PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	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.
AUTHORS	Matt Egan, Alice MacLean, Helen Sweeting and Kate Hunt

VERSION 1 - REVIEW

REVIEWER	Mark Petticrew, Faculty of Public Health and Policy, LSHTM, UK.
REVIEW RETURNED	14/03/2012

RESULTS & CONCLUSIONS	I've answered "no" to 2 questions, as they need further discussion because some more analysis would be useful - see comments
GENERAL COMMENTS	This is a useful paper and the findings are pretty remarkable if they are generalisable – i.e. that using either generic or specific searches risks missing half the evidence, as the title states. I have some reservations about whether this really is the case however. For half the evidence to be missed one would have to assume that reviewers would only search electronic databases, whereas in reality surely researchers, particularly those conducting complex reviews, would also conduct handsearches and searches of bibliographies, (e.g. because of the risk of publication bias). Would (did) further handsearching, and checking of bilbiographies pick up the missing studies anyway? If this is the case, then the evidence is not really "missed", it is simply picked up in a different part of the search. Also, this is not a "typical" systematic review. The review question is more complex than most, and it is not a review of the effectiveness of an intervention, which means that specificity cannot be easily increased by including simple study design search terms, and the outcomes are also very complex. There is also no intervention, which probably makes searching both less sensitive and specific. Perhaps all this could be clarified in the limitations section. It also means that the real implication is that reviewers should not confine themselves to electronic searches alone. This has particular implications for those who do particular types of review that might rely on electronic searches alone, because of limitations in time or resources – e.g. scoping reviews for example. The statement that failure to identify half of the studies would have compromised the accuracy of the reviews findings probably needs more analysis to support it. This statement would only be true if the review's findings could be shown to be affected by only including one or other sets of studies, but this is not part of the analysis. In the absence of such an analysis, it could be argued that it might not matter which search was run – for example, it might not matter if all there were only 3 high quality, reliable studies anyway, and these were the ones which were already picked up by both searches. This seems unlikely for this review, but is entirely possible for reviews of

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

	RCTs for example. It would be useful to know at least whether the quality of the studies differs between the two groups. This information is already available, according to the protocol. In the results section, there is not much analysis of the differences between the studies identified by the two searches, just an impressionistic comment that the search tended to be more focussed on a single health outcome. Do studies differ in other respects? (Design, quality, country, what types of journals they appear in?) Some further analysis here would be informative to back up the general statement about the difference in number of health outcomes between groups - e.g. an average (or median) number of outcomes per study for the two groups is calculable (i.e. approx 100 outcomes in the 21 studies; compared to 22 outcomes across the 17 studies; though the number depends on whether one counts individual symptoms or not). This would give a more concrete sense of the difference between the groups of studies. Similarly, do the countries (or journals) the studies tend to come from differ? It is difficult to tell as the country is included for some studies and not others; however this might affect “locatability” of evidence. Despite the above comments the study is publishable, with probably just some further clarification or discussion needed. Minor points As I was reading the paper, an alternative explanation for the findings seemed to be that no information scientist was involved in designing or conducting the searches – so perhaps the reason for the lack of overlap was that searches simply had not been well done. It wasn’t until page 14 (in the strengths and limitations section) that it became clear that an information scientist had been involved after all, so this explanation could be ruled out. It would be useful to include this information earlier in the methods section.

REVIEWER	Julie Glanville Project Director - Information Services York Health Economics Consortium University of York Heslington York YO10 5NH UK I am an information professional with over twenty year's experience of identifying evidence for systematic reviews in a wide range of topics including health, social care, criminal justice, education and public health.
REVIEW RETURNED	26/03/2012

THE STUDY	This is not a patient study and does not require assessment of a patient population. The methods are clearly described with the exception of the literature search to underpin the research question and the actual search strategies employed in the review - these are in the accompanying review but really need to be in the main paper. The limitations of the paper need to include a discussion of the limitations of the whole search strategy adopted: the choice of all concepts, the search terms in each concept and the limits employed,

