

Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: a prospective comparative study of literature search methods

Matt Egan, Alice MacLean, Helen Sweeting, Kate Hunt

To cite: Egan M, MacLean A, Sweeting K, *et al.* Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: a prospective comparative study of literature search methods. *BMJ Open* 2012;**2**: e001043. doi:10.1136/bmjopen-2012-001043

► Prepublication history and additional materials for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-001043>).

Received 2 March 2012
Accepted 22 May 2012

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

Social and Public Health Sciences Unit (SPHSU), MRC/CSO, Glasgow, UK

Correspondence to

Dr Matt Egan;
m.egan@sphsu.mrc.ac.uk

ABSTRACT

Objective: To compare the effectiveness of systematic review literature searches that use either generic or specific terms for health outcomes.

Design: Prospective comparative study of two electronic literature search strategies. The 'generic' search included general terms for health such as 'adolescent health', 'health status', 'morbidity', etc. The 'specific' search focused on terms for a range of specific illnesses, such as 'headache', 'epilepsy', 'diabetes mellitus', etc.

Data sources: The authors searched Medline, Embase, the Cumulative Index to Nursing and Allied Health Literature, PsycINFO and the Education Resources Information Center for studies published in English between 1992 and April 2010.

Main outcome measures: Number and proportion of studies included in the systematic review that were identified from each search.

Results: The two searches tended to identify different studies. Of 41 studies included in the final review, only three (7%) were identified by both search strategies, 21 (51%) were identified by the generic search only and 17 (41%) were identified by the specific search only. 5 of the 41 studies were also identified through manual searching methods. Studies identified by the two ELS differed in terms of reported health outcomes, while each ELS uniquely identified some of the review's higher quality studies.

Conclusions: Electronic literature searches (ELS) are a vital stage in conducting systematic reviews and therefore have an important role in attempts to inform and improve policy and practice with the best available evidence. While the use of both generic and specific health terms is conventional for many reviewers and information scientists, there are also reviews that rely solely on either generic or specific terms. Based on the findings, reliance on only the generic or specific approach could increase the risk of systematic reviews missing important evidence and, consequently, misinforming decision makers. However, future research should test the generalisability of these findings.

ARTICLE SUMMARY

Article focus

- Providing evidence-based guidance to improve electronic literature searches (ELS): an often overlooked but vital stage in our efforts to inform policy and practice with the best available evidence.
- During systematic review literature search we conducted two ELS and compared the results: one ELS included search terms for a range of specific health conditions, while the other included only generic terms for health and illness.

Key messages

- Future systematic reviews that involve multiple health outcomes should include both generic and specific health terms in their literature search.
- Based on our findings, previous reviews that have only used one of these approaches may have failed to identify relevant evidence and this in turn could have affected the reviewers' conclusions.
- Systematic reviews that miss important evidence risk causing harm by misinforming practitioners and other decision makers.

Strengths and limitations of this study

- The relatively novel application of a prospective comparative study design to the issue of electronic literature searching is a key strength.
- Although the searches identified over 10 000 initial records, they could have been made more sensitive through greater use of techniques such as truncation, synonyms and by searching additional databases.
- The study is based on searches conducted for a specific review, so the generalisability of our findings should be tested in the context of other reviews and different types of literature search, including more sensitive searches.

INTRODUCTION

Electronic literature searches (ELS) are an essential stage in most systematic reviews.^{1 2} As such, they have a crucial role in the scientific community's attempts to inform and improve policy and practice with the best available evidence.^{3 4} Designing ELS can be challenging and it is widely recognised that specialist skills and knowledge, such as those provided by an information scientist, are important for best practice in this field.¹⁻³ The trade-off between screening out irrelevant evidence while identifying relevant evidence (sometimes discussed in terms of a search's 'precision' and 'sensitivity') is a well-known challenge for information scientists and researchers who work on systematic reviews. In this paper, we present a worked example of how an empirical study comparing different ELS can be conducted to explore the effects that different search strategies may have on the identification of studies for a systematic review and how this in turn may affect the review's conclusions.

