‘it has no meaning to me’. How do researchers understand the effectiveness of literature searches? A qualitative analysis and preliminary typology of understandings.

- Mr Chris Cooper: Department of Clinical, Educational and Health Psychology, University College London, London, UK
- Professor Ruth Garside: European Centre for Environment and Human Health, University of Exeter Medical School, Truro, UK
- Dr Joanna Varley-Campbell: Department of Clinical, Educational and Health Psychology, University College London, London, UK
- Mr Juan Talens-Bou: Independent researcher
- Dr Andrew Booth: University of Sheffield, School of Health & Related Research
- Professor Nicky Britten: University of Exeter Medical School, Institute of Health Research

Abstract: This study aimed to address the question: what does 'effectiveness' mean to researchers in the context of literature searching for systematic reviews?

We conducted a thematic analysis of responses to an e-mail survey. Eighty-nine study authors, whose studies met inclusion in a recent review (2018), were contacted via e-mail and asked three questions; one directly asking the question: in literature searching, what does effective (or effectiveness in) literature searching mean to you?

Thirty-eight (46%) responses were received from diverse professional groups, including: literature searchers, systematic reviewers, clinicians and researchers. A shared understanding of what effectiveness means was not identified. Instead, five themes were developed from data: 1) effectiveness is described as a metric; 2) effectiveness is a balance between metrics; 3) effectiveness can be categorised by search purpose; 4) effectiveness is an outcome; and, 5) effectiveness is an experimental concept. We propose that these themes constitute a preliminary typology of understandings.

No single definition of effectiveness was identified. The proposed typology suggests that different researchers have differing understandings of effectiveness. This could lead to uncertainty as to the aim and the purpose of literature searches and confusion about the outcomes. The typology offers a potential route for further exploration.

Background

The aim of literature searching in systematic reviews is broadly understood as systematically identifying all relevant studies for review and reporting this process in a transparent way (1-4). The skill of the literature searcher is to identity the relevant studies, with as few irrelevant studies as possible, so that the literature searches meet the needs of researchers in study identification and the results of the literature search can be processed within the available resources (5-7).

The success of literature searching, in achieving the above-mentioned aim, is generally evaluated in terms of effectiveness (8). Researchers have traditionally used the methods of diagnostic test accuracy evaluation to evaluate literature searches and its terminology to describe their findings. These evaluations compare the operating characteristics of new searches or search filters (index tests) against existing searches or gold-standard data sets (the reference standard), commonly reporting evaluations in terms of sensitivity (recall) and precision (3, 8). These types of evaluation constitute the greatest investment in the development of study design literature search filters, in particular the development and validation of filters to identify studies reporting randomised controlled trials (RCT) (3, 5, 9-12).

Recently, researchers have been challenged to consider new ways to report the performance results of effectiveness evaluations in literature searching (8, 13). This is not only in response to reviews of search filter performance (5), and calls to improve reporting standards (3), but also to account for how new and emerging styles of evidence synthesis engage with different understandings of effectiveness or an effective search result (14-18). Rapid reviews, where researchers consider effectiveness constrained by limited time and resources (19-23), may embody such differential understandings of what constitutes effectiveness or an effective search result. Equally diverse understandings may also apply to qualitative evidence syntheses, reviews of theory, or public health systematic reviews (24-28). This shift in understanding may, in turn, require that new outcomes (e.g. time taken (1, 29, 30)) and different methods of evaluation (e.g. incremental cost-effective ratios (ICER), are used when analysing effectiveness (3, 28, 30-33). All of the above suggests that it is timely to reconsider what effectiveness means in literature searching since, on the basis of findings from a recent systematic review (undertaken in 2018), this remains unclear (3).

Study aim and methods

A recent systematic review identified a lack of consensus as to how effectiveness of literature searching is understood or what constitutes an effective literature search result (3). None of the studies included in the systematic review explicitly defined what effectiveness meant or what would constitute effective search performance (3).

The aim of this study is to address the following research question from the perspectives of literature searchers, systematic reviewers, clinicians and researchers. The research question is: what does effectiveness mean to researchers in the context of literature searching? The study reports primary work undertaken to address this research question.

Ethics approval

We obtained ethics approval from the University of Exeter Medical School ethics committee under reference number: 17/10/140.

Sampling and recruitment

The sampling frame for this study consisted of 119 studies included in the systematic review which directly compared the effectiveness of bibliographic database searching to a non-database search method (30). We identified 89 individual corresponding authors as potentially eligible from these 119 studies (some authors were corresponding authors on multiple studies).

The 89 corresponding authors were chosen since they had undertaken and reported published evaluation studies of literature search effectiveness. We recruited these study authors via an initial e-mail which included details of the study (see below) and the University of Exeter's standard consent form together with details of the ethics approval.

We chose e-mail as the method for author contact as it has been shown to be effective and efficient as a method of contacting study authors (30, 34, 35). We sent an initial e-mail to the corresponding author from an institutional e-mail address (@exeter.ac.uk) since this has been linked to improving the rate of author reply when compared to a Hotmail account (29, 34). Where it was not possible to identify an e-mail address for a corresponding author, or where the address was no longer valid, we searched Google for a current e-mail address. We were unable to identify a current e-mail address for six authors.

Data collection

Data for this study comprised e-mail replies from study authors whose studies met inclusion in the systematic review and who studies directly compared the effectiveness of a bibliographic database search to a non-database search method (30). These replies provided descriptive context from the authors as they reflected on their study to address our research question.