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

	impact on the sensitivity and precision of the searches. Systematic review searches of asthma for example, typically employ a range of synonyms and related terms to find that topic, but that does not seem to have occurred with this review. The strategy employed does not seem extensive enough for the purposes of a systematic review, but there may be all sorts of reasons for the choice of approach and search terms which the authors could share with us to explain the approach chosen. The authors don't report the sensitivity, precision and number needed to read for their searches - these search strategy performance measures are standard when reporting searches and can all be calculated from the results presented in the searches which they undertook.
RESULTS & CONCLUSIONS	The authors claim novelty but the approach of using generic and specific search terms to achieve sensitivity has been established for a long time in information retrieval in systematic reviews. The experiment which the authors have conducted is more novel, but their conclusions are rather undermined by the searches which were employed. Those searches (reported in the accompanying paper) are not extensive and do not follow some of the basic methods of systematic searching undertaken for systematic reviews: i.e. using both subject headings and words in the title and abstract to search for each concept in the search strategy (see the Cochrane Handbook, CRD guidance and Petticrew monograph cited by the authors). The authors tend to only use subject headings, and where they do use textwords tend to use words in the abstract. This means the sensitivity of their strategy as whole is not optimal for any of the concepts in their strategy. The authors use a three concept approach in their strategies, with the 'sex factors' concept in particular being a very limited concept in terms of its sensitivity: ie only a few search terms are relied on to ensure this concept is retrieving effectively. Relying on the database indexers to spot the focus of the papers is optimistic especially given such a difficult topic. The authors do not discuss the impact of having three concepts and how the other concepts were captured, on the overall sensitivity of their strategies: their focus is on one concept, but the whole strategy is interdependent. For example the limit to Humans in MEDLINE carries with it further reliance on the database indexers and there are records in MEDLINE about human issues without the Humans tag.
GENERAL COMMENTS	1. The authors refer to the their strategy as a filter. Filters tend to be used for the parts of search strategies which search for study design methods (see ISSG search filter resource at CRD, University of York). Filters also carry with them the connotation of a search strategy which has been designed and tested (even validated) to perform in a specific way each time it is run. In the light of this, the authors may wish to refer to their 'filters' as search strategies. 2. p.7 para 2. The authors should present their literature review methods to support their assumptions. 3. The authors should present their search strategies and their rationale for the choice of concepts and limits in the search strategies because the concept they have focused on is only one part of a whole. The PRESS checklist may prove helpful for this purpose (http://cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf4.)

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

	4. The authors refer to specificity repeatedly but in fact precision (and number needed to read) is the core issue which they are trying to optimise. They want the highest possible sensitivity for the highest precision. Clarifying this will be helpful especially for statements such as line 30 in the introduction section (p4) where the authors suggest the 'search strategies that are too sensitive' might be a bad thing: However, high sensitivity is nearly always desirable in systematic reviews, it is low levels of precision which are undesirable. I suspect that the authors would ideally like highly sensitive searches but would prefer them to be highly precise as well. 5. In the discussion the authors make useful points but they have rather reinvented the wheel. The use of the generic and specific approaches together, to compensate for author variability in reporting research is standard in systematic reviews. If the authors look at some systematic review reports or NICE public health guidance informed by systematic searches they will see this approach in common use. The usefulness of the authors' experiment is in providing a worked example.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Mark Petticrew, Faculty of Public Health and Policy, LSHTM, UK.

MP: For half the evidence to be missed one would have to assume that reviewers would only search electronic databases, whereas in reality surely researchers, particularly those conducting complex reviews, would also conduct handsearches and searches of bibliographies, (e.g. because of the risk of publication bias). Would (did) further handsearching, and checking of bilbiographies pick up the missing studies anyway? If this is the case, then the evidence is not really “missed”, it is simply picked up in a different part of the search.

AUTHORS: We agree with the handsearching point raised. The two electronic searches each identified around half the studies included in the review – but our discussion should clarify the importance of using both generic and specific electronic searches along with handsearching. Before conducting the ELS, we handsearched private collections (two of the authors have each been researchers in this field for around two decades, another author for several years), conducted a serendipitous internet search and used tracked papers that had cited the seminal early paper for this topic (Helen Sweeting’s earlier review): from this handsearch four studies were finally included in the review (see Methods p.10). All four of these studies were also identified using the generic search. Bibliography checking of included studies makes little difference to our results: one study currently classed as being identified from our specific ELS was also referenced in a study identified through our general search. Similarly, one study we identified from the generic ELS was also referenced in a study identified from the specific search and was also one of the 4 papers we found in our initial handsearch. We have revised the article to include this information. It means that the generic ELS in combination with the manual search and bibliography check would have identified 24 of the 41 included studies. The specific ELS in combination with the handsearch would have identified 25 of the 41 included studies. Either way, this still represents a substantial number of studies that would have been missed by adopting only one of the two ELS approaches (see results page 13).

Interestingly, studies that were identified by the generic search tended to reference other studies identified by the generic search. Similarly, studies that were identified by the specific search sometimes referenced other studies identified by the specific search. We speculate (and therefore have not put this in the paper) that this suggests our two ELSs identified not only different studies, but

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

also different streams of research – with relatively little cross-referencing between streams.