Systematic reviews vary in terms of subject matter and approach,³ and this can have implications for how ELS are designed. Some systematic reviews are based on comprehensive searches, which aim to have high sensitivity and retrieve references to all relevant papers, whereas others are based on more restricted searches, which may limit the number of relevant papers identified.⁵ Search strategies that are insufficiently sensitive may risk encouraging potentially harmful decisions based on the findings of reviews that have failed to identify important evidence. Search strategies that aim to comprehensively identify all the relevant evidence can present challenges in situations where reviewers have limited time or other resources (eg, as a result of research funding requirements or because findings are considered to be needed urgently) or when extending a search fails to identify relevant evidence and might therefore represent an ineffective allocation of scarce resources.³⁻⁵

Previous research exploring how to improve the effectiveness and efficiency of search strategies has tended to focus on issues such as how to optimise search outputs from 'frontline' electronic databases (ie, databases that are frequently searched for systematic reviews of medical interventions such as Medline and Embase) and how to identify randomised control trials (RCTs).⁶⁻¹² This research focus may in part reflect the influence of the Cochrane Collaboration, which has helped to stimulate considerable interest in systematic reviews of clinical trials.¹

However, not all systematic reviews (nor indeed all Cochrane Reviews¹³) focus on RCTs of clinical interventions. Interest in broader, non-clinical systematic reviews has steadily increased within the social and public health sciences and other disciplines.^{3 5} As some of these non-clinical reviews tackle relatively under-researched topics, they often combine a scoping and hypothesis testing function by asking relatively broad

research questions that, for example, cover a range of outcomes (eg, *what are the health impacts of intervention x?*; *what health outcomes are associated with risk-factor y?*).¹⁴⁻²⁷ Evidence-informed guidance on how to conduct searches for this broader range of systematic reviews is therefore an emerging priority.

There are few examples of research that can help guide information scientists and reviewers to develop efficient but effective search strategies for these broader/non-clinical systematic reviews. The research that is available illustrates how searches for such reviews can become lengthy and complex.²⁸ For example, Greenhalgh *et al* recommended the development of iterative search strategies to search for complex evidence (eg, multiple study designs). Ogilvie *et al*⁴ suggested that cross-disciplinary reviews may necessitate searching databases across a range of disciplines rather than focusing on frontline health databases.

From our own experiences of conducting systematic reviews of non-clinical public health research, the authors of this paper can identify additional challenges that have led to large and complex ELS. For example, search terms that involve commonly used words are likely to identify large numbers of irrelevant papers, and non-clinical public health reviews often rely on commonly used terms to describe everyday settings, activities and outcomes (eg, 'walking', 'obesity', 'stress', 'workplace health', 'health promotion' and 'general health'). In comparison, an ELS for a clinical review will often involve very specific medical terminology that can help to focus the search on papers relevant to a particular field.³

Furthermore, the Cochrane Handbook¹ (section 6.4.2) states that a search strategy to identify studies for a Cochrane Review 'typically has three sets of terms: (1) terms to search for the health condition of interest, that is, the population; (2) terms to search for the intervention(s) evaluated and (3) terms to search for the types of study design to be included (typically a 'filter' for randomised trials)'. Each of these sets of terms can help to filter out unwanted studies from the search, but it is not always appropriate or possible to structure an ELS in this way. Systematic reviews do not always include populations defined by a health condition (they may, eg, focus on studies of the general population). As stated earlier, not all systematic reviews are based on evaluations of interventions. Furthermore, not all systematic reviews focus on RCTs, and some include a range of study designs. Systematic reviewers recognise that it is sometimes appropriate to deviate from this typical search structure: for example, the Cochrane Handbook states that in some circumstances it may be necessary to search 'only for the population or the intervention' (Cochrane Handbook¹ section 6.4.2).

The chances of an ELS identifying irrelevant studies could be increased if the search includes both specialist and non-specialist databases, or uses search terms based on unspecialised vocabulary, or cannot include terms for

population types or interventions or study designs to help screen out irrelevant literature. Searches characterised by a large number of search results and low precision may be resource intensive and this could become a problem if the resources required for a search outstrip what is available for a particular review. In such circumstances, reviewers may look for alternative means of increasing precision. However, for the broader public health reviews of the kind we have described here, there is relatively little evidence-based guidance on how greater precision can be achieved without compromising sensitivity (compared with the guidance on clinical/RCT systematic reviews).

Including search terms that relate to health outcomes is one commonly used technique for increasing precision in broader reviews.^{14–24} However, if a review question is broad enough to include multiple health outcomes, it is not obvious how an ELS that includes health outcomes can best accommodate this breadth of scope. Some reviews have used generic health terms (eg, ‘health’, ‘illness’, ‘morbidity’) to search for evidence that includes a range of health outcomes.^{14–17} In other cases, reviewers have used more specific search terms to identify a number of diseases or symptoms considered to be of particular relevance to the review question.^{18–21} Both approaches may be hypothesised to have risks. Generic search terms may either be too inclusive (virtually every study on Medline is about ‘health’) or may miss studies that only use more specialist vocabulary to describe a particular illness. Specific search terms are problematic if the reviewers want to avoid pre-specifying which health outcomes are relevant to the review (eg, scoping reviews). Some reviews combine both generic and specific approaches,^{22–24} but the extent to which this either adds value to the search or merely adds to the workload is not known.

