Method

WHICH DATABASES SHOULD BE USED TO IDENTIFY STUDIES FOR SYSTEMATIC REVIEWS OF ECONOMIC EVALUATIONS?

Mick Arber
York Health Economics Consortium (YHEC)
mick.arber@york.ac.uk

Julie Glanville
Jaana Isojarvi
Erin Baragula
Mary Edwards
Alison Shaw
Hannah Wood
York Health Economics Consortium (YHEC)

Objectives: This study investigated which databases and which combinations of databases should be used to identify economic evaluations (EEs) to inform systematic reviews. It also investigated the characteristics of studies not identified in database searches and evaluated the success of MEDLINE search strategies used within typical reviews in retrieving EEs in MEDLINE.

Methods: A quasi-gold standard (QGS) set of EEs was collected from reviews of EEs. The number of QGS records found in nine databases was calculated and the most efficient combination of databases was determined. The number and characteristics of QGS records not retrieved from the databases were collected. Reproducible MEDLINE strategies from the reviews were rerun to calculate the sensitivity and precision for each strategy in finding QGS records.

Results: The QGS comprised 351 records. Across all databases, 337/351 (96 percent) QGS records were identified. Embase yielded the most records (314; 89 percent). Four databases were needed to retrieve all 337 references: Embase + Health Technology Assessment database + (MEDLINE or PubMed) + Scopus. Four percent (14/351) of records could not be found in any database. Twenty-nine of forty-one (71 percent) reviews reported a reproducible MEDLINE strategy. Ten of twenty-nine (34.5 percent) of the strategies missed at least one QGS record in MEDLINE. Across all twenty-nine MEDLINE searches, 25/143 records were missed (17.5 percent). Mean sensitivity was 89 percent and mean precision was 1.6 percent.

Conclusions: Searching beyond key databases for published EEs may be inefficient, providing the search strategies in those key databases are adequately sensitive. Additional search approaches should be used to identify unpublished evidence (grey literature).

Keywords: Cost-benefit analysis, Information storage and retrieval, Databases bibliographic

In the context of health care, economic evaluation (EE) can be defined as the analysis of the costs and effects of alternative interventions in a defined population (1). It is an important element of decision making about reimbursement or implementation of interventions in many countries (2–6). Systematic reviews (SRs) of EEs can play a key role in this process. Conducting SRs requires significant resources and the search approaches used, including the choice of databases to search, can impact on these resources (7). Ideally, researchers need to identify as many relevant records as possible, with maximum efficiency. The number of databases searched is a key factor in achieving an efficient approach. To inform the selection of databases for an efficient SR of EEs, evidence is needed on the yield of specific databases and database combinations.

Few studies have previously investigated this topic (8–11). Three are over 10 years old (8;10;11), and predate the 2015

closure of two key databases that specifically indexed EEs: NHS Economic Evaluation Database (NHS EED) (freely available, closed to new records) and Health Economic Evaluations Database (HEED) (subscription resource, no longer available). Royle and Waugh (10) and Alton et al. (8) both identified NHS EED as a key source for the identification of EEs, and its use is recommended in methods guidance (12;13). The importance of NHS EED is also reflected in SR practice. We previously reported that 79 percent of a sample of published SRs of EEs searched NHS EED, the second most frequently searched source after MEDLINE (14). In the context of these closures, the appropriate selection of databases and the quality of search strategies will become increasingly important (15;16). For studies published from 2015, researchers can no longer rely on searches of NHS EED or HEED to compensate for a sub-optimal selection of databases or inadequate search strategies in other databases.

More recent research related to the choice of search sources for EEs is available. Thielen et al. (16) make recommendations on the identification of studies for SRs of economic evidence

for guideline development, but the selection of databases is based on a summary of recommendations and research predating the closure of NHS EED and HEED. A 2016 bibliometric analysis of the yield of fourteen databases for a reference set of EEs (9), reported that a combination of Scopus, MEDLINE and Global Health identified 91 percent of the EEs (17). Rather than identifying a reference set of EEs through a hand-search of relevant journals, or through the lists of included studies in SRs of EEs, the authors used a set of focused database searches developed specifically for the study. These searches lacked the sensitivity of established search filters for EEs. These limitations may mean that all EEs available to be found in a database were not retrieved by the authors' searches, impacting on the reliability of their conclusions on database yield (18).