As some authors were known to author CC, with the potential to influence the style and content of replies, a graduate trainee (JTB), unknown to all participants, sent emails and collected the data.

The e-mail asked the following questions, and the authors' responses formed data for this analysis:

1. in your study, what did you categorise as effective or what was your measure of search effectiveness?
2. can you report any advantages or disadvantages you experienced in using the method in your study/studies, to evaluate literature search effectiveness?
3. in literature searching, what does effective (or effectiveness in) literature searching mean to you?

We sent a follow-up e-mail 14 days later, if no reply had been received. We allowed one month for responses, based upon findings of studies which have evaluated the effectiveness of contacting study authors (30, 35). We received no replies after this

period. A copy of the e-mail is included as supplementary material. An evaluation of the effectiveness, efficiency and value of author contact in this study is published elsewhere (30).

Analysis

We used thematic analysis to analyse e-mail responses, and followed the step-by-step guide reported by Braun and Clarke (36):

Data were anonymised responses, which we cut and pasted from the body text of each e-mail into a word document. We then analysed the data thematically.
Familiarisation: we 'repeatedly read' the data in an 'active way,' looking for patterns ('themes')'. On the second reading, we made preliminary notes to develop codes;
Generating initial codes: we produced initial codes from the data, which we drafted using pen and paper;
Searching for themes: we grouped the initial codes into themes and sub-themes;
Reviewing themes: we read and re-read the initial codes and initial themes against the data to test validity. At this stage we began to identify themes as reflecting a typology of understandings of search effectiveness.
Defining and naming themes: we named themes to reflect the five different categories in this typology and organised the data by theme.
Producing the analysis: we prepared an initial analysis. We read the first draft alongside the data to test the validity of the analysis.

Author CC developed the coding and the grouping of themes in discussion with author RG. As set out above, these themes became categories which we formed to develop a preliminary typology to describe what effectiveness might mean in literature searching (37). We describe this as a 'typology' since categories are descriptive rather than predictive (38), and they represent concepts rather than empirical data (37), in contrast with a taxonomy. This typology, whilst preliminary, offers a systematic basis for comparison between categories and for future evaluation and testing (37, 38).

Findings

Response and participants

We received thirty-eight (46%) responses from a possible 89 e-mail contacts relating to 119 studies. The length of e-mail replies varied between participants, with some replies directly addressing the questions posed with short sentences and other responses offered fuller descriptions of the author's study and opinions on the questions we asked. Participants represented different professional groups, including: literature searchers, systematic reviewers, clinicians and researchers in general. Where reported below, we assigned codes to the replies from participants as follows: L for literature searcher, S for systematic reviewer, C for clinical, or a more general category of R for researcher.

what does effectiveness mean?

The systematic review had not identified a clear understanding of what effectiveness 'means' in a literature searching context and, therefore, what constitutes an effective result for a literature search. Respondent replies offered initial insight into why this was the case:

[effectiveness] It has no meaning to me.
R11

Effective literature searching may mean several things: it may mean missing no studies; it may mean seeking the right balance between missing studies and NNR (but then you need a sort of cutpoint or algorithm); it may mean only retrieving high quality studies (but what is high quality); it may mean that the search does not lead to biased results or biased conclusions.
R 26

The contrast between the above two replies was mirrored in other respondent replies. It initially suggests that there is not a singular - or even shared - understanding of 'effectiveness' between respondents but rather a spectrum of differing understandings. These understandings were evident in more detailed respondent replies, none of which were mutually exclusive, or definitive. We identified five main themes and two sub-themes from the data (e-mail replies) (Table 2):

1. effectiveness is described as a metric;
2. effectiveness is a balance between metrics;
3. effectiveness can be categorised by search purpose;
4. effectiveness is an outcome;
    - 4.1 outcomes of effectiveness
    - 4.2 job role to determine effectiveness
5. effectiveness is an experimental concept.

We describe these themes below, together with respondent replies to support each theme. The main themes represent different understandings of what effectiveness means in literature searching. We present descriptions of each theme alongside our interpretation, including reference to wider literature where appropriate, for ease of reference.

1. Effectiveness is described as a metric

In this theme, respondent replies were characterised by the language used to report effectiveness and the way in which respondents described effectiveness. These replies often constituted the shortest responses to the questions we asked.

Respondents reported the name of the metric(s) used to calculate effectiveness, often without supporting context or definition. Alternatively, respondents used the language of diagnostic test accuracy evaluation to *describe* the effectiveness of a literature search again, often without context or definition:

[effectiveness means] Sensitivity and specificity.
R4

effectiveness was based on sensitivity, specificity, precision and accuracy
R9

> *Our measures of effectiveness were recall (sensitivity), precision and number needed to read*
> *I4*

Respondent replies (such as those above) use diagnostic terminology to describe effectiveness: a literature search filter was described as having 'good sensitivity' (S12), meaning that it is broad in scope and unlikely to miss potentially relevant studies. Conversely, a literature search was described as having 'low precision' (S12), meaning that the number of retrieved articles, eligible for inclusion in a systematic review, was low in comparison to the number of irrelevant articles identified. Reporting the name of a metric appears to address the question of what is effectiveness: effectiveness is, for example, sensitivity.

These responses locate the meaning of effectiveness somewhere within the metrics of evaluation, where terms like 'sensitivity' or 'precision' form a descriptive short-hand, and the relationship with effectiveness is considered self-explanatory. This technical terminology, paradoxically, obscures a clear understanding of what effectiveness is or means. To locate the understanding, and also the definition of effectiveness, a researcher must understand the meaning of these terms and how they apply to literature searching. In addition, and as explored in the second theme below, it is not clear what represents a good result from a metric and therefore a good outcome in the evaluation of a literature search, unless the methods of evaluations and metrics are clearly and transparently reported, or a balance between metrics is explored.