We know of no study that has compared the relative merits of ELS strategies that focus on either generic terms for health or specific terms for particular health issues or illnesses. Nor do we know of any evidence to help reviewers determine whether these two approaches are likely to identify a similar or a different set of publications (both of the above observations are based on a non-systematic exploration of the literature rather than a systematic review). When the authors of this paper recently conducted a systematic review that included multiple health outcomes, we felt that guidance on this issue would have been helpful. As there was an absence of evidence upon which to base such guidance, we ran two separate literature searches for our review: one that included generic health terms and one that used more specific health terms. Our aim was to see which approach was most effective in identifying studies that were included in the final review.

Hence, we examined whether the included studies tended to be identified from the generic search only, the specific search only or both searches. We also explored efficiency by comparing the size of the searches (ie, the

number of references initially identified from the ELS—sometimes referred to as the number of ‘hits’) for each approach. Finally, we explored the extent to which the ‘generic search’ and the ‘specific search’ identified studies with different or similar types of health outcome.

Our review was conducted within a limited time frame (originally planned as 9 months and then extended to 18 months), and we believe that the implications of this study are of particular relevance to reviews of broader public health topics and reviews with time or other resource limitations.

METHODS

This paper focuses on one specific, but crucial, stage of a systematic review: the literature search. We developed two contrasting strategies for searching electronic databases and compared their effectiveness in identifying studies for a specific systematic review. The systematic review itself is summarised in [box 1](#), and described more fully in the publically available Protocol document (available as a supplemental document), and the full report of the review that will be published separately to this methodological paper.

Box 1 Summary of the systematic review used as the basis of this methodological study

Title: How robust is the evidence of an emerging or increasing female excess in physical morbidity rates between childhood and adolescence? Results of a systematic literature review.

Hypothesis: That the incidence of physical morbidity among children tends to be higher among men in pre-adolescent childhood, but this male excess is replaced by an emergence of higher rates in females during the transition to adolescence.

Inclusion/exclusion criteria: These criteria are summarised using the PICOS statement below. For full details of the inclusion and exclusion criteria, see the protocol: supplemental document.

Included studies must have the following characteristics

Population: men and women between the ages of 4 and 17;

Intervention: none;

Comparator: sex and age (at least two age groups);

Outcome: gender patterning, by age, in measures of physical morbidity;

Study design: longitudinal, cross-sectional and repeat cross-sectional studies (including analysis of study-specific data or routinely collected data).

Methods: The systematic review included methodological components suggested by the PRISMA guidelines (eg, protocol, literature search, study selection, flow chart, data extraction, critical appraisal and synthesis) and was designed to meet the standards of that guidance. More details are provided in the protocol.

Data sources and search strategy

We searched five electronic databases (Medline, Embase, the Cumulative Index to Nursing and Allied Health Literature, PsycINFO and the Education Resources Information Center) for studies published in English between 1992 and the date of search (April 2010). As it was our intention to update a previous review conducted around 20 years previously,²⁹ we searched for studies published from 1992 to the present. Supplemental document 1 describes the review methods and search strategy in more detail. Following test searches using pre-identified papers, an information scientist advised on database selection and search terms. As the review's time frame was limited, the information scientist advised on a search strategy that limited the number of records retrieved by the searches so that they could be processed within the time frame. Prior to the electronic search, we manually searched private collections (one of the reviewers has worked in the field of gender and adolescent health for several years and two for approximately 2 decades), conducted a relatively unstructured internet search and also identified papers that had cited the earlier review.²⁹ At the end of our study selection process, we manually checked the bibliographies of included studies.

We searched each database twice: once using 'generic' health subject headings and keywords and once using 'specific' subject headings and keywords relating to the health conditions we had selected for review (see table 1). In this paper, we refer to these searches as the 'generic search' and the 'specific search'. The precise search strategy differed between databases if different search facilities and search engines made it necessary to adapt our approach.