In light of the need for updated guidance on the most appropriate information sources to identify EEs, this study aims to provide further evidence on the relative yield of databases (both alone and in combination) to inform database choice for SRs of EEs. We also investigated the characteristics of studies not retrieved from the databases.

The inclusion of a record in a database does not mean that it will be identified by a search strategy; recommendations on resources to search lack value if common search practices mean that relevant records are not retrieved from these resources. Therefore, we also evaluated the quality of the search strategies used in recent reviews of EEs, using the performance of the MEDLINE strategy to assess if strategies were sufficiently sensitive.

OBJECTIVES

The study objectives were to:

1. Assess the yield of nine databases as sources of EEs;
2. Identify the most efficient combinations of databases to search when conducting a SR of EEs;
3. Determine the characteristics of studies not retrieved from any of the databases; and
4. Evaluate the success of MEDLINE search strategies reported in a sample of SRs in retrieving studies included in the SR and available in MEDLINE.

METHODS

Objectives 1, 2, and 3

We used relative recall methodology (19) to build a quasi-gold standard (QGS). A QGS is a set of relevant records against which the performance of a search strategy, or coverage of a database, is tested to determine how effective it is at retrieving particular record types. Although a QGS can be formed by a hand-search for relevant records, a QGS formed from relative recall is often used to test the performance of search strategies

and database yield (20–22). In this study, we used the QGS to assess the yield of each database, and combinations of databases.

The QGS comprised EEs included in reviews commissioned or carried out by the English National Institute for Health and Care Excellence (NICE). We selected SRs that were either: (i) commissioned and funded by the health technology assessment (HTA) program on behalf of NICE and published in the journal *Health Technology Assessment* or (ii) conducted as part of NICE guideline development in the Public Health work-stream and published on the NICE Web site.

Searches undertaken in this context are shaped by methodological standards, reporting guidelines, and requirements for this type of evidence (3;5;23). Therefore, these reviews might be assumed to be of good quality and likely to clearly describe their methodology. No additional quality assessment of individual search strategies was undertaken. The inclusion of reviews from the NICE Public Health work-stream reflected our intention to include nonclinical topics in the QGS, increasing generalizability.

Reviews were identified by hand-searching the journal Web site of *Health Technology Assessment*, starting at the most recent publication and working back in date. Reviews undertaken to inform published NICE Public Health guidance were identified by browsing the Guidance section of the NICE Web pages, filtered by guidance type and starting with the most recent. The identification of candidate reviews took place in February 2017.

Eligible candidate reviews had to meet prespecified criteria (Table 1). Results were screened by one reviewer; any reviews where a clear inclusion decision could not be made were

Table 1. Eligibility Criteria to Be Included in the Sample of Candidate Reviews

Eligible review
Published in *Health Technology Assessment* journal as part of the work of the NIHR HTA Programme on behalf of NICE OR undertaken to inform guidance developed by the NICE Public Health work-stream.
The review methods are explicitly described in the title, abstract or text as 'systematic' (this will include for example, "a systematic review", "we systematically reviewed the literature", "a systematic literature review", "a systematic search").
Objectives of the review clearly include systematically reviewing economic evaluations of healthcare interventions in the title, abstract, or text. For the purposes of this research this includes the explicit review of "economic evaluations", "cost-benefit analysis", "cost-effectiveness analysis", "cost-minimization analysis", "cost consequence analysis" or "cost-utility analysis" research. We will include reviews that searched for both clinical and cost effectiveness evidence together if the studies included for the economic review are reported separately.
The search reporting must include at least one search strategy (the actual strategy as run, not just a narrative description).

discussed with a second reviewer (or third reviewer if necessary) and agreement was reached.

We aimed to harvest a minimum of 350 studies from eligible reviews, with approximately 280 (80 percent) sourced from reviews published in *Health Technology Assessment*, and approximately seventy (20 percent) sourced from reviews produced as part of the NICE Public Health work-stream. The 80 percent/20 percent split reflected the approximate ratio of technology appraisals to public health reviews on the NICE Web site in 2017.