2. Effectiveness is a balance between metrics

The second theme closely resembled the first theme in the language used to report and describe effectiveness but the respondents provided greater detail on how they had determined and then interpreted these metrics:

> *[effectiveness means] getting the right balance between sensitivity and precision. The balance would depend on the type of project we are working on. For example, for systematic reviews, sensitivity is of primary importance. For rapid reviews, precision becomes more important – as the time available for screening records is reduced*
> *I6*

> *[An] [e]ffective literature search is the one with the optimal combination of recall (sensitivity) and precision (specificity). However what constitutes the optimal combination of these two will differ depending on the purpose of the search*
> *R20*

These respondents identify that the understanding of effectiveness is based on the optimal 'balance' between metrics to inform a comparison and determine the effectiveness of a literature search. They observe that it is not clear how a 'balance' is determined, indicating that it potentially differs by the type of the review under evaluation. It may also differ between researchers or research groups.

It was not clear from the replies in this theme exactly what constitutes an optimal balance, or how an optimal balance between metrics is determined or perceived. An 'optimal balance' appears to be determined by tacit knowledge, as opposed to comparison with an *a priori* threshold defined for the purpose of evaluation, or by reference to anticipated or validated gold-standard threshold categorised by review type. Respondent replies illustrated such tacit knowledge by reporting a graded approach for interpreting effectiveness against informal percentage thresholds. These thresholds were not defined *a priori* but based on the intuition of respondents about how their studies and literature searches generally corresponded to a putative norm:

> *The study relied on measures of sensitivity and specificity to determine if the search results garnered what they were expected to garner. Eleven years ago when the study was published, this was the common way of doing it. Numbers above 90% were considered good while results above 95% were considered excellent. Achieving results above these thresholds for both calculations was the ideal result*
> *R6*

> *I think a sensitivity of anything less than 90%-95% (in the specific case of the overviews filter) wouldn't have been acceptable*
> *R22*

These replies suggest that some respondents hold an informal threshold for effectiveness in their minds when evaluating the performance of retrieval in their own study. Respondents are able to locate a point between a 'good' result or an 'excellent' result, but the evidence for this being a 'common approach' amongst researchers, or an approach which is empirically derived, is unclear. It is also not clear if this graded approach has changed over time, and if the idea that findings over 90% are good and 95% are excellent, remains valid.

3. Effectiveness can be categorised by the type of review or search purpose

The idea that the type of review or purpose of the search determines an understanding of effectiveness was evident in respondent replies. This theme is partially linked to the preceding theme.

> *For Cochrane-style reviews, like intervention effectiveness, diagnostic or prognostic reviews, I am aiming not [to] miss anything so I aim for sensitivity. The reviewers understand to expect a lot of studies to screen and they don't want to miss anything too.*

> *For non-Cochrane reviews, it doesn't matter if I miss a paper. I'm trying not to overload the reviewers and they know we won't find everything. I focus on yield*
> *I2*

This reply broadly typifies other responses that link the type of review, or search purpose, to an understanding of search effectiveness. The distinction between the previous theme and this theme lies in the language used to describe 'effectiveness'. The shift is partly semantic. 'Cochrane-style reviews' is the phrase used to describe

literature searches which, in the previous theme, would have been characterised as searches with 'high sensitivity', but it is also partly descriptive too, with the association of 'Cochrane-style reviews' used as shorthand to describe literature searches which are perceived as those which do not miss studies.

Respondents appeared to locate effectiveness within a binary classification between types of review which are either comprehensive or non-comprehensive. Cochrane reviews and Cochrane-style reviews which aim to identify the effectiveness of interventions are regarded as requiring comprehensive and highly sensitive literature searches even if the precision of the literature search is low. This is because these reviews aim to give a precise estimate of intervention effect in statistical meta-analysis. Other review types may not need this comprehensiveness and so precision is given greater weight in designing searches. This binary classification is supported in the replies below and, to some extent, it begins to associate the idea of comprehensiveness and non-comprehensiveness with different 'types' of review, or the purpose of a literature search, each carrying different understandings of what constitutes an effective result:

> *I think it depends on what you are looking for, systematic literature review or 'just' for clinical purposes*
> *R12*

> *Depends on the question I have to answer for the patron. For Cochrane I have to identify all studies (within the limits of what is possible), for guidelines the requirements are less strict, and for a quick and dirty search I use as everybody, only one or two search terms*
> *I3*

> *Now, if you are doing a systematic review, you must find ALL studies, so you have time to screen thousands and thousands of abstracts. So you want high sensitivity but at the expense of low specificity*
> *R21*

These replies broadly suggest that, for systematic reviews, comprehensiveness is the marker of effectiveness associated with 'Cochrane-style reviews'; broadly, if the search misses a study, it is ineffective or less effective. In literature searches for clinical purposes, or the development of guidelines (based on these replies), effectiveness equates to successful delivery of a manageable number of relevant studies. This suggests that, whilst 'effectiveness' is the lens though which the success of the literature search is evaluated, it is the context of the literature search which defines what effectiveness means, and that this differs across types of review. This implies that a singular definition of effectiveness is inadequate for all types of literature search effectiveness.