Study selection

One reviewer (AM) screened all the publications identified by both literature searches to exclude obviously irrelevant titles. The remaining (ie, not excluded) publications were retrieved and, on reading, AM screened out those that were clearly not eligible for inclusion in the review (see figure 1, 'First Sift'). Studies of uncertain eligibility were checked by two other reviewers (KH and HS) so that a decision to exclude or retrieve the full paper could be reached (see figure 1, 'Second Sift'). Some retrieved papers were excluded at the initial reading ('Third Sift'), while others were excluded at the data extraction and appraisal stage (based on agreement from all the reviewers). At this final stage, we also excluded studies that only explored asthma-related outcomes after finding a review that already applied our research question to this health outcome.

Outcomes

Our main outcomes measures for this analysis were the number and proportion of studies included in the systematic review that were identified from each ELS. We also collected data on (1) the number of studies identified by each ELS at all stages of the reviews' search and selection process, (2) the types of health outcomes

identified by each ELS and (3) the number of studies identified by manual searches.

Comparing the two searches

We produced a series of Venn diagrams for each stage of the review process, showing the number of studies identified only by the specific literature search, the number identified only by the general literature search and the number identified by both searches (see figure 1). The purpose was to see if the two searches identified similar or different sets of documents. Studies that were included in the final review were then tabulated in more detail to help us assess whether there was any systematic variation in the types of health outcome identified by the different searches. Comparisons involved the calculation of frequencies and percentages.

RESULTS

Figure 1 shows for each stage of the review the number of studies identified exclusively by either the specific or the generic search and (in each intersect) the number of studies identified by both searches.

The diagram makes two points apparent. First, there was relatively little duplication between the two searches. For example, of the 11 509 total hits identified from both literature searches, only 413 (3.6%) were duplicates between the two searches. Throughout each stage of the study selection process, duplication between the two searches remained low, so that only three (7.3%) of the 41 studies selected for final inclusion in the review were identified by both search strategies (further details of the 41 included studies are available in a supplemental document).

Second, we note that the specific search led to less than half the number of initial hits compared with the generic search (3299 vs 8210, respectively), but both searches identified a similar number of studies included in the final review (17 vs 21 and three duplicates).

Four final inclusion studies were identified from our initial manual search but the generic ELS also identified each of these four studies. Further bibliographic checking revealed that one of the studies identified from both the manual search and the generic ELS could also have been found by checking the bibliographies of included studies identified from the specific search. One study identified from the specific ELS could also have been found by checking the bibliographies of included studies identified from the generic search. This means that the generic ELS in combination with the manual search and bibliography check would have identified 25 of the 41 included studies. The specific ELS in combination with the manual search and bibliography check would have identified 24 of the 41 included studies.