Finally, we note that searches can always be broadened (at least that is our experience from conducting broader public health reviews – there always seem to be other electronic and handsearch sources that could be checked, and other ways of widening the search criteria). With this in mind, we think it could be hypothesized that if we had only made our handsearch and ELS even more extensive, then this may have potentially led to greater overlap between the different search strategies. We have added this point to our limitations section (see p.17-18).

MP: Also, this is not a "typical" systematic review. The review question is more complex than most, and it is not a review of the effectiveness of an intervention, which means that specificity cannot be easily increased by including simple study design search terms, and the outcomes are also very complex. There is also no intervention, which probably makes searching both less sensitive and specific.
Perhaps all this could be clarified in the limitations section. It also means that the real implication is that reviewers should not confine themselves to electronic searches alone. This has particular implications for those who do particular types of review that might rely on electronic searches alone, because of limitations in time or resources – e.g. scoping reviews for example.

AUTHORS: We agree that the study has particular relevance to reviewers who do 'atypical' reviews, and reviews with time or other resource limitations. The original article made some reference to these issues. In the revised introduction we make clear the following "Our review was conducted within a limited time frame (originally planned as nine months and then extended to 18 months), and we believe the implications of this study are of particular relevance to reviews of broader public health topics and reviews with time or other resource limitations."(p.8 – and the limitations section also refers to this issue). Eighteen months is, from our experience, not an unusual time frame for a systematic review but we are aware that some reviews (e.g. many Cochrane and Campbell reviews) take longer (a point we make in the limitations section p.17).

MP: The statement that failure to identify half of the studies would have compromised the accuracy of the reviews findings probably needs more analysis to support it. This statement would only be true if the review's findings could be shown to be affected by only including one or other sets of studies, but this is not part of the analysis. In the absence of such an analysis, it could be argued that it might not matter which search was run – for example, it might not matter if all there were only 3 high quality, reliable studies anyway, and these were the ones which were already picked up by both searches. This seems unlikely for this review, but is entirely possible for reviews of RCTs for example. It would be useful to know at least whether the quality of the studies differs between the two groups. This information is already available, according to the protocol.

AUTHORS: Table 1 of the study shows that the two different ELSs tended to identify different types of health outcome. In the revised version we have added that using only one type of ELS would have meant missing all or most of the evidence for some of these health outcomes and that this would have affected the review's conclusions. Most notably, the specific search (i.e. not the generic, nor the handsearch/bibliography check) identified all three included studies of epilepsy and all but one of the 7 studies on diabetes. The review's findings for these two health outcomes were markedly different to findings for the all the other included outcomes: findings for diabetes and epilepsy tended to run counter to our review's main hypotheses, where as findings for other types of health outcome tended to support the hypothesis or were inconclusive. Therefore, a review that omitted the specific ELS would have provided us with plenty of evidence to support our hypothesis but would have left us unaware of evidence suggesting that the inverse of our hypothesis may apply to certain health conditions – a key conclusion for our review. (we include this information on p.14-15).

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

MP: In the results section, there is not much analysis of the differences between the studies identified by the two searches, just an impressionistic comment that the search tended to be more focussed on a single health outcome. Do studies differ in other respects? (Design, quality, country, what types of journals they appear in?) Some further analysis here would be informative to back up the general statement about the difference in number of health outcomes between groups - e.g. an average (or median) number of outcomes per study for the two groups is calculable (i.e. approx 100 outcomes in the 21 studies; compared to 22 outcomes across the 17 studies; though the number depends on whether one counts individual symptoms or not). This would give a more concrete sense of the difference between the groups of studies. Similarly, do the countries (or journals) the studies tend to come from differ? It is difficult to tell as the country is included for some studies and not others; however this might affect “locatability” of evidence.

AUTHORS: Following MP’s advice we have revised the tables in the supplemental document to include journal title, study design, appraisal score and country. The studies that would have been missed had there been a handsearch and specific ELS but NOT a generic ELS included several of the more robust studies identified by the review (e.g. 3 longitudinal studies and 3 studies classed as high scoring following the study appraisal). The studies that would have been missed had there been no specific ELS included 5 of the higher scoring studies (no longitudinal studies would have been missed). The general search identified studies from a wider range of European and North American countries compared to the specific search, which only identified European studies. All the studies identified were from medical/health journals. (see supplemental document 2, and main document p.15).

MP: Despite the above comments the study is publishable, with probably just some further clarification or discussion needed.

AUTHORS: we thank the reviewer for this opinion - the study has been revised to clarify.