4. Effectiveness can be described as an outcome

Some respondents focused on 'outcomes' of literature searching when seeking to determine or explain effectiveness. Whilst some outcomes relate to existing metrics (3), the language used by respondents, and the content of their responses, differs from the metrics described in theme one. Different professional groups sought different outcomes when producing and evaluating literature searches and

interpreting effectiveness. This theme is therefore split into two contributing sub-themes.

4.1 Outcomes or impacts of effectiveness

These respondents described effectiveness as not only about producing an effective result but also an efficient outcome or impact from different strategies, which indicate why “effectiveness” matters to them.

> *I don't think effectiveness is the right term. I think it's a balance between sensitivity and specificity. So efficiency would be better*
> *I7*

The idea of an optimal balance between metrics from theme 2 is again present (see above) but here the respondent indicates that 'efficiency' may better explain what effectiveness means in a literature search context. Other replies support this proposition adding weight to the idea of identifying efficiency outcomes when exploring and evaluating effectiveness:

> *I think effectiveness means reducing work load, saving cost and improving efficiency of study on the premise that the purpose of searching is achieved.*
> *R11*

> *Effective literature searching to me is minimizing the time and resources needed to get the target articles to answer the question. Ideally effective literature searching means returning the articles of interest (all of them in the case of a systematic review) without having to spend too much time filtering through all of the returned articles that are not on topic*
> *R10*

Reducing review team work load, saving costs, and minimising time, offer quantifiable outcomes to be used when comparing the effectiveness of one literature search to another. The presence of such outcomes suggests that researchers may have differing requirements when attempting to understand effectiveness, and differing needs in interpreting choice-making between one literature search approach and another. Furthermore, this implies that such needs are not captured, or go beyond, the values found in the metrics currently used to evaluate effectiveness in literature searching.

4.2 Job role to determine effectiveness

In the published literature, Brettle *et al*. suggest that different users might have differing needs of literature searches and the replies in this research offer some empirical support for this (39).

> *if you are a clinician or a researcher some measures [of literature search effectiveness] are more important: for instance comprehensiveness is very important for a researcher, while efficiency is more relevant to a clinician.*
> *C1*

This reply illustrates how comprehensiveness may be linked to researcher needs, with efficiency being linked to clinicians. The researcher aims not to miss anything

relevant, while the clinician is busy and wants to see relevant studies quickly. The association between needs from literature searching and job role was evident in other replies (R26). Replies extended this idea by associating outcomes of interest in literature search evaluation to job roles. This indicates how job role may influence an understanding of effectiveness, or an understanding of an effective literature search, although such understandings, as with job roles, undoubtedly overlap. The replies revealed two distinct groups: literature searchers and researchers/systematic reviewers.

Literature searchers

> *If you're doing a systematic review, you must find ALL studies…so you want high sensitivity but at the expense of low specificity*
> *R21*

The idea that the literature searcher should identify 'all' studies – as indicated in the reply above – was common in respondent replies. Respondents identified the challenges evident in this task, namely: that the literature searcher 'can never really know how many studies are available in the world literature' (I4) yet they are 'challenged' to create a literature search which captures 'ALL' (S12) the potentially relevant studies and as few irrelevant studies as possible (C2), so as to not be 'wasting the researchers time' (R16).

Literature searching is one of the first tasks undertaken in a systematic review with literature searches commonly undertaken at a time when an understanding of the available evidence is least clear. The 'pressure' to 'capture the right amount of everything' constitutes an area of uncertainty with respondents being unclear as to what level of 'completeness is sufficient' (I3).

> *My personal bias is in favour of sensitivity, so I am happy to tolerate low levels of precision*
> *I9*

How literature searches respond to this uncertainty is best evidenced in the reply above, where they appear to favour an approach based on sensitivity, so as to minimise the risk of missing potentially relevant studies. The risk of missing studies manifested in mistrust in the literature searches and the ability of the literature searcher to undertake the literature searches.

> *The worst is over-looking or missing something relevant. It might just [be] one thing, and not a key study, but suddenly your whole search approach is question[ed] and no one trusts it anymore*
> *S12*

These replies suggest how sensitivity and sensitivity-focused outcomes for the literature searcher link to the risk of missing a potentially relevant study. This focus could offer a diametrically opposite understanding of effectiveness between the literature searcher and researcher. The literature searcher considers an effective search as one where no studies are 'missed' whereas the researcher who screens the studies identified by the same literature search may question its effectiveness as determined by its precision. Furthermore, as the reply above suggests, a search

which misses a relevant study undermines trust in the reliability of the literature searches, with implications for the literature searcher and their confidence. The impact of 'missed' studies on synthesis is not always considered (28).

Researchers/ systematic reviewers

> *As a systematic review[er] I need to take the total yield into account because we have limited resources and need to be mindful of what is doable. No search strategy is perfect but we can't search millions of citations to make sure that we have found everything relevant*
> *R9*

Other replies from reviewers articulated similar views to the above. Words such as 'manageable' (C2) and 'feasible' (R9) were used by respondents to describe how they sought to understand effectiveness where 'time and resource constraints are important to factor in' (R22). This understanding of effectiveness is perhaps closer to the concept of efficiency.

Such an understanding focuses on the pragmatic task of processing the studies identified by a literature search. An effective search could be seen as one where the number of studies identified by the literature search can be processed within the resource limits of the research team. This understanding articulates a contrast with the idea of effectiveness based on sensitivity, as set out above: an effective literature search based on sensitivity could yield 100,000 studies to screen. Such a search can be 'effective' in the sense of not missing a study, but may not be effective in terms of processing studies within resource limits.