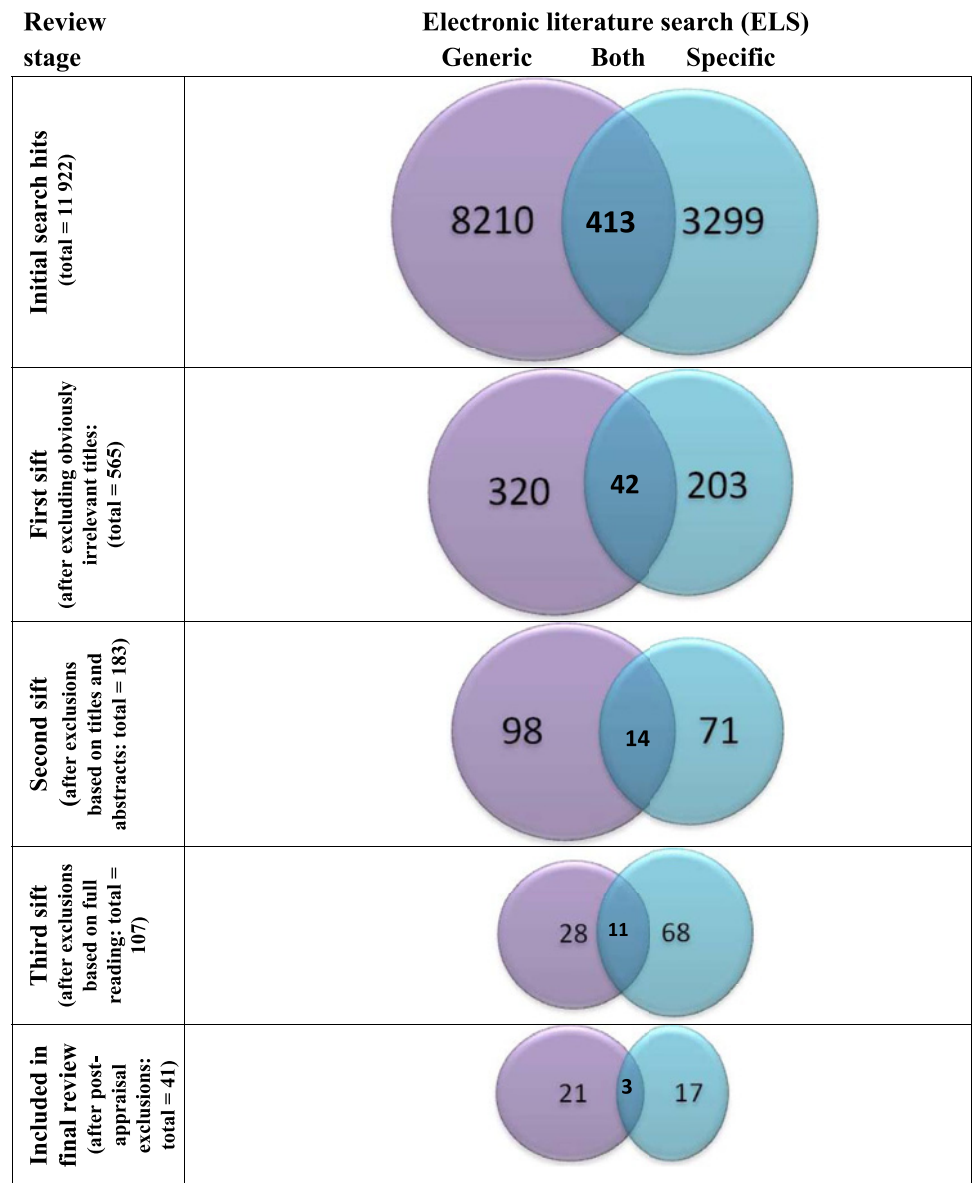
We then examined the 41 studies included in the final review, categorising them by the health outcomes each one investigated (see table 2). The findings suggest some systematic differences in the health outcomes of studies identified using each of the two search strategies.

Table 1 Search history

| Database | Generic search | Specific search |
|--|--|---|
| Medline (Ovid interface) | Date: 16 May 2010 (child, preschool or child or adolescent).sh. and (sex factors or sex distribution).sh. and (health status or attitude to health or health surveys or mental health†).sh. limit to (english language and humans and yr="1992 -Current") (.sh. = MeSH subject headings) Results – 3587 | Date: 06 May 2010 (child, preschool or child or adolescent).sh. and (sex factors or sex distribution).sh. and (asthma or epilepsy or headache disorders, primary or diabetes mellitus).sh. limit to (english language and humans and yr="1992 -Current") (.sh. = MeSH subject headings) Results – 1426 |
| Embase (Ovid interface) | Date: 19 April 2010 (child or school child or adolescent or preschool child).sh. and (sex difference or gender).sh. and (adolescent health or health survey or health status).sh. limit to (human and English language and yr="1992 -Current") (.sh.= subject headings) Results – 2652 | Date: 06 May 2010 (child or school child or adolescent or preschool child).sh. and (sex difference or gender).sh. and (asthma or primary headache or migraine or diabetes mellitus or epilepsy).sh. limit to (human and English language and yr="1992 -Current") (.sh.= subject headings) Results – 1526 |
| CINAHL (Cumulative Index to Nursing and Allied Health) (EBSCO Host interface) | Date: 19 April 2010 (adolescence or child or child, preschool).sh. and (sex factors).sh. and (health status or health status indicators or attitude to health or symptoms or morbidity or child health or adolescent health).sh. limit to (english language and yr="1992 -Current") (.sh.= word in subject heading) Results – 1467 | Date: 07 May 2010 (adolescence or child or child, preschool).sh. and (sex factors).sh. and (asthma or diabetes mellitus or headache or epilepsy).sh. limit to (english language and yr="1992 -Current") (.sh.= word in subject heading) Results – 498 |
| PsycINFO (EBSCO Host interface) | Date 19 April 2010 (child* or adolesc* or young pers* or teenage* or pupil* or school child* or minor*).kw. and (human sex differences or sex).sh. and (health or health attitudes or health complaints or general health questionnaire or well being or self report or morbidity or symptoms).sh. limit to (english language and yr="1992 -Current") (.kw. – keywords, .sh.= subjects) Results – 1136 | Date: 07 May 2010 (child* or adolesc* or young pers* or teenage* or pupil* or school child* or minor*).kw. and (human sex differences or sex).sh. and (asthma or diabetes mellitus or headache or epilepsy).sh. limit to (english language and yr="1992 -Current") (.kw. – keywords, .sh.= exact subjects) Results: 38 |
| ERIC (Education from US Department of Education, and Institute of Education Sciences) (Ovid interface) | (19 April 2010) ((children or young children or adolescents or early adolescents or late adolescents).sh. or (pupil* or school child* or minor*).ab.) and ((sex or gender differences).sh. or (sex or gender).ab.) and ((health or child health or adolescent health or well being or wellness).sh. or (morbidity or symptom*).ab.) limit to (english language and yr="1992 -Current") (.sh.= ERIC subject headings, .ab. = abstract) Results: 593 | Date: 07 May 2010 ((children or young children or adolescents or early adolescents or late adolescents).sh. or (pupil* or school child* or minor*).ab.) and ((sex or gender differences).sh. or (sex or gender).ab.) and ((asthma or headache or migraine).ab. or (diabetes or epilepsy).sh.) limit to (english language and yr="1992 -Current") (.sh.= ERIC subject headings, .ab. = abstract) Results: 22 |

