the number of relevant full text articles obtained from the database search results, which was calculated as a percentage of the total number of relevant results found in the database);
- retrievability (the number of relevant full text articles obtained from the database search results, which was calculated as a percentage of the total number of relevant full text articles found from all four database searches);
- effectiveness (the probable odds that a database will obtain relevant search results);
- efficiency (the probable odds that a database will obtain both unique and relevant search results); and
- accessibility (the probable odds that the full text of the relevant references obtained from the database search are available electronically or in print via the user's library).

Students decided whether the search results were relevant to their topic by using a "yes/no" scale. Only record titles were used to make relevancy judgments.

Main Results – Friedman's Test and odds ratios were used to compare the performance of BNI, CINAHL, MEDLINE, and EMBASE when searching for information about nursing topics.

These two statistical measures demonstrated the following:

- BNI had the best average score for the precision, availability, effectiveness, and accessibility of search results;
- CINAHL scored the highest for the novelty, retrievability, and efficiency of results, and ranked second place for all the other criteria;
- MEDLINE excelled in the areas of recall and originality, and ranked second place for novelty and retrievability; and
- EMBASE did not obtain the highest, or second highest score, for any of the criteria.

Conclusion – According to the authors, these results suggest that none of the databases studied can be considered the most useful for searching undergraduate nursing topics. CINAHL and MEDLINE emerge as consistently good performers, but both databases are needed to find relevant material on a topic.

Friedman's Test clearly differentiated between the databases for the accessibility of search results. Odds ratio testing may assist librarians to make decisions about database purchases. BNI scored the highest for availability of results and CINAHL ranked the highest for retrievability. Statistical measures need to be supplemented with qualitative data about user preferences in order to determine which database is the most useful to our users.

Commentary

This study contributed to the existing literature in that it was the first study to compare BNI, CINAHL, MEDLINE, and EMBASE, and the first one to combine the novelty, originality, availability, and retrievability of search results with the traditional testing criteria of precision and recall to compare database performance. Its findings confirmed what is already known, i.e., "that searching a single database is likely to miss relevant articles, and that some databases may be in general good performers" (pp. 230).

The statistical measures used for comparative database evaluation, i.e., Friedman's Test and odds ratios, could not determine which database was the most useful. This reviewer questions whether odds ratio was an appropriate statistical test to compare BNI, CINAHL, MEDLINE, and EMBASE, since the authors state that "the odds ratio is comparing each database individually against the pool of data; it does not compare the four databases with each other" (pp. 229-230).

The authors also suggest that odds ratio may assist in the selection of databases to purchase, but do not explain how.

The study would have benefited from including a brief description of Friedman's Test and odds ratios, as well as an explanation of how the data from both tests were combined. A table with the raw data from the searches could also have included in the article itself. Unfortunately, the appendix containing the search data is no longer available on the publisher's website. The missing appendix and the lack of sufficient explanatory details in the article make it difficult to replicate, or completely understand, the study's research methodology.

It is also difficult to generalize the study's findings due to: its small sample size (i.e., nine students' topics); the use of keyword searching in the title field to obtain relevant results, which may not be a user's typical searching behaviour; and the use of database testing criteria that are dependent on an individual library's subscriptions rather than on database search performance (i.e., the use of the availability, retrievability, and accessibility criteria).

Despite its weaknesses, this study reminds librarians that precision and recall are not the only criteria that should be used to measure the performance of a database.

References

British Nursing Index (n.d.). *About BNI.* Retrieved 20 Feb. 2011 from http://www.bniplus.co.uk/about_bni.html