Minor points
As I was reading the paper, an alternative explanation for the findings seemed to be that no information scientist was involved in designing or conducting the searches – so perhaps the reason for the lack of overlap was that searches simply had not been well done. It wasn’t until page 14 (in the strengths and limitations section) that it became clear that an information scientist had been involved after all, so this explanation could be ruled out. It would be useful to include this information earlier in the methods section.

AUTHORS: The information scientist is now referred to in the methods section.

Reviewer: Julie Glanville
Project Director - Information Services
York Health Economics Consortium
University of York
UK

I am an information professional with over twenty year's experience of identifying evidence for systematic reviews in a wide range of topics including health, social care, criminal justice, education and public health.

This is not a patient study and does not require assessment of a patient population.

The methods are clearly described with the exception of the literature search to underpin the research

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

question and the actual search strategies employed in the review - these are in the accompanying review but really need to be in the main paper.

AUTHORS: BMJ OPEN is designed to be accessed online, so supplemental documents, tables and figures should be accessible by a single click. Keeping the search strategy in the supplemental protocol document rather than moving it to the main paper ensures that this document is consistent with other versions of the protocol that have been made available when the report of the actual systematic review is published. We have added an additional statement in the methods section directing readers to the supplemental document to obtain further details about the search (see main document, p.10)

JG: The limitations of the paper need to include a discussion of the limitations of the whole search strategy adopted: the choice of all concepts, the search terms in each concept and the limits employed, impact on the sensitivity and precision of the searches. Systematic review searches of asthma for example, typically employ a range of synonyms and related terms to find that topic, but that does not seem to have occurred with this review. The strategy employed does not seem extensive enough for the purposes of a systematic review, but there may be all sorts of reasons for the choice of approach and search terms which the authors could share with us to explain the approach chosen.

AUTHORS: The limitations section has been revised following comments from both reviewers (see p.18) – we now suggest that the extent to which different components of a search identify similar or different studies may be influenced by both the subject matter being reviewed and the extensiveness of each search component.

The first paragraph of the introduction now discusses different approaches to systematic reviewing, pointing out that not all systematic reviews aim for complete comprehensiveness (reasons for this include resource constraints, efficiency and urgency). We also note there exists a variety of opinion amongst reviewers (and others interested in research synthesis and evidence informed policy and practice) on the opportunity cost of always aiming for comprehensiveness – particularly if there are cases when a less sensitive search is able to identify the most robust evidence available (as suggested by MP's hypothetical argument - "it could be argued that it might not matter which search was run – for example, it might not matter if all there were only 3 high quality, reliable studies anyway, and these were the ones which were already picked up by both searches." – see also Bambra, JECH 2011;65:14-19). On the other hand, our paper presents evidence to support reviewers and information scientists who emphasise the importance of sensitivity: our findings highlight the danger of missing important evidence if reviewers fail to include both generic and specific health terms. As we know of previous reviews that have not included both approaches, we think this is a useful message for future reviewers.

JG: The authors don't report the sensitivity, precision and number needed to read for their searches - these search strategy performance measures are standard when reporting searches and can all be calculated from the results presented in the searches which they undertook.

AUTHORS: Figure 1 has been annotated to include these details.

JG The authors claim novelty but the approach of using generic and specific search terms to achieve sensitivity has been established for a long time in information retrieval in systematic reviews. The experiment which the authors have conducted is more novel, but their conclusions are rather undermined by the searches which were employed. Those searches (reported in the accompanying paper) are not extensive and do not follow some of the basic methods of systematic searching undertaken for systematic reviews: i.e. using both subject headings and words in the title and abstract

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

to search for each concept in the search strategy (see the Cochrane Handbook, CRD guidance and Petticrew monograph cited by the authors). The authors tend to only use subject headings, and where they do use textwords tend to use words in the abstract. This means the sensitivity of their strategy as whole is not optimal for any of the concepts in their strategy. The authors use a three concept approach in their strategies, with the 'sex factors' concept in particular being a very limited concept in terms of its sensitivity: ie only a few search terms are relied on to ensure this concept is retrieving effectively. Relying on the database indexers to spot the focus of the papers is optimistic especially given such a difficult topic. The authors do not discuss the impact of having three concepts and how the other concepts were captured, on the overall sensitivity of their strategies: their focus is on one concept, but the whole strategy is interdependent. For example the limit to Humans in MEDLINE carries with it further reliance on the database indexers and there are records in MEDLINE about human issues without the Humans tag.