† Following this initial search, the decision was made to focus the review on physical rather than mental health.

Figure 1 Review study selection flow chart: studies identified by the 'generic' search only (purple circle), 'specific' search only (light blue circle) and by both searches (dark blue intersect). From the above figures the following can be calculated. Generic ELS: sensitivity =58.5%; precision =0.3%; number needed to read =112 (1.3%). Specific ELS: sensitivity =48.8%; precision =0.5%; number needed to read =85 (2.3%). An initial manual search identified 10 articles for full reading, of which four were included in the final review. These studies were also identified from the generic ELS and are included as such in the figure.



The specific search tended to be the more successful at identifying studies that focused on a single type of health outcome (ie, those that related to the search terms). The opposite was found for the generic search strategy, which tended to be more successful at identifying studies with multiple health outcomes.

Most notably, we found that the specific ELS alone (ie, not the generic ELS or manual search) identified all three included studies of epilepsy and all but one of the seven studies on diabetes. Therefore, failure to run the specific search would have meant that our review would have missed most of the evidence relating to these two outcomes. Within the context of our review's findings, this omission would have been important because, while the evidence for the other health outcomes presented in table 2 tended to support our review's main hypothesis, findings for diabetes and epilepsy uniquely suggested a counter hypothesis. Failing to identify evidence to support the counter hypothesis would have directly affected our review's conclusions.

The tables in supplemental document 2 describe the studies identified by the different ELS by summarising information on health outcome, journal, study design, appraisal score and country. Three longitudinal studies and six studies classed as higher scoring following the study appraisal were among those identified by the generic ELS (although three of these were also identified using the manual search). Five higher scoring studies (but no longitudinal studies) were among those only identified by the specific ELS. Both searches identified evidence from a similar (but not identical) range of European countries but only the generic search identified any North American studies. All the studies identified were published in medical/health journals.

DISCUSSION

We have compared two strategies for conducting an electronic literature search for a systematic review. One strategy used generic health terms, while the other used more specific health terms. The purpose was to explore

Table 2 Studies included in the systematic review (n=41) by summary health outcomes and by the search strategy used to identify each study

| Outcomes | Generic search | Specific search | Both searches |
|------------------------------------|----------------|-----------------|---------------|
| Abdominal pain | | 1 | |
| Back pain | 2 | | |
| Diabetes | 1 | 6 | |
| Epilepsy | | 3 | |
| Headache | 2 | 7 | 1 |
| General physical health/well-being | 5 | | |
| Multiple physical health outcomes* | 11 | | 2 |
| Total (for each search) | 21 | 17 | 3 |

*A range of health outcomes were included in these studies: usually involving measures of general health and bodily pain. See tables in supplemental document 2.

whether literature searches with relatively broad inclusion criteria (in terms of health outcomes) are better served by generic or specific health terms or whether both are needed.

We found that both specific and generic health terms were necessary. They each uniquely identified some of the review's more robust studies. They also identified different types of health outcome. Failure to identify some of those outcomes would have directly affected our review's conclusions. Had we only used generic health terms in our search we would have missed around half the studies that we finally included in the review. Likewise, focusing exclusively on specific health terms in the literature search would have failed to identify around half the included papers. A small proportion of these studies would have been identified by our manual search and bibliography check but failing to conduct either of the ELS approaches would still have led to a serious 'loss' of data (or, more correctly, a failure to find data) that would have compromised the integrity and accuracy of our review's findings.