We selected reviews using the eligibility criteria and harvested the included EEs from the reviews. Reviews were selected from the *Health Technology Assessment* journal and the NICE Web pages until we reached the target number of studies. Once we achieved the target number we continued harvesting studies from any remaining eligible reviews published in that same year so we would have a complete year.

EEs included in each identified review were extracted and added to an Excel spreadsheet. Duplicates (records included in more than one review) were removed. Material submitted to NICE by the manufacturer as part of the HTA process and cited as evidence was excluded and did not form part of the QGS. References where the citation details were ambiguous and where we could not confidently identify the item being cited were also excluded. The remaining references formed our QGS set of relevant studies.

We searched nine databases for each QGS reference, to ascertain which databases included each reference. The databases comprised:

- Five healthcare databases:
    - Ovid MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, MEDLINE Daily and MEDLINE 1946 to Present
    - Embase (Ovid)
    - HTA Database (CRD Database)
    - CEA Registry (http://healtheconomics.tuftsmedicalcenter.org/cear4/home.aspx)
    - PubMed (https://www.ncbi.nlm.nih.gov/pubmed/).
- One general economics database:
    - EconLit (Ovid).
- Three multidisciplinary databases:
    - Science Citation Index (Web of Science)
    - Social Sciences Citation Index (Web of Science)
    - Scopus (Elsevier).

The databases represent the range of types of database that might be searched for EEs and that previous research and available guidance suggested were important for EEs (8;10–12;17). We also chose resources we could access and that provided suitable functionality for efficient searching in the context of a SR.

We did not include NHS EED as we wanted to identify the best sources of EEs in the context of the closure of NHS EED to new records.

The presence or absence of each reference in each database was recorded in an Excel spreadsheet.

Results were analyzed in Excel to identify:

- The yield ((number of QGS references found in each database / total number of QGS references) × 100) for each database alone and for all databases combined.
- The number of unique references retrieved from each database.
- The most efficient combination of databases in three scenarios. We defined 'efficient' as the fewest databases that could be combined to find the largest number of QGS records. The three scenarios were:
    - The most efficient combination overall
    - The most efficient combination of *healthcare* databases in the event that searchers do not have access to multidisciplinary resources
    - The most efficient combination of *free* resources in the event that searchers do not have access to subscription databases.
- The number and characteristics of references not found in any of the nine databases.

Objective 4

We evaluated the success of MEDLINE search strategies reported in a sample of SRs in retrieving studies included in the SR and available in MEDLINE.

Each eligible review was checked to see if it included EEs available in MEDLINE and reported a MEDLINE strategy in enough detail to enable reproduction. We reran the MEDLINE strategy in each review that met these criteria to see whether it retrieved the QGS records available in MEDLINE. We then calculated sensitivity, precision, and number needed to read (NNR) for each strategy. Sensitivity, precision, and NNR were defined as:

- Sensitivity % = (number of QGS records available in MEDLINE retrieved by reported MEDLINE strategy / total number of QGS records found in MEDLINE) × 100
- Precision % = (number of included QGS records available in MEDLINE retrieved by reported MEDLINE strategy / total number of MEDLINE records retrieved) × 100
- NNR = total number of MEDLINE records retrieved / number of included QGS records available in MEDLINE retrieved by reported MEDLINE strategy.

We analyzed:

- The number of MEDLINE strategies that missed at least one of the QGS records found in MEDLINE
- The total number of QGS records missed across all strategies

- Mean sensitivity, precision, and NNR across all strategies
- The reasons for nonretrieval of any studies not identified by the MEDLINE strategies.

RESULTS

Objectives 1, 2, and 3

Forming the QGS. We identified forty-six eligible reviews of EEs. Thirty reviews from the *Health Technology Assessment* produced 288 EE references and sixteen reviews from the NICE Public Health guidance Web pages produced 74 EE references. Five of the Public Health reviews included zero studies or only duplicate records. The QGS totalled 362 EE references.

Eleven of 362 QGS references were removed because they were ambiguous citations, duplicates, or material submitted to NICE by the manufacturer as part of the HTA process. A total of 351 references formed the final QGS (280 from the NIHR reviews and 71 from the NICE Public Health reviews).

Objective 1: What is the yield of the nine databases as a source of EEs? Results are given in Table 2. Embase was the highest yielding database (89 percent). Unique references were found in Embase (two references, both conference abstracts), Scopus (one reference, a health management journal article) and HTA Database (thirteen references, all nonjournal HTA agency publications).