AUTHORS: We agree that the use of generic and specific searches are not novel, and that our use of the word 'novel' applies to the experimental approach that compares outcomes from the two searches. We have revised the limitations section to include the reviewer's comments about search sensitivity.
"It may also be hypothesised that conducting a more extensive ELS and handsearch 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. Therefore, it is worth testing our findings in the context of other reviews and different types of literature search, including more sensitive searches." (see p.18)

1. The authors refer to the their strategy as a filter. Filters tend to be used for the parts of search strategies which search for study design methods (see ISSG search filter resource at CRD, University of York). Filters also carry with them the connotation of a search strategy which has been designed and tested (even validated) to perform in a specific way each time it is run. In the light of this, the authors may wish to refer to their 'filters' as search strategies.

AUTHORS: we have revised our use of the word 'filters'.

2. p.7 para 2. The authors should present their literature review methods to support their assumptions.

AUTHORS: we were unsure what this comment refers to, but please note that we have included the review protocol as a supplemental document describing the review and its methods in more detail.

3. The authors should present their search strategies and their rationale for the choice of concepts and limits in the search strategies because the concept they have focused on is only one part of a whole. The PRESS checklist may prove helpful for this purpose (http://cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf4.)

AUTHORS: We believe the reviewer has already helpfully identified several ways to increase the extensiveness of our searches and these are now detailed in the limitations sections (see earlier response). We also emphasise that the review was time limited (18 months) and that the search strategy was devised to be conducted within that time (see page 10 and 17). Nonetheless we reiterate that these kinds of constraints are quite common for systematic reviews.
Furthermore, our searches did identify over 10000 hits – and given the broad research question, we

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

think it reasonable to assume that increasing the sensitivity of every concept within the entire search could easily have expanded that figure exponentially.

Although approaches to literature searching may vary, a key issue for the design of this study is that both the specific and generic search strategies are comparable in all aspects except for the aspect of the search that we are testing (i.e. the health terms). This comparability is the basis of our claim to internal validity. The next question (implied by JG and also raised by MP) is about generalisability – i.e. would different information scientists and reviewers running different but comparable searches to our own reach a similar general conclusion? Our limitations section already makes clear that more research is needed to test whether our findings are generalisable "to other reviews" – to this we have now added "and different types of literature search, including more sensitive searches."

4. The authors refer to specificity repeatedly but in fact precision (and number needed to read) is the core issue which they are trying to optimise. They want the highest possible sensitivity for the highest precision. Clarifying this will be helpful especially for statements such as line 30 in the introduction section (p4) where the authors suggest the 'search strategies that are too sensitive' might be a bad thing: However, high sensitivity is nearly always desirable in systematic reviews, it is low levels of precision which are undesirable. I suspect that the authors would ideally like highly sensitive searches but would prefer them to be highly precise as well.

AUTHORS: We have revised the text accordingly.

5. In the discussion the authors make useful points but they have rather reinvented the wheel. The use of the generic and specific approaches together, to compensate for author variability in reporting research is standard in systematic reviews. If the authors look at some systematic review reports or NICE public health guidance informed by systematic searches they will see this approach in common use. The usefulness of the authors' experiment is in providing a worked example.

AUTHORS: Like the reviewer, we are aware of reviews that use both generic and specific approaches. However, we are also aware of reviews that use just one of these approaches. We hope our worked example can provide evidence to help convince future reviewers that using both approaches is the more sensible strategy to adopt – although we hope future reviewers will also test that assumption just as we have done, so that more information about generalisability can be obtained. We also think our general methodological approach is of interest because it might be adapted to shed light on other aspects of literature searching that currently lack evidence.

VERSION 2 – REVIEW

REVIEWER	Julie Glanville Project Director - Information Services York Health Economics Consortium University of York York United Kingdom
REVIEW RETURNED	25/04/2012

THE STUDY	I have answered 'No' to some of the questions because the questions do not all really fit this paper: the paper is not about a clinical situation. 1. In the title 'electronic search terms' might be better expressed as 'search terms in electronic databases' 2. The running title is different to the article title: this could be confusing? 3. The abstract conclusion ("Future systematic reviews that involve multiple health outcomes should include both generic and specific health

