

# **EXECUTIVE SUMMARY**

## **Background**

Health care decision-makers and health care professionals need efficient access to the best evidence. Health technology assessment (HTA) agencies and guideline developers need access to evidence from economic evaluations to identify the cost-effectiveness of a technology and to inform models.

Access to economic evaluations has improved with the development of the National Health Service Economic Evaluation Database (NHS EED) and the Health Economic Evaluations Database. Despite the availability of these resources, however, there are still issues with currency because of lags between the time when studies are identified and the time when full abstracts are published in the databases. This means that large biomedical databases such as EMBASE and MEDLINE still need to be searched for economic evaluations. Data are lacking on how well published search filters perform in finding economic evaluations in the databases.

Our objective was to develop search filters to identify economic evaluations in the MEDLINE and EMBASE databases that would maximize sensitivity and achieve levels of precision to meet the needs of health technology assessment researchers. This project was also designed to obtain data on the relative performance of new and published search filters.

## **Method**

A gold standard set of economic evaluations was identified from the NHS EED. Records were selected from 2000, 2003, and 2006. These records were then identified in MEDLINE and EMBASE. A comparator set of non-economic evaluation records was obtained from MEDLINE and EMBASE using random selection from among those records that were retrieved using the NHS EED search filter. A 50% random sample of the gold standard records was analyzed to identify the occurrence of terms in the title, abstract, subject headings, and publication type fields. The terms that met a range of occurrence levels were identified. Data on the presence or absence of those terms in all gold standard and comparator records were analyzed using classification trees. Classification trees were developed for the records from the MEDLINE and EMBASE databases. For each database, 50% of the data were used to develop the tree (the training set), and 50% of the data were used to test the effectiveness of the tree (the validation set).

The candidate search filters that were identified from the search analysis, other published and unpublished filters, and combinations of some of these filters were tested in the Ovid interface to MEDLINE and EMBASE. Published and unpublished filters were identified from the InterTASC Information Specialists' Sub-Group search filter website, provided by the Canadian Agency for Drugs and Technologies in Health (CADTH), and provided by the reviewers of a first draft of this report. The filters that were tested were produced by CADTH, Emory University (the Grady EBM filter), McKinlay et al., NHS EED, National Health Service Quality Improvement Scotland (NHS QIS), Royle and Waugh, Sassi et al., the Scottish Intercollegiate Guidelines Network, and Wilczynski et al.

## **Results**

A total of 2,070 full economic evaluations were identified from NHS EED for 2000, 2003, and 2006. Overall, 1,957 of these had corresponding records in MEDLINE and 1,876 had corresponding records in EMBASE. These two sets of records were the gold standards. The MEDLINE comparator set of non-economic evaluation records comprised 4,136 records and the EMBASE comparator set comprised 3,750 records.

After analysis of the occurrence of terms in the records, 347 terms (single words, subject headings, and publication types) were selected for analysis from the MEDLINE records, and 528 terms were selected for analysis from the EMBASE records. Eight candidate MEDLINE filters (MEDLINE A to MEDLINE H) and eight candidate EMBASE filters (EMBASE A to EMBASE H) emerged from the analysis. All candidate filters, and 13 published and unpublished filters were tested in the Ovid interface to the databases.

The MEDLINE filters with high sensitivity (more than 0.99) were the NHS QIS full and brief filters, NHS EED, and Royle and Waugh. The NHS EED filter had the highest precision (0.04) among these four filters. Achieving higher levels of precision thereby reduced sensitivity. Combining the NHS EED filter (using “AND”) with the MEDLINE D filter achieved more than 0.092 precision with more than 0.943 sensitivity. The Wilczynski best optimization of sensitivity and specificity filter achieved 0.093 precision with 0.923 sensitivity.

The MEDLINE G filter had the maximum precision (0.72 sensitivity and 0.257 precision). The Emory University Grady filter best met the objective of good precision with a sensitivity greater than 0.80 (0.845 sensitivity and 0.133 precision).

Four EMBASE filters (NHS QIS, CADTH, Royle and Waugh, and NHS EED) all had greater than 0.99 sensitivity (the precision ranged between 0.015 and 0.029). The highest precision, with more than 0.90 sensitivity (sensitivity 0.931 and precision 0.133) was achieved by the EMBASE G filter combined with the NHS EED filter.

The EMBASE H filter had the highest precision (0.716 sensitivity and 0.266 precision). The EMBASE G filter combined with the NHS EED filter best met the objective of good precision with a sensitivity of more than 0.80 (0.931 sensitivity and 0.133 precision).

## **Discussion**

This research provides new performance data on published search filters used for the identification of economic evaluations in MEDLINE and EMBASE. Many publicly available filters have not been validated or compared.

Filters that are produced by NHS EED, NHS QIS, and Royle and Waugh continue to perform with high sensitivity in MEDLINE. None of the published or new filters could meet the objectives of high sensitivity (more than 0.95) with a precision of 0.20, or lower sensitivity (greater than 0.79) and enhanced precision (at least 0.50).

The search filters developed for this project were discriminating in the analysis. When tested in the Ovid interface, however, they did not perform as well. This indicates that the text words and

indexing terms in economic evaluation records do not discriminate evaluation records from other records that deal with economic issues in health care. The text words with high sensitivity for identifying economic evaluations (such as “cost” and “economics”) are used in many contexts other than economic evaluations. Therefore, maximizing precision remains a challenge.

This analysis shows that it is still difficult to efficiently identify economic evaluations using the indexing terms applied by the database producers. Even though there are indexing terms available that are specific to economic evaluations (such as “Cost-benefit Analysis/” in MEDLINE), they do not seem to be assigned to all relevant records. This information may help database producers make research evidence more accessible.

Further analysis of the data and gold standard records in this report may lead to the improved filters. Additional analysis could focus on the performance of lower frequency terms (below the selected cut-off frequency that was used in this project) and the analysis of phrases and terms in close proximity in the title and abstracts of records.

The filters have been tested in Ovid MEDLINE and EMBASE, but they require more extensive validation in other gold standard sets of records.

## **Conclusions**

This research shows that searchers have several sensitive filters to use to identify economic evaluations in MEDLINE: NHS QIS (full and brief), NHS EED, and Royle and Waugh. Searchers may select filters based on highest precision (NHS EED) or conciseness (Royle and Waugh). Increased precision can be achieved by choosing filters with lower sensitivity. All the filters for EMBASE should be used without exclusion search lines that remove specific publication types and animal studies.

For searchers conducting scoping studies or rapid reviews, this research identified filters that offer higher levels of precision: the new MEDLINE G filter (more than 0.25 precision) and the new EMBASE H filter (more than 0.26 precision).

Searchers now have new comparative information on the performance of a range of filters, which can assist during the planning of the workload for technology assessments.