Objective 2: What are the most efficient combinations of databases to search when conducting a SR of EEs? Results for the three different scenarios are shown in Table 3.

A combination of three databases (Embase + HTA Database + (MEDLINE OR PubMed)) identified 95 percent of the QGS (333/351 records). Four extra records (96 percent; 337/351) could be identified by additionally searching Scopus. This was the most efficient combination to find all 337 available references. The four additional references were all journal articles; one was from a clinical review on hip replacement, and three were from a single public health review and related to domestic heating and energy use (24).

Objective 3: What were the characteristics of studies not identified in any of the databases? Four percent of records (14/351) could not be found in any database. Six references were from clinical reviews and eight were from public health reviews. The fourteen references comprised: (i) one non-English language journal article; (ii) three conference abstracts; (iii) three nonjournal reports of technology assessments; and (iv) seven nonjournal reports produced by universities and research organizations.

Objective 4. How successful were MEDLINE search strategies reported in the SRs in retrieving studies included in the SR and available in MEDLINE?

Five of the reviews from the original sample of forty-six eligible SRs included zero studies or only duplicate records so were excluded from further analysis.

Twenty-nine of forty-one (71 percent) reviews included EEs that were available in MEDLINE and reported a reproducible MEDLINE strategy. Twenty-one of these reviews were on clinical topics, and the remaining eight were from NICE public health guidance.

Ten of twenty-nine (34.5 percent) strategies missed at least one of the QGS records that were in MEDLINE. Across all twenty-nine searches, the strategies failed to retrieve 17.5 percent of the records (25/143). Mean sensitivity was 89 percent, mean precision was 1.6 percent, and mean NNR was 633. Six of the twenty-nine rerun strategies were designed to identify both economic and clinical evidence. Mean sensitivity for the twenty-three rerun strategies designed to identify only economic evidence was 86 percent, mean precision was 2 percent, and mean NNR was 197. Mean sensitivity was similar in both the clinical reviews (89 percent) and public health reviews (88 percent), but the searches conducted for the public health reviews were much less precise (0.2 percent versus 2.2 percent) and had a much higher NNR (1,751 versus 206)

Only one (4 percent; one of twenty-five) of the missed records was not retrieved due to search terms used for the economics concept. This record did not include any terms to indicate economic outcomes. The majority of the records (twenty-one of twenty-five records; 84 percent) were missed because search terms used for the population or intervention concepts were insufficiently sensitive. Several strategies were designed to identify a specific drug, but the missed database records only explicitly referenced the broader drug class. Other reasons for nonretrieval included illogical combinations of search concepts (one of twenty-five records; 4 percent), and publication type or date limits (two of twenty-five records; 8 percent).

DISCUSSION AND CONCLUSIONS

Our results suggest that searching Embase, the HTA Database, and either PubMed or MEDLINE will identify the majority of EEs relevant for inclusion in SRs. In the absence of NHS EED, searchers should not rely on PubMed or MEDLINE alone, as suggested by earlier studies (8;11).

Previous research did not test the value of the HTA Database in identifying EEs. However, the HTA Database identified more unique records from our QGS (13) than any other database tested. This is likely to be because HTA Database indexes literature published by HTA agencies that is not routinely included in journal-focused bibliographic databases. HTAs from many agencies often include an EE. From May 2018 the HTA Database is not being updated while the production process transfers to INAHTA. It is currently unclear

Table 2. Yield and Number of Unique References Identified for Each Database

	Embase[a]	Scopus[c]	MEDLINE[a]	PubMed[a,d]	Science Citation Index[c]	CEA Registry[a,d]	Social Sciences Citation Index[c]	HTA Database[a,d]	EconLit[b]
No. of QGS retrieved (QGS total = 351)	314	295	285	285	271	119	63	35	12
Yield* (%)	89	84	81	81	77	34	18	10	3
No. of unique references	2	1	0	0	0	0	0	13	0

*(Number of QGS retrieved/Total number of QGS records) x 100.
[a]Healthcare database.
[b]General economic database.
[c]Multidisciplinary database.
[d]Freely available database.

whether this will result in differences in functionality and coverage: any changes may impact on database utility. The uncertainty around the future of HTA Database is concerning as it is an important resource for identifying EEs and should be searched to identify publications related to healthcare decision making not available from other bibliographic databases. At present, other than the HTA Database, this material must be identified by a time-consuming process of searching the Web pages of individual HTA agencies or by means of a general Web search engine. Producers of HTAs may consider exploring alternative methods to enhance the visibility and accessibility of their publications to researchers.