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

	terms in their literature search") is quite strong based on this single research study. It might be better to indicate that the findings of this study are in agreement with searching conventions.
RESULTS & CONCLUSIONS	1. The authors should define 'recall' or better still replace it with 'sensitivity' which is more frequently used these days in information retrieval research. 2. The authors state: "Either way, there is a pressing need to learn more about how best to negotiate the competing demands of specificity and sensitivity." It's really 'precision' that is the issue for searchers and reviewers, rather than specificity. The trade-off between sensitivity and precision is a well known and well rehearsed challenge in information retrieval for systematic reviews. 3. The authors state: "systematic reviews that focus on a more general population sample, have no intervention, and/or are not limited to a single study design, lack one or more of these three search components, and so result in a less specific ELS." There are a lot of issues embedded in this sentence and the authors should tease these more clearly. Not all searches or searchers follow the PICO conceptual breakdown and there are approaches which are used in the circumstances described. 4. The authors state "All these challenges increase the chances of a search becoming lengthy and costineffective." These are two issues. The length of a search strategy does not mean it is inefficient - perhaps the authors are referring to the number of results? What is a cost-ineffective search? 5. The authors state: 'look for alternative means of increasing search specificity but there has been relatively little guidance on how this can be achieved without compromising sensitivity." Again the issue is really about improving precision rather than specificity and the tradeoff between sensitivity and precision is a long discussed issue within information retrieval for systematic reviews. 6. The authors state: "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" 7. To be able to make this statement the authors must have undertaken a literature review to look for studies – the authors should report their literature review. 8. The authors discuss the "size of the searches". Please clarify if this relates to the number of search lines, the number of resources searched or the number of search results. 9. The authors don't report the search strategies they tested – in a paper which focuses on search strategies it is essential that the reader can see them as part of the paper, so that s/he can understand what is going on in the paper. If the strategies are included in the paper the authors can save some of the explanatory text and directions to the supplementary documents 10. The authors note that: "The precise search strategy differed between databases if different search facilities and search engines made it necessary to adapt our approach." – adapting strategies can require further tradeoffs and can have quite radical effects on the search results ,so it would be helpful to have an example of adaptation and a brief discussion of the issues.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

	11. The authors mention handsearches at various points and it would be helpful to clarify what these involve. Typically handsearching involves the selection of a number of journals and the assessment of relevance of papers, page by page. If handsearching was undertaken it is usual to report which journals, how they were selected, and how the handsearch was accomplished as well as the results? It may be that by 'handsearching' the authors are referring to checking the references at the end of relevant papers. If this is the case then 'handsearching' should not be referred to, especially not in the abstract where it may mislead searchers looking for studies where handsearching has been used. 12. The authors mention "robust studies": it would be helpful to know how robustness was defined or measured. 13. The authors state that "This may appear intuitive, but we contend that the finding is actually surprising." I think the authors are struggling with this claim. What they have is a useful empirical example of what has been known and acted on for a long time in SR searching but not formally explored. There are few concrete published examples of a parallel exploration of different search approaches, so my suggestion is that the authors focus on this aspect rather than claiming novelty for a widely used technique. 14. The search strategy is a major weakness of this paper as it is not sensitive, lacking in the use of truncation, synonyms and related terms. However, as a worked example, with a candid statement about its weaknesses, this is still useful. 15. The authors suggest a 'more extensive search' would have been ideal. I am not sure whether this refers to searching more resources, but I suggest that before moving on to more resources, a more sensitive (extensive) search strategy within the databases they did search would be the first priority.
REPORTING & ETHICS	There is no reporting guidance for papers of this type, but search strategies can be assessed using the PRESS checklist: http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402

VERSION 2 – AUTHOR RESPONSE

3. The abstract conclusion ("Future systematic reviews that involve multiple health outcomes should include both generic and specific health terms in their literature search") is quite strong based on this single research study. It might be better to indicate that the findings of this study are in agreement with searching conventions.

Authors: We take on board the reviewer's point about our findings fitting conventions for good practice. We are also aware of reviews that have only used generic or specific health outcomes terms (rather than both) and we think it worthwhile to warn others against this approach – whilst acknowledging that further research is needed to test the generalisablity of our findings. Therefore we have made the following revision to abstract conclusion.

"Whilst 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 our 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. "

4. The authors should define 'recall' or better still replace it with 'sensitivity' which is more frequently used these days in information retrieval research.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

Authors: we changed ‘recall’ to ‘sensitivity’.

5. The authors state: “Either way, there is a pressing need to learn more about how best to negotiate the competing demands of specificity and sensitivity.” It’s really ‘precision’ that is the issue for searchers and reviewers, rather than specificity. The trade-off between sensitivity and precision is a well known and well rehearsed challenge in information retrieval for systematic reviews.

Authors: We have changed ‘specificity’ to ‘precision’ and have noted that the trade off between precision and sensitivity is a well known challenge in information retrieval for systematic reviews. We made the following revisions.

“The trade-off between screening out irrelevant evidence whilst 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.” (page 4)

6. The authors state: “systematic reviews that focus on a more general population sample, have no intervention, and/or are not limited to a single study design, lack one or more of these three search components, and so result in a less specific ELS.” There are a lot of issues embedded in this sentence and the authors should tease these more clearly. Not all searches or searchers follow the PICO conceptual breakdown and there are approaches which are used in the circumstances described.