We found that the specific search tended to miss studies with general or multiple health outcomes, while the generic search tended to miss studies with single specific health outcomes. This may appear intuitive, but we contend that the finding is actually surprising. It suggests, for example, that studies that look specifically at young people's diabetes, epilepsy and headache tend not to be identifiable by search terms such as 'health status', 'health surveys', 'child health', 'adolescent health', 'health status indicators', 'symptoms', 'morbidity', 'health complaints', etc. It also suggests that some studies that, for example, included headache as one of a number of different health outcomes may be identified by a search strategy that includes generic health terms but could be missed by an ELS that specifically focuses on the term 'headache.'

This finding is at odds with what some authors of this paper initially expected. Prior to our exploring this issue, the authors assumed that the generic health search

would identify the vast majority of included studies, while the specific search would mainly identify a subset of those studies. If other systematic reviewers also make this assumption, then their reviews are at risk of being based on poor-quality (highly insensitive) searches.

Strengths and limitations

We have conducted a prospective comparative study of two electronic literature search strategies that have been field tested while we conducted a systematic review. This kind of study is uncommon and hence novel, while the prospective and comparative design is a key strength.

The main limitations of this study are that it is based on a single review and the search was not sensitive (ie, lacking in the use of truncation, synonyms and related terms). The review that we based the study on does not focus on the effectiveness of an intervention, which means that precision cannot be easily increased by including simple study design search terms, and the outcomes are also very complex, which probably increases the difficulty of sensitive and specific searching. These may be regarded as unusual features affecting the generalisability of our findings but we have argued in our introduction that 'unusual' (ie, not clinical intervention) reviews are becoming more common and hence are an emerging priority in terms of review methods. The same may be said about time-limited reviews. Ours took 18 months to complete—not an unusual time frame in our experience—but we are aware that some systematic reviews (eg, many Cochrane and Campbell reviews) take longer and involve more comprehensive searches.

It may also be hypothesised that conducting a more extensive ELS and manual search could have led to a greater number of, and possibly more overlap between, studies identified by each component of our search strategy. Ways to achieve a more extensive search could have included using more electronic databases and other relevant data sources, identifying a wider number of synonyms for both the health outcomes and other concepts included in the review, using both subject headings and words in the title and abstract to search for every concept in the search strategy and minimising reliance on the accuracy of database indexers. There is also some existing evidence that the effectiveness of different search strategies may vary depending on the subject of the review.⁵ Therefore, it is worth testing our findings in the context of other reviews and different types of literature search, including more sensitive searches. Missing out health outcomes altogether is an alternative means of increasing search sensitivity but we note that our initial search identified well in excess of 10 000 hits. Given the broad review question, attempts to vastly expand the search risked increasing the number of hits to unmanageable levels.

Implications and conclusions

Literature searching has a vital role to play in evidence-informed policy and practice, and it is plausible to theorise a direct pathway by which a poor search may

lead to harmful decisions. Conducting research that may assist information scientists and reviewers to improve their search strategies should therefore be a priority. Such research can be nested within the processes of conducting systematic reviews: from our own experience, this requires minimal additional resources to the cost of the overall review and can therefore be considered an inexpensive way of conducting useful research in an important field. We therefore hope that other reviewers will make use of similar opportunities to explore how best to optimise electronic searching.

In light of our findings, we recommend that future systematic reviews of topics that involve multiple health outcomes include both generic and specific health terms in their literature search (if a health outcome search is considered necessary), along with manual searching. Choosing only one of these search components could, based on our findings, increase the risk of reviewers missing robust evidence and making misleading conclusions.

Acknowledgements Candida Fenton (MRC/CSO SPHSU) provided advice as information scientist. Mary Robins (MRC/CSO SPHSU) helped retrieve papers. As Director of MRC/CSO SPHSU, Sally Macintyre read and approved the manuscript.

Contributors ME helped to plan and conduct the study, analyse the findings, led on writing the manuscript and is guarantor for the study. AM, HS and KH helped to plan the study and conduct the study, analyse the findings and provide content and comments on the manuscript. All authors, external and internal, had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding ME, AM, HS and KH are core funded by the Medical Research Council (5TK50; 5TK40). ME is also core funded by the Chief Scientist Office (part of the Scottish Government Health Directorates). The authors declare that the research was conducted independently from the funders: the funders played no part in the study design; in the collection, analysis and interpretation of data; in the writing of the report and in the decision to submit the article for publication.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The review protocol (including search strategies) and a list of studies included in the final review are available in the supplemental documents submitted with this article. Further data related to the searches are available from the corresponding author at matt@sphsu.mrc.ac.uk.