In addition to HTA Database, unique records were also found in Embase (two records) and Scopus (one record). Both of the unique records found in Embase were conference abstracts. This highlights the potential value of including

Table 3. Yield for Specific Scenarios

	No. of QGS retrieved (Total QGS = 351)	Yield* (%)
All databases	337	96
Most efficient combinations to find all 337 retrievable references		
Embase + Scopus + HTA Database + MEDLINE	337	96
Embase + Scopus + HTA Database + PubMed	337	96
Most efficient combinations of healthcare databases only		
Embase + HTA Database + MEDLINE	333	95
Embase + HTA Database + PubMed	333	95
Most efficient combination of only freely available (non-subscription) databases		
PubMed + HTA Database + CEA Registry[a]	299	85

*(Number of QGS retrieved/Total number of QGS records) x 100.
[a]CEA Registry is free for Basic Search. Advanced Search is only available as part of premium access.

databases that index conference abstracts if this type of material is eligible for inclusion in the review.

Embase was the highest yielding database. This is likely to partly reflect Embase's policy of indexing conference abstracts. For example, twenty-eight of the thirty-five QGS records found in Embase but not in MEDLINE or PubMed were conference abstracts.

The high yield of Embase can also be explained by Elsevier's project, launched in 2010, to include all MEDLINE records in Embase (25). Embase, therefore, contains two databases. Despite this ambition to include all of MEDLINE, six of our QGS records could be found in both MEDLINE and PubMed but not in Embase. Searching both resources was necessary to achieve the highest yield with the fewest possible databases. Our results, therefore, support the current recommendation to search both MEDLINE and Embase in the context of SRs (26). Searching both databases also allows searchers to exploit the differences in indexing, record structure, and search functionality between resources to maximize retrieval.

There was no difference in the performance of MEDLINE and PubMed: they retrieved the same records. This was because we searched all available segments of MEDLINE (including In-Process & Other Non-Indexed Citations). Searching Ovid MEDLINE without these segments would have resulted in a lower yield than PubMed. All available segments in Ovid MEDLINE should be searched to maximize sensitivity: the newly released segment MEDLINE All provides a simple way to achieve this.

We note that the QGS did not contain records for relatively recent publications. Searching PubMed in addition to MEDLINE has been suggested as beneficial for identifying very recent papers not yet fully indexed for MEDLINE (27). However, this research predates the expansion of MEDLINE with the addition of new segments such as MEDLINE All. We do not currently have sufficient evidence to say whether this impacts on the conclusions of the previous research. Any

additional value from searching PubMed in terms of yield, must be balanced against the comparatively limited search functionality in this interface. The inability to use advanced search syntax such as proximity searching makes it difficult to construct a strategy with the desired balance of sensitivity and precision.

Searching databases other than the core group had limited incremental yield, retrieving only four additional QGS references (1 percent). This small incremental yield does not allow for strong conclusions on the value of searching specific additional resources. However, the additional four references could all be retrieved by searching Scopus and three of the four references could be found in Science Citation Index and Social Sciences Citation Index. We suggest there is some evidence that researchers should consider searching a multidisciplinary database, particularly for nonclinical research topics. Pitt et al. have also suggested Scopus is a potentially useful resource for EEs (9).

Searching only freely available (nonsubscription) databases resulted in the identification of fewer QGS records (85 percent compared with 96 percent). Researchers who only have access to freely available databases should place an increased emphasis on supplementary search methods such as reference-list checking and citation searching to maximize retrieval of relevant studies.

Records for fourteen QGS references (4 percent) were not included in any database tested and largely comprised grey literature. Coverage of grey literature in supplementary search methods designed to retrieve this type of evidence (e.g., searches of HTA agency Web sites, conference proceedings, online sources of nonjournal reports, reference list checking, and expert contact) may be a more efficient and effective use of resources than extensive database searching. This supports the conclusion by Royle and Waugh (10) that the majority of published EEs can be identified in a small number of core databases, and beyond this, supplementary search approaches may be most productive in finding additional studies.