This understanding echoes the reply above in suggesting that researchers may focus first on the yield of a literature search and perhaps subsequently on the number of studies needed to read to identify relevant studies or data. This would co-locate effectiveness with the resources available to process the studies identified by the literature search. This may create an issue where the number of studies identified exceeds the resources available but the search strategy is considered effective.

Replies from reviewers demonstrated a shared understanding of effectiveness with literature searchers to the extent that literature searches should 'contain what I want' (R6) and identify 'all the studies that address your research question' (R17).

This theme locates literature search effectiveness in two ways: first in the outcomes measured where effectiveness approximates "efficiency". Secondly, it supports the idea, identified by Harbour *et al*, that researchers may have specific metrics of interest in evaluation (40). This issue could influence the practice of literature searching. Where different professional groups have differing understanding of effectiveness, or differing views on what an effective literature search means, potential confusion could be usefully explored at the start of the literature searching process through discussion within the review team. As identified in theme 1 and theme 3, some understandings of effectiveness are described in shorthand assuming a shared and common meaning for all members of the review team. It seems wise to clarify these meanings, clearly, at the start of a review.

5. Effectiveness might be an experimental information retrieval construct (it might bear little relation to the task of 'real-world' literature searching)

Replies identified within this theme appeared to differentiate effectiveness on the basis of evaluation studies, which some respondents considered as experimental and 'theoretical' (R32), and literature searching in 'real life' (I4). The latter was broadly interpreted to mean literature searching for the purpose of evidence synthesis and not necessarily evaluation or validation.

> *Our method was a retrospective examination, so it did not reflect the true experience of someone searching to answer a research question.*
> *R14*

This respondent reply directly acknowledges and highlights this distinction. Other respondents questioned if experimental constructs, which necessarily form part of the experimental evaluation studies, are relevant to real-life literature searching.

> *We can never know how many studies are available in the world literature to be found, so it is not very really possible to gauge effectiveness confidently.*
> *I4*

> *Assessments of acceptable levels of irrelevance [in a literature search] are subjective.*
> *I4*

> *The other main problem in my opinion is to define the statistic measures: sample size calculation probability, and so on.*
> *I3*

These replies highlight how respondents recognise that, in an experimental study, it is possible to know the number of relevant studies since this forms the basis of selecting a relevant gold standard: it is not possible to know this for certain in a 'real-life' literature search. Other parameters which help develop an understanding of effectiveness, or which are used to report effectiveness outcomes, are study/context specific and it is not clear how these generalise to literature searching in real-life.

Replies distinguish two types of effectiveness in literature searching. There is: i) empirical effectiveness, as reported in search evaluation studies and where the purpose of the study is to determine the effectiveness of the literature searches and, ii) effectiveness in real-life literature searching, where the primary pragmatic purpose is to identify studies for review, but where it is valuable to understand the effectiveness with which this has been achieved. The understanding, methods and data available, and the need for 'proof' of effectiveness, may vary between these experimental or evaluation studies and 'real life', creating a potential duality in understanding.

Discussion

This study addressed the research question 'what does effectiveness mean in literature searching from the point of view of literature searches, systematic reviewers, clinicians and researchers?'. In addressing this question, we have identified five different understandings. None of these understandings is definitive

and none are mutually exclusive. Viewed collectively, they capture a preliminary and nuanced understanding of what effectiveness might mean in literature searching, as well as the potential for misunderstanding.

We suggest that this work supports a preliminary typology from which further work to address the broad topic of effectiveness in literature searching can stem. Limitations to this study, and therefore in the typology, suggest a need for further research. We first set out these limitations, to situate a proposed agenda for further work within the limitations themselves, and within the wider literature.

Limitations

The sample of respondents contacted in this study was constrained by the scope of the underpinning systematic review (3). The systematic review explored literature search effectiveness in the context of evaluations which compared bibliographic database searching to non-database methods of searching. This specific context may have influenced findings from which we have developed the preliminary typology. Furthermore, respondents were chosen specifically from authors of published studies that have themselves explored the question of effectiveness in literature searching. We acknowledge that the sample of respondents is not likely representative of all researchers undertaking literature searching or processing literature searches, and that respondents who have evaluated effectiveness in literature searching may have formed views based on their experience which are not representative of common practice. Nonetheless, we identified diverse understandings even within this group with prior familiarity with concepts of search effectiveness Each of these points holds the potential to influence our findings and to shape development of the preliminary typology.

The use of e-mail replies to generate data for analysis may represent some limitations to the work we present. The response rate was 46%. While this may be considered high for an email survey it does still mean that over 50% of eligible authors did not contribute to our survey. The responses we received were based on a single round of written questions, so we were unable to explore respondent replies in great depth.

Finally, the position of the authors in analysing and developing the typology should be noted. The authors of this study themselves have experience in literature searching and in undertaking systematic reviews, with a track record of working on systematic reviews where the question of effective literature searching is frequently considered. Whilst this has undoubtedly helped with interpreting the replies we cannot discount its likely influence on the development of the typology.

With these limitations in mind, we reflect on a proposed research agenda based on the typology.

Developing the typology

The typology represents a preliminary framework from which further work can be undertaken to explore the topic of effectiveness in literature searching. As observed above, the sample of respondents was identified within a specific context. This could be addressed in further work. The sample of respondents could be broadened

through an on-line survey and the typology could be tested for validity, allowing for development or refinement of the typology.