REFERENCES

- Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2*. (updated September 2009). The Cochrane Collaboration, 2009.
- NHS Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare*. York: CRD University of York, 2009.
- Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell Publishing, 2006.
- Ogilvie D, Hamilton V, Egan M, et al. Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *J Epidemiol Community Health* 2005;59:804–8.
- Bambra C. Real world reviews: a beginner's guide to undertaking systematic reviews of public health policy interventions. *J Epidemiol Community Health* 2011;65:14–19.
- Glanville JM, Lefebvre C, Miles JN, et al. How to identify randomized controlled trials in MEDLINE. *J Med Libr Assoc* 2006;94:130–6.
- Golder S, McIntosh HM, Duffy S, et al; Centre for Reviews and Dissemination and UK Cochrane Centre Search Filters Design Group. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J* 2006;23:3–12.
- Montori VM, Wilczynski NL, Morgan D, et al. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68.
- Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. *J Med Libr Assoc* 2006;94:41–7.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286–91.
- Eisinga A, Siegfried N, Clarke M. The sensitivity and precision of search terms in Phases I, II and III of the Cochrane Highly Sensitive Search Strategy for identifying reports of randomized trials in MEDLINE in a specific area of health care - HIV/AIDS prevention and treatment interventions. *Health Info Libr J* 2007;24:103–9.
- Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care* 2003;19:591–603.
- Joyce K, Pabayo R, Critchley JA, et al. Flexible working conditions and their effects on employee health and wellbeing. *Cochrane Database Syst Rev* 2010;(2):CD008009.
- Porta D, Milani S, Lazzarino A, et al. Systematic review of epidemiological studies on health effects associated with management of solid waste. *Environ Health* 2009;8:60.
- Uscher-Pines L. Health effects of relocation following disaster: a systematic review of the literature. *Disasters* 2009;33:1–22.
- D'Onise K, Lynch J, Sawyer M, et al. Can preschool improve child health outcomes? A systematic review. *Soc Sci Med* 2010;70:1423–40.
- Egan M, Petticrew M, Ogilvie D, et al. New Roads and Human health: a systematic review. *Am J Public Health* 2003;93:1463–71.
- Dangour A, Aikenhead A, Hayter A, et al. *Comparison of Putative Health Effects of Organically and Conventionally Produced Foodstuffs: A Systematic Review: Report for the Food Standards Agency*. London: Nutrition and Public Health Intervention Research Unit London School of Hygiene & Tropical Medicine, 2009.
- Vickers AJ. Can acupuncture have specific effects on health? A systematic review of acupuncture antiemesis trials. *J R Soc Med* 1996;89:303–11.
- Daykin N, Orme J, Evans D, et al. The impact of participation in performing arts on adolescent health and behaviour: a systematic review of the literature. *J Health Psychol* 2008;13:251–64.
- Rogers G, Elston J, Garside R, et al. The harmful health effects of recreational ecstasy: a systematic review of observational evidence. *Health Technol Assess* 2009;13:iii-iv, ix-xii, 1-315.
- Passchier W, Knottnerus A, Albering H, et al. Public health impact of large airports. *Rev Environ Health* 2000;15:83–96.
- Egan M, Petticrew M, Ogilvie D, et al. "Profits before people"? A systematic review of the health and safety impacts of privatising public utilities and industries in developed countries. *J Epidemiol Community Health* 2007;61:862–70.
- McCartney G, Thomas S, Thomson H, et al. The health and socioeconomic impacts of major multi-sport events: systematic review (1978-2008). *BMJ* 2010;340:c2369.
- Paradies Y. A systematic review of empirical research on self-reported racism and health. *Int J Epidemiol* 2006;35:888–901.
- Vartanian LR, Schwartz MB, Brownell KD. Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. *Am J Public Health* 2007;97:667–75.
- Bambra C, Whitehead M, Sowden A, et al. "A hard day's night?" The effects of Compressed Working Week interventions on the health and work-life balance of shift workers: a systematic review. *J Epidemiol Community Health* 2008;62:764–77.
- Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;331:1064–65.
- Sweeting H. Reversals of fortune? Sex differences in health in childhood and adolescence. *Soc Sci Med* 1995;40:77–90.