Despite relatively high mean sensitivity, the MEDLINE search strategies developed by the authors in the included reviews had weaknesses, resulting in nonretrieval of relevant studies. As researchers can no longer rely on searches of NHS EED or HEED to retrieve EEs missed by suboptimal searches elsewhere, it is important that search strategies in large bibliographic databases are of high quality to maximize the likelihood of identifying all relevant studies. The efficient combinations of databases we have identified will only retrieve relevant EEs if researchers conduct adequately sensitive searches.

Only one of the twenty-five (4 percent) of the records missed by MEDLINE strategies was not retrieved because of search terms used for the economics concept. This perhaps reflects the availability of published filters designed to identify EEs, such as that developed by the Centre for Reviews and Dissemination (28). Our findings suggest that improving the sensitivity of searches for EEs in these large bibliographic databases is likely to be more complex than simply encouraging the use of appropriate search filters for economic study designs. The reasons studies were missed (e.g., insufficiently sensitive search terms, illogical combinations of concepts, and the use of limits) suggest searchers need to be aware of a range of issues. We recommend that researchers designing strategies to identify EEs in general bibliographic databases such as MEDLINE use a published filter designed to identify EEs, ensure that terms for population and intervention concepts are sufficiently sensitive, consider whether date or publication type limits are appropriate, and check their strategies carefully to identify syntax errors.

The use of quality assessment tools for search strategies, such as the PRESS checklist (29), may help to achieve this. Research has also suggested that the involvement of a suitably experienced librarian or information specialist improves the quality of searches conducted as part of SRs (30–32).

The MEDLINE strategies that we tested demonstrated variable precision and NNR. NHS EED and HEED allowed the construction of relatively precise search strategies as their content was prefocused by study type. The closure of NHS EED and HEED and the subsequent reliance on larger bibliographic databases suggests that the ability of searchers to construct strategies that can achieve high sensitivity with reasonable precision will become more important. Precision can best be achieved by searching using sophisticated interfaces that allow the use of phrase searching, proximity operators and other techniques to introduce focus to strategies. Precision may be a particular challenge in reviews related to public health and other nonclinical topics.

In conclusion, we suggest searching for published EEs can be limited to key databases, as long as these databases are searched using methodologically appropriate strategies. Searchers should concentrate on developing suitable search strategies for key databases to ensure high sensitivity and adequate precision, in addition to using supplementary search approaches to retrieve evidence that is unpublished or unlikely to be identified by bibliographic databases.

Limitations of this Study

Several limitations should be taken into account when considering this study's results and conclusions.

The candidate reviews were screened against prespecified eligibility criteria by one reviewer. If a clear inclusion decision could not be made, agreement was reached with a second or third reviewer. Using a single reviewer for screening increased the risk of selection errors and bias.

Although the nine databases tested were chosen to represent the range of databases that might be searched for EEs, many other databases are available to researchers. Our

study can only provide information on the yield of the included databases.

Our QGS comprised 351 EEs. Although this is a reasonably sized QGS, a larger reference set would have increased the generalizability of our research. Sourcing the QGS from a wider range of reviews could also have improved generalizability. All reviews were produced in the context of United Kingdom decision making and focused on clinical medicine or public health. Our findings may be less generalizable to reviews specifically relevant to other topics (e.g., mental health, health management) or healthcare contexts (e.g., low- and middle-income countries [LMICs]).

The robustness of these findings depends on the extent to which the QGS is representative of all relevant EEs. Representative QGS sets result from high quality search strategies that can be expected to retrieve a high proportion of all available relevant studies. Ideally, the quality of searches conducted by the SRs from which the QGS is harvested should be assessed. However, we took the pragmatic decision not to quality assess each search strategy in the reviews from which the QGS was sourced. The assumption that searches conducted in the context of NICE decision making would be of sufficient quality to provide a representative QGS is a potential limit of our methodology. Weaknesses in the search methods of source reviews could have failed to retrieve eligible EEs and lessen the degree to which our QGS was representative of all relevant studies.