The language used to describe effectiveness within the typology may benefit from further development. As identified in the associated systematic review (3), and by other researchers (13, 40), the language used to describe effectiveness in literature searching is used interchangeably and often incorrectly. In this study, we identified that effectiveness is commonly described in shorthand using the name of a specific metric (theme one) or defined by the type of review which the literature search informed (theme three). This suggests that researchers know what they understand by effectiveness, but such understandings have not hitherto been identified, codified or set out for wider examination. The typology might fulfil a useful function as a framework to generate discussion about the language used by researchers and to make explicit what exactly they mean by search effectiveness.

Processing different understandings of effectiveness

Replies revealed that different professional groups have different understandings of effectiveness leading to potentially different motivations in developing or working with literature searches (theme four). Furthermore, different professional groups may be interested in different outcomes from a literature search (theme three) and may view literature searches against different conceptual, pre-established understandings (theme three). This suggests a need for greater dialogue between professional groups before a literature search is developed to explore the purpose of the review and therefore the requirements of the literature search. Such a discussion would seek to generate a consensus on what a ‘successful’ literature search would look like for the whole research team in context, and potentially how to measure or evaluate this, if this is important to the wider work.

Our findings suggest a need for improved understanding of the work required in literature searching, and in reviewing the studies identified by the literature search, between professional groups. Dialogue between professional groups would likely facilitate a shared understanding of the literature search. Tools such as search narratives, to explain conceptual and contextual detail of the construction of search strategies (2), or checklists such as the PRESS checklist (41), could be used to include researchers in the development and review of a search strategy. More broadly, training on the methods of literature searching and including researchers in developing the literature searching, and training on screening and including literature searchers in the screening of the review, may achieve cross-over between disciplines (42, 43). This may add rich context to the experience, for example, how the addition of one ‘crucial’ but seemingly nebulous search term can add a large number of ineligible studies to screen whilst also developing an understanding of what it means to screen these studies. These ideas, already common in some contexts, are confirmed by our findings, and they may help to address concerns of team members, even within experienced research teams, as identified in connection with theme four.

Improving transparency in effectiveness evaluation: thresholds

Theme two focused on identifying an optimal balance between metrics to determine effectiveness. None of the studies in the accompanying systematic review prospectively set thresholds for comparison with search results in order to determine

an effective result. There was evidence in theme two that respondents held informal thresholds in mind to evaluate their findings and this represents one potential area for development.

Whilst the prospective setting of benchmark thresholds on a study-by-study basis would be optimal (3), generic thresholds determined by the purpose of a literature search might offer an initial, visual benchmark. For instance, for reviews of intervention effectiveness guidance suggests that high sensitivity in literature searching is to be preferred (1, 5, 44); the literature search could aim for sensitivity above 95%. Searches that fall below this threshold could require further examination.

Developing benchmarks for literature searching would not be easy but it would represent an initial marker for further work. Such aspirations should be considered in the context of recent reviews which aim to address identified shortcomings in methodology and transparency of reporting for effectiveness evaluation of literature searching (c.f. (3, 8)).

Improving measurement: outcomes

Evaluation of literature search effectiveness to date has largely focused on 'comprehensiveness' as a measure of effectiveness and of success (1, 24, 27, 28). Whilst comprehensive literature searches are considered important for reviews of intervention effectiveness (1), other types of review may have differing requirements of their literature searching. New ways to consider or to evaluate effectiveness are needed (1, 3, 13, 30, 44, 45). Further discussion on these metrics, and which outcomes to capture and compare in effectiveness evaluation, is an important topic for discussion.

The proposed typology can contribute to this discussion. Some outcomes identified by respondents in this study are, to some extent, already present in the metrics identified above and reported in the systematic review. Respondents also reported outcomes which do not yet form part of the published literature on literature search evaluation, namely: client satisfaction with the searches (C3) and meeting the expectations of researchers who rely on the searches (I7). As literature searchers continue to locate their role in systematic reviews, and the effectiveness of their contribution to evidence synthesis (6, 45-47), co-locating satisfaction-based outcomes with effectiveness-based outcomes, could prove valuable.

Considering outcomes in addition to - or in place of - current approaches to effectiveness evaluation is reflected in the academic literature on evaluation of literature search filters. Researchers have called for new ways to report evaluation of literature search effectiveness (13). Combining metrics AND outcomes, or enhancing methods from metric AND outcome data, for example, using incremental cost effectiveness ratios (ICERs) offer alternative ways of meeting this call (for example, Schimilt et al who have explored this idea when screening for systematic reviews (32)). Literature searching is rarely evaluated against time or cost (29) and yet researchers and literature searches continually look to improve effectiveness and efficiency in rapid reviews, as well as to understand the potential risks carried by the different search approaches and methods (18, 48). Quantifying time and/or cost could also help researchers to more clearly enumerate the resources needed to identify studies.

Effectiveness in real life literature searching

Respondents flagged a distinction between their own evaluation studies and understanding the effectiveness of real-life literature searches (theme five). This may lead to further discussion on how effectiveness is evaluated and the extent to which the methods and techniques of effectiveness evaluation are fit for evaluation of effectiveness in day-to-day literature searching. This represents a key question for researchers in information science or retrieval disciplines. The potential role for demonstrating the effectiveness of literature searches when reporting literature searches in systematic reviews can be linked to the need to improve confidence in the findings of the systematic review. It may also lead researchers to consider how searches are evaluated and whether the language of diagnostic techniques is appropriate for 'real life' literature searches, given the issues raised in this theme. It may be that diagnostic-based approaches are only appropriate in selective cases (8).

Conclusions

We identified no consensual definition of effectiveness in literature searching. Instead, we identified five themes which explore the idea of what effectiveness means in literature searching. We propose that these five themes form a preliminary typology of how to interpret effectiveness in literature searching.