Authors: our intention is not to suggest that all searches follow the PICO conceptual breakdown but rather to suggest that a three concept approach (population, intervention and study design) is often used in reviews with a clinical focus and is a well known approach advocated in systematic review guidance including the Cochrane Handbook. See revised paragraph.

“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, i.e. 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 randomized 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, for example, 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).” (page 6-7)

7. The authors state “All these challenges increase the chances of a search becoming lengthy and cost ineffective.” These are two issues. The length of a search strategy does not mean it is inefficient - perhaps the authors are referring to the number of results? What is a cost-ineffective search?

Authors: we have revised to clarify.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

“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.” (page 7)

8. The authors state: ‘look for alternative means of increasing search specificity but there has been relatively little guidance on how this can be achieved without compromising sensitivity.” Again the issue is really about improving precision rather than specificity and the tradeoff between sensitivity and precision is a long discussed issue within information retrieval for systematic reviews.

Authors: we have revised to use the term ‘precision’. The tradeoff between sensitivity and precision is a long discussed issue within information retrieval for systematic reviews – but for a long time much of the guidance on ELS for systematic reviews focused on the most common type of systematic review (clinical interventions, RCTs). Even now (because there has been an accumulation of work over the years), we would suggest this clinical/RCT focus is still discernable in much of the systematic review guidance. Reviewers conducting alternative types of systematic review are not always going to find ‘long discussed’ methods of improving precision useful – e.g . the work that has gone into developing RCT filters over the years is not going to be directly useful to reviewers who are not looking for RCTs. Different kinds of systematic review may, we believe, require different types of guidance about how to negotiate the sensitivity-precision trade off – hence, as systematic reviews continue to develop there will be an ongoing need to consider and reconsider this trade off.

In our introduction we make the case that more guidance is needed for other types of systematic review (e.g. more broadly focused reviews that have emerged from the field of public health). In our view, alternative kinds of systematic review have particular challenges of their own. In our introduction we have outlined some of these challenges (see previous response). See further revision below:

“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 to the guidance on clinical/RCT systematic reviews).” (page 7)

9. The authors state: “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” To be able to make this statement the authors must have undertaken a literature review to look for studies – the authors should report their literature review.

Authors: The statement in question is correct in so far as the authors do not know of any study that fits this description. However, we base this on a non-systematic exploration of the literature. We have now revised to make clear that we have not systematically reviewed this issue.
“(both of the above observations are based on a non-systematic exploration of the literature rather than a systematic review).” (page 8)

We think it reasonable to suggest that studies using similar methods to ours are not common or not commonly known and we note that the reviewer also states that: “There are few concrete published examples of a parallel exploration of different search approaches.”

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

10. The authors discuss the “size of the searches”. Please clarify if this relates to the number of search lines, the number of resources searched or the number of search results.

Authors: We refer to the number of search results: “(i.e. the number of references initially identified from the ELS – sometimes referred to as the number of ‘hits’)” (page 8-9)

11. The authors don’t report the search strategies they tested – in a paper which focuses on search strategies it is essential that the reader can see them as part of the paper, so that s/he can understand what is going on in the paper. If the strategies are included in the paper the authors can save some of the explanatory text and directions to the supplementary documents

Authors: we have included the searches in a new table – table 1 (pages 11 to 15). However, we still consider it necessary to include the review protocol as a supplemental document to give those interested the opportunity to obtain further details about all aspects of the review. This means the search strategy now appears twice – i.e. in the table and in supplemental document 1. We are still unsure as to whether this repetition is necessary – particularly as the table for the online article will, we assume, require a mouse click before it can be viewed, similar to the supplemental document. If the journal editors have a view about that, we would be happy to accommodate that view.

12. The authors note that: “The precise search strategy differed between databases if different search facilities and search engines made it necessary to adapt our approach.” – adapting strategies can require further tradeoffs and can have quite radical effects on the search results ,so it would be helpful to have an example of adaptation and a brief discussion of the issues.

Authors: this is related to later points about providing more details and discussion of the search and its limitations. See our response to point 18.

13. The authors mention handsearches at various points and it would be helpful to clarify what these involve. Typically handsearching involves the selection of a number of journals and the assessment of relevance of papers, page by page. If handsearching was undertaken it is usual to report which journals, how they were selected, and how the handsearch was accomplished as well as the results? It may be that by ‘handsearching’ the authors are referring to checking the references at the end of relevant papers. If this is the case then ‘handsearching’ should not be referred to, especially not in the abstract where it may mislead searchers looking for studies where handsearching has been used.