We define "efficient" as the fewest databases that could be combined to find the largest number of QGS records. The number of databases is one measure of search efficiency, but not the only one. We do not take into account, for example, time taken to search a database, time taken to export records, or the number of irrelevant records retrieved. Our study also does not consider the impact of database interfaces on efficiency. Many bibliographic databases are available on more than one platform, each providing different functionality that can impact on retrieval. When viewing the database combinations reported as most efficient, the limitations of our definition should be considered.

The sensitivity of each review's MEDLINE search was used as a proxy for search quality. This provides only limited information on the quality of search methodology. Although a record for a relevant study may be missed by the MEDLINE strategy, this does not mean the strategy, when translated, will also fail to identify that record in other databases searched. Additionally, the MEDLINE strategy may not reflect the search approaches used in the other databases searched by the review. There are other methods to assess quality of searches, such as the PRESS checklist (29). As we were only concerned with whether the strategy could retrieve QGS records, elements of search methodology assessed by PRESS (e.g., errors in search syntax, missing search terms, and inappropriate limits) were not of interest unless they impacted

on retrieval. Reasons for nonretrieval did closely map to several aspects of search development covered by PRESS elements.

Implications for Practice

Our findings can inform researchers' decisions on database choice when searching for EEs following the closure of NHS EED and HEED. Although our research was carried out in the context of searching for SRs, our findings on database yield and search quality are also likely to be relevant to those searching for EEs for other purposes.

Implications for Research

We can only provide information on the value of the nine databases tested. There is scope for analysis of further databases, particularly in the context of economic reviews with a specific focus (for example nursing, mental health, or health care in LMICs). (9) The bibliometric analysis by Pitt et al. of EEs from a global perspective suggested that Global Health, a database that has particular relevance for LMIC research (33), merits further exploration.

Similarly, we can only provide information on the value of the included databases in relation to our QGS set. QGS records were harvested from reviews with a particular focus (mainly clinical, with some public health) and research context (United Kingdom health care). Further research to investigate how these findings relate to QGS sets harvested from SRs with a different focus or research context would be valuable.

Records for fourteen QGS references (4 percent) were not included in any database and we suggest supplementary search methods are used to identify these types of studies. Evidence-based research on the relative value of different supplementary search methods is needed, particularly as many methods can be resource intensive.

CONFLICTS OF INTEREST

M Arber, J Glanville, M Edwards, E Baragula and H Wood are employed by YHEC. J Isojarvi and A Shaw are former employees of YHEC. YHEC is a consultancy company conducting systematic reviews of the effectiveness and cost-effectiveness of interventions and provides search services and training in search conduct.

REFERENCES

1. York York Health Economics Consortium. Economic evaluation [online]. 2016. http://www.yhec.co.uk/glossary/economic-evaluation/ (accessed September 30, 2018).
2. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. 2017. https://www.cadth.ca/dv/guidelines-economic-evaluation-health-technologies-canada-4th-edition (accessed September 30, 2018).

3. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance. Process and methods [PMG4]. 3rd ed. 2012. https://www.nice.org.uk/process/pmg4/chapter/introduction (accessed September 30, 2018).

4. National Institute for Health and Care Excellence. Developing NICE guidelines: The manual. Process and methods [PMG20]. 2014. https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview (accessed September 30, 2018).

5. National Institute for Health and Care Excellence. Guide to the processes of technology appraisal. 2018. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf (accessed September 30, 2018).

6. Scottish Medicines Consortium. How we make our decisions. 2018. https://www.scottishmedicines.org.uk/how-we-decide/ (accessed September 30, 2018).

7. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*. 2017;7:e012545.

8. Alton V, Eckerlund I, Norlund A. Health economic evaluations: How to find them. *Int J Technol Assess Health Care*. 2006;22:512-517.

9. Pitt C, Goodman C, Hanson K. Economic evaluation in global perspective: A bibliometric analysis of the recent literature. *Health Econ*. 2016;25(Suppl 1):9-28.

10. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technol Assess*. 2003;7(34):iii, ix-x, 1-51.

11. Sassi F, Archard L, McDaid D. Searching literature databases for health care economic evaluations: How systematic can we afford to be? *Med Care*. 2002;40:387-394.

12. Centre for Reviews and Dissemination. Undertaking the review. In: Centre for Reviews and Dissemination, editor. Systematic reviews: CRD's guidance for undertaking reviews in health care. York. University of York. 2009. https://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/1_3_UNDERTAKING_THE_REVIEW.htm (accessed September 30, 2018).