The typology illustrates that it is not sufficient to use common terms to describe effectiveness, or even that researcher teams do not have shared understandings of search effectiveness. Definitions and expectations differ, from those who assume that the terminology is "self-explanatory" through to those who problematise or reject "effectiveness" as an appropriate way of thinking about the usefulness and appropriateness of searching within the context of a review project. These issues would benefit from further consideration informed by our proposed typology.

Acknowledgments

We would like to thank the respondents for taking the time to reply to our questions

Funding

Data collection was undertaken as part of the systematic review of effectiveness reported elsewhere (3). That systematic review formed part of a project funded by the PenTAG health technology assessment grant (PI Chris Hyde: Project Number 16/54/11) and data collection was undertaken in 2016 when Chris Cooper, Joanna Campbell and Juan Talens-Bou were employed by PenTAG.

The analysis reported here, and the writing up of this study, was undertaken in 2018/19 without any specific funding for Chris Cooper, Joanna Campbell and Juan Talens-Bou.

Professor Garside and Professor Britten were partially supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) South West Peninsula, now recommissioned as NIHR Applied Research Collaboration (ARC) South West Peninsula. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Dr Booth contributed under his University of Sheffield contract.

Ethics

Ethics approval was gained from the University of Exeter Medical School. Approval Number: 17/10/140.

Author contributions

Mr Chris Cooper
Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Supervision; Writing - original draft.

Prof. Ruth Garside
Supervision; Data curation; Formal analysis; Writing - review & editing.

Dr Jo Varley-Campbell
Data curation; Writing - review & editing.

Mr Juan Talens-Bou
Data curation; Writing - review & editing.

Dr Andrew Booth
Supervision; Writing - review & editing.

Prof. Nicky Britten
Supervision; Writing - review & editing.

Highlights

What is already known:
in a recent systematic review, it was not possible to determine what effectiveness meant in literature searching. This could lead to uncertainty as to the aim and the purpose of literature searches and confusion about the outcomes. The preliminary typology presented here offers a potential route for further exploration.

What is new: a preliminary typology to describe what effectiveness might mean in literature searching is presented. Five understandings were developed from data:

1) effectiveness is described as a metric;
2) effectiveness is a balance between metrics;
3) effectiveness can be categorised by search purpose;
4) effectiveness is an outcome; and,
5) effectiveness is an experimental concept.

Potential impact for Review Synthesis Methods readers outside the authors' field:
This work forms a preliminary typology from which further work to address the broad topic of effectiveness in literature searching can stem. It may help research teams to reflect on their diverse expectations of literature reviews in specific reviews and syntheses.

There are numerous limitations in the work. The preliminary typology presented here would benefit from further analysis by the research community. The sample of respondents could be broadened and the typology could be further developed or refined.

The proposed typology suggests that different researcher groups have differing understandings of what effectiveness means. Discussion between researchers, and research groups, may prove valuable to developing shared and common understandings of effectiveness.

The preliminary typology may serve as a framework for these discussions.

Data Availability Statement

The ethics application for this study (University of Exeter Medical School reference number: 17/10/140) only covered access to data by the study authors. The data is not publicly available beyond where anonymised sections are reported in the study.

References

1. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Medical Research Methodology. 2018;18(1):85.
2. Cooper C, Dawson S, Peters J, Varley-Campbell J, Cockcroft E, Hendon J, et al. Revisiting the need for a literature search narrative: A brief methodological note. Research synthesis methods. 2018;9(3):361-5.
3. Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. J Clin Epidemiol. 2018;99:53-63.
4. Rader T, Mann M, Stansfield C, Cooper C, Sampson M. Methods for documenting systematic review searches: a discussion of common issues. Research synthesis methods. 2014;5(2):98-115.
5. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. 2011 [cited Accessed 7th December 2017]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. The Cochrane Collaboration, [cited Accessed 7th December 2017]. Available from: http://handbook.cochrane.org/.
6. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. Journal of clinical epidemiology. 2015;68(6):617-26.
7. McGowan J, Sampson M. Systematic reviews need systematic searchers. Journal of the Medical Library Association. 2005;93(1):74-80.
8. Lefebvre C, Glanville J, Beale S, Boachie C, Duffy S, Fraser C, et al. Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study. Health technology assessment (Winchester, England). 2017;21(69):1-148.
9. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. Journal of the American Medical Informatics Association. 1994;1(6):447-58.
10. McDonald SJ, Lefebvre C, Clarke MJ. Identifying reports of controlled trials in the BMJ and the Lancet. BMJ: British Medical Journal. 1996;313(7065):1116-7.
11. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. International Journal of Epidemiology. 2002;31(1):150-3.
12. Watson RJ, Richardson PH. Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, Medline and PsycINFO bibliographic databases. The British journal of medical psychology. 1999;72 (Pt 4):535-42.
13. Jenkins M. Evaluation of methodological search filters--a review. Health information and libraries journal. 2004;21(3):148-63.
14. Anderson LM, Oliver SR, Michie S, Rehfuess E, Noyes J, Shemilt I. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. Journal of clinical epidemiology. 2013;66(11):1223-9.
15. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. Systematic Reviews. 2012;1(1):28.

16. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC medical research methodology. 2018;18(1):5.
17. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, et al. A research and development agenda for systematic reviews that ask complex questions about complex interventions. Journal of clinical epidemiology. 2013;66(11):1262-70.
18. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Systematic Reviews. 2015;4(1):36.
19. Kaltenthaler E, Cooper K, Pandor A, Martyn-St James M, Chatters R, Wong R. The use of rapid review methods in health technology assessments: 3 case studies. BMC medical research methodology. 2016;16(1):108.
20. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Systematic Reviews. 2012;1(1):10.
21. Nussbaumer-Streit B, Klerings I, Wagner G, Titscher V, Gartlehner G. Assessing the validity of abbreviated literature searches for rapid reviews: protocol of a non-inferiority and meta-epidemiologic study. Systematic Reviews. 2016;5:197.
22. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC medicine. 2015;13:224.
23. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al. Rapid reviews versus full systematic reviews: An inventory of current methods and practice in health technology assessment. International journal of technology assessment in health care. 2008;24(2):133-9.
24. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. International journal of technology assessment in health care. 2010;26(4):431-5.
25. Booth A. The number needed to retrieve: a practically useful measure of information retrieval? Health information and libraries journal. 2006;23(3):229-32.
26. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Systematic Reviews. 2016;5:74.
27. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health technology assessment. 2003;7(1):1-76.
28. Cooper C, Lovell R, Husk K, Booth A, Garside R. Supplementary search methods were more effective and offered better value than bibliographic database searching: A case study from public health and environmental enhancement. Research Synthesis Methods. 2018;9(2):195-223.
29. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. Systematic Reviews. 2017;6(1):234.
30. Cooper C, Bou JT, Varley-Campbell J. Evaluating the effectiveness, efficiency, cost and value of contacting study authors in a systematic review: a case study and worked example. BMC Medical Research Methodology. 2019;19(1):45.
31. Cooper C, Varley-Campbell J, Carter P. Established search filters may miss studies when identifying randomised controlled trials. Journal of Clinical Epidemiology. 2019;112(August):12-9.

32. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Research Synthesis Methods. 2014;5(1):31-49.
33. Cooper C, Snowsill T, Worsley C, Prowse A, O'Mara-Eves A, Greenwood H, et al. Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study. Journal of Clinical Epidemiology. 2020;123:39-48.
34. O'Leary F. Is email a reliable means of contacting authors of previously published papers? A study of the Emergency Medicine Journal for 2001. Emergency Medicine Journal : EMJ. 2003;20(4):352-3.
35. Gibson CA, Bailey BW, Carper MJ, LeCheminant JD, Kirk EP, Huang G, et al. Author contacts for retrieval of data for a meta-analysis on exercise and diet restriction. International Journal of Technology Assessment in Health Care. 2006;22(2):267-70.
36. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
37. Smith KB. Typologies, Taxonomies, and the Benefits of Policy Classification. Policy Studies Journal. 2002;30(3):379-95.
38. Bailey K. Typologies and taxonomies (paper): an introductionto classification techniques: An Introduction to Classification Techniques (Quantitative Applications in the Social Sciences). New York: Sage; 1994.
39. Brettle AJ, Long AF, Grant MJ, Greenhalgh J. Searching for information on outcomes: do you need to be comprehensive? Quality in health care : QHC. 1998;7(3):163-7.
40. Harbour J, Fraser C, Lefebvre C, Glanville J, Beale S, Boachie C, et al. Reporting methodological search filter performance comparisons: a literature review. Health Information & Libraries Journal. 2014;31(3):176-94.
41. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of Clinical Epidemiology. 2016;75:40-6.
42. Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. Research Synthesis Methods.0(0).
43. Polanin JR, Pigott TD, Espelage DL, Grotpeter JK. Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. Research Synthesis Methods.0(0).
44. Centre for Reviews and Dissemination. Systematic reviews – CRD's guidance for undertaking reviews in healthcare. York: Centre for Reviews and Dissemination, University of York; 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
45. Beverley CA, Booth A, Bath PA. The role of the information specialist in the systematic review process: a health information case study. Health information and libraries journal. 2003;20(2):65-74.
46. Harris MR. The librarian's roles in the systematic review process: a case study. Journal of the Medical Library Association. 2005;93(1):81-7.
47. Koffel JB. Use of Recommended Search Strategies in Systematic Reviews and the Impact of Librarian Involvement: A Cross-Sectional Survey of Recent Authors. PLoS ONE. 2015;10(5):e0125931.
48. Pham MT, Waddell L, Rajić A, Sargeant JM, Papadopoulos A, McEwen SA. Implications of applying methodological shortcuts to expedite systematic reviews: three case studies using systematic reviews from agri-food public health. Research Synthesis Methods. 2016;7(4):433-46.

Table

Table 1: description of themes and the preliminary typology

Themes	Sub-themes (if any)	Description
1) effectiveness is described as a metric	n/a	Respondents reported the name of the metric(s) used to calculate effectiveness, often without supporting context or definition.
2) effectiveness is a balance between metrics	n/a	Respondents identify that the understanding of effectiveness is based on the optimal 'balance' between metrics to inform a comparison and determine the effectiveness of a literature search.
3) effectiveness can be categorised by search purpose	n/a	Respondents appeared to locate effectiveness within a binary classification between types of review which are either comprehensive or non-comprehensive.
4) effectiveness is an outcome	4.1 outcomes of effectiveness 4.2 job role determines an understanding of effectiveness	Respondents focused on 'outcomes' of literature searching as a mechanism to determine or explain effectiveness. This theme was subdivided: 4.1 respondents described effectiveness as not only about producing an effective result but also an efficient outcome or impact; 4.2 Researchers may have specific metrics of interest in evaluation (40). Job role may influence an understanding of effectiveness or what represents and effective result.
5) effectiveness is an experimental concept	n/a	Effectiveness was considered to bear little relation to the tasks of real-world literature searching