Authors: we used handsearch as this was the term used by the other peer reviewer – Mark Petticrew – when he asked us to provide more information about this in the paper. We do not believe that MP was specifically or solely referring to page by page, volume by volume, journal reading. We have now revised the text to say manual search rather than handsearch – and we explain what we mean by this. Note that we assume that ‘manual searching’ is only an approximate term and does not need to be limited to searches of paper sources – so for example an article’s references can be checked from either a printed journal or an online copy.

“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 two 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.” (page 11)

14. The authors mention “robust studies”: it would be helpful to know how robustness was defined or measured.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

Authors: the review protocol (supplemental document 1) describes our quality appraisal criteria.

15. The authors state that “This may appear intuitive, but we contend that the finding is actually surprising.” I think the authors are struggling with this claim. What they have is a useful empirical example of what has been known and acted on for a long time in SR searching but not formally explored. There are few concrete published examples of a parallel exploration of different search approaches, so my suggestion is that the authors focus on this aspect rather than claiming novelty for a widely used technique.

Authors. We share the reviewer’s belief that this issue has not been formally explored until now. We assume that some readers may not be surprised by the findings, but we also think that some readers will be surprised. Some of the authors of this paper were surprised (as the paper states), and presumably authors of other reviews that have only included generic or specific (but not both) search terms for health outcomes would be surprised to learn how this could be problematic. The journal’s intended audience is diverse and not limited to specialist information scientists. People who are less knowledgeable about information science (but still have an interest in reviewing evidence) might well assume that searching for generic health terms could sufficiently cover all health issues of interest and therefore render specific health terms redundant –we want to make clear to such people that our study’s findings run counter to that assumption.

16. The search strategy is a major weakness of this paper as it is not sensitive, lacking in the use of truncation, synonyms and related terms. However, as a worked example, with a candid statement about its weaknesses, this is still useful.

Authors. The introduction now makes reference to this study’s status as a worked example. The limitation section now says: “The main limitations of this study are that it is based on a single review and the search was not sensitive (i.e. lacking in the use of truncation, synonyms and related terms).” (page 21)

The limitations section says:
“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.” (page 21 - 22)

17. The authors suggest a ‘more extensive search’ would have been ideal. I am not sure whether this refers to searching more resources, but I suggest that before moving on to more resources, a more sensitive (extensive) search strategy within the databases they did search would be the first priority.

Authors: The passages quoted above (point 16) makes clear that a more extensive search could include both more databases and a more sensitive search strategy (e.g. synonyms, subject headings and words in title and abstract for every concept, etc).

18. There is no reporting guidance for papers of this type, but search strategies can be assessed using the PRESS checklist: http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402

Authors: We note that there is no reporting guidance that perfectly fits a paper of this type. Strictly speaking, it is a kind of observational study so we have completed the STROBE checklist.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.

We think a post-hoc application of the PRESS checklist to a completed review at the minor corrections stage of submission is of limited usefulness – the checklist would be more useful when planning and revising searches. We do not think this article is the place to debate the applicability of the PRESS checklist to different types of review but we do wonder if all the same standards apply to broad focused reviews, as to narrow focused reviews . Our searches did identify over 10000 hits – and given the broad research question, we think it reasonable to assume that increasing the sensitivity of every concept within the entire search could easily have expanded that figure exponentially.

We believe the crucial point about the search that the reviewer wants to get across is that it was not sensitive. We have made this point clear and provided some examples of how it could have been more sensitive. We emphasise that the review was time limited (18 months) and that the search strategy reflects this time limit (see page 11 and 21). We reiterate that these kinds of constraints are common for many systematic reviews. We have made clear that more sensitive specific and generic searches may potentially have identified more studies in common. We have also made clear our recommendation that more research is needed to test whether our findings are generalisable to other reviews “and different types of literature search, including more sensitive searches.” With this in mind, we are now keen to ensure the article stays focused on the issues most central to its key messages, rather than devote further space to more detailed general description and discussion (with examples) of all aspects of the search. We are confident that the current draft now details the search and its limitations in sufficient detail for the reader to understand how search methods (and in particular, limits to search sensitivity) may potentially influence our main findings and affect generalisability.

VERSION 3 - REVIEW

REVIEWER	Julie Glanville Project Director - Information Services, York Health Economics Consortium University of York York UK
REVIEW RETURNED	21/05/2012

The reviewer completed the checklist but made no further comments.

BMJ Open: first published as 10.1136/bmjopen-2012-001043 on 25 June 2012. Downloaded from http://bmjopen.bmj.com/ on December 8, 2023 by guest. Protected by copyright.