13. Shemilt I, Mugford M, Byford S, et al. Incorporating economics evidence. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration. 2011. 449-80 www.handbook.cochrane.org (accessed September 30, 2018).

14. Wood H, Arber M, Glanville JM. Systematic reviews of economic evaluations: How extensive are their searches? *Int J Technol Assess Health Care*. 2017;33:25-31.

15. Briscoe S, Cooper C, Glanville J, Lefebvre C. The loss of the NHS EED and DARE databases and the effect on evidence synthesis and evaluation. *Res Synth Methods*. 2017;8:256-257.

16. Thielen FW, Van Mastrigt G, Burgers LT, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: Database selection and search strategy development (part 2/3). *Expert Rev Pharmacoecon Outcomes Res*. 2016;16:705-721.

17. Kaunelis D, Glanville J. Costs and economic evaluation. Summarized research in information retrieval for HTA. 2018. http://vortal.htai.org/?q=sure-info (accessed September 30, 2018).

18. Kaunelis D, Farrah K. Appraisal of: Pitt C, Goodman C, Hanson K. Economic evaluation in global perspective: A bibliometric analysis of the recent literature. *Health Econ*. 2016;25(Suppl 1):9-28.

19. Sampson M, Zhang L, Morrison A, et al. An alternative to the hand searching gold standard: Validating methodological search filters using relative recall. *BMC Med Res Methodol*. 2006;6:33.

20. Aagaard T, Lund H, Juhl C. Optimizing literature search in systematic reviews - Are MEDLINE, EMBASE and CENTRAL enough for identifying effect studies within the area of musculoskeletal disorders? *BMC Med Res Methodol*. 2016;16:161.

21. Ayiku L, Levay P, Hudson T, et al. The medline UK filter: Development and validation of a geographic search filter to retrieve research about the UK from OVID medline. *Health Info Libr J*. 2017. doi:10.1111/hir.12187.

22. Selva A, Sola I, Zhang Y, et al. Development and use of a content search strategy for retrieving studies on patients' views and preferences. *Health Qual Life Outcomes*. 2017;15:126.

23. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal [PMG9]. London: National Institute for Health and Care Excellence; 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword (accessed September 30, 2018).

24. London School of Hygiene & Tropical Medicine, Public Health England, University College London. Evidence review & economic analysis of excess winter deaths for the National Institute for Health and Care Excellence (NICE). Review 2: Interventions and economic studies. Undated. https://www.nice.org.uk/guidance/NG6/documents/excess-winter-deaths-and-illnesses-review-22 (accessed September 30, 2018).

25. Lam MT, De Longhi C, Turnbull J, Lam HR, Besa R. Has Embase replaced MEDLINE since coverage expansion? *J Med Libr Assoc*. 2018;106:227-234.

26. Lefebvre C, Manheimer E, Glanville JM. Chapter 6: Searching for studies. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions* [Version 5.1.0] [updated March 2011]. The Cochrane Collaboration. 2011. http://handbook-5-1.cochrane.org/.

27. Duffy S, de Kock S, Misso K, et al. Supplementary searches of PubMed to improve currency of MEDLINE and MEDLINE In-Process searches via Ovid. *J Med Libr Assoc*. 2016;104:309-312.

28. University of York, Centre for Reviews and Dissemination. Search strategies: NHS EED. http://www.crd.york.ac.uk/crdweb/searchstrategies.asp#nhseedmedline (accessed September 30, 2018).

29. McGowan J, Sampson M, Salzwedel DM, et al. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75:40-6.

30. Meert D, Torabi N, Costella J. Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. *J Med Libr Assoc*. 2016;104:267-277.

31. Koffel JB. Use of recommended search strategies in systematic reviews and the impact of librarian involvement: A cross-sectional survey of recent authors. *PLoS One*. 2015;10:e0125931.

32. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015;68:617-626.

33. Effective Practice and Organisation of Care (EPOC) Cochrane Group. A collection of databases, web sites and journals relevant to Low-and Middle-Income Countries (LMICs). 2013. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/LMIC%20Databases%20August%202013.pdf (accessed September 30, 2018).