Testing filter term performance in PsycINFO to identify evidence syntheses in crime reduction, using the relative recall method

Lisa Tompson[1] (iD)

Abstract

Objectives To test filter term performance against an original search strategy to identify evidence syntheses with a crime reduction outcome in the PsycINFO database, with a view to maximising efficiency and/or effectiveness in the search phase of a systematic review.

Methods A search strategy was developed to identify evidence syntheses with crime reduction outcomes. A 'quasi-gold standard' set of 255 relevant studies that were indexed in the PsycINFO database was derived from this initial work and was used to test various filter terms available in the database using the relative recall method. Precision and sensitivity statistics were generated for each search strategy.

Results Seven search strategies were tested using three clusters of index terms, on (1) method filter terms, (2) topic filter terms and (3) method *and* topic filter terms. These were applied as filters for the original search strategy and, to facilitate comparison, against all records in PsycINFO. The most sensitive filter scored 74.1%, the most precise scored 44.1% and the best compromise between sensitivity and precision scored 53.7% sensitivity and 16.3% precision.

Conclusions Filter term performance in PsycINFO can be used to inform search strategies used within criminology and allied fields for systematic reviews. The variety of filter terms tested here, in the absence and presence of a keyword search, caters for researchers with different information requirements. Using an evidence-based approach to systematic searching can yield considerable resource savings in conducting a systematic review.

Keywords Search filter · Search hedge · Search strategy · Systematic search · Systematic review

✉ Lisa Tompson
  l.tompson@ucl.ac.uk

Extended author information available on the last page of the article

🖄 Springer

Introduction

Systematic reviews are seen as critical tools for evidence-based practice as they synthesise evidence on a topic to provide a contemporary portrait (Gough et al. 2013). The development of a search strategy is a crucial and time-consuming component of a systematic review. The results of a search strategy are analogous to the data collection phase of a primary study, since they provide a collection of citations from which eligible studies are identified for later synthesis. The minimisation of bias at this stage is central to the reliability of the results produced in a systematic review (Wilson 2009).

In keeping with the scientific philosophy to which they contribute, systematic reviews ought to use evidence-based methods (Thompson et al. 2014). 'Search filters', often called hedges[1] in the early literature, are search strategies that are deployed in electronic bibliographic databases to retrieve particular types of records (Sampson et al. 2006), typically those with a specific research design, such as randomised control trials, or focus, such as a medical concept, population or topic. Filters are usually comprised of tested combinations of electronic database search terms. So many now exist that organisations such as The InterTASC Information Specialists' Sub-Group collate filters on their website,[2] along with information on how they have been found to perform under empirical scrutiny.

Electronic bibliographic databases vary in the sophistication of their search functions. At the simpler end of the spectrum, they depend solely on keyword searches, using what is known as 'natural language'. At the other end are vast databases containing many millions of citation records, where records are indexed; that is, they are manually assigned 'controlled vocabulary' to classify them further (see Tompson and Belur, 2016). Such index terms range from methods descriptors, to population details (e.g., male, female, animal, human), to topic descriptors. These larger databases sometimes have in-built limits for publication year, language and methods (White et al. 2001).

Many electronic reference databases index criminological journal articles. Some of these have a bespoke topic orientation (e.g., Criminal Justice periodicals), others have a broad social science focus (e.g., Sociological Abstracts), and others still are vast multi-disciplinary databases such as SCOPUS or Web of Knowledge. Partial lists of useful databases for criminology literature are provided by Kugley et al. (2016) and Tompson and Belur (2016). Each database has a unique way of indexing, with a great variety of terms and frequencies of terms seen when surveying them 'en masse'. Complicating matters, each database may be hosted by different vendors, and the index terms may similarly vary across these vendors (e.g. PsycINFO is hosted by Ovid and EBSCO, to name two) (Kugley et al. 2016: 22). Some topic index terms are exclusive to specific databases, whereas methods index terms are more consistent across different databases hosted by the same vendor. Commensurately, the performance of different index terms (for example, those searching for 'randomised control trials') is reported to vary across different databases too (see InterTASC website[2]).

---

[1] They are also known as optimal search strategies, optimal search filters, quality filters, search filters or clinical queries (Jenkins 2004).

[2] https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/

Search filters are database-specific and generated through a combination of index terms that have been empirically shown to identify relevant records for a given research question. Moreover, it is also possible to use the Boolean term 'NOT' to filter the results of a database search; for example, the author has previously used it to remove studies containing 'heart arrest' when 'arrest' has been one of the keywords. However, when advising Campbell Collaboration systematic reviewers, Kugley et al. (2016) caution that the NOT term can lead to unstable results, so it should be applied with caution. Kugley et al. (2016: 29) further note that the thoroughness of indexing in social science databases is not comparable with those in the health sciences. Methods are often not mentioned in social science abstracts—which is where index terms are often drawn from—and therefore search filters seeking to identify particular methodological studies may miss many potentially relevant studies. This echoes others who have observed the lack of common vocabulary across the social sciences and lack of structured abstracts which are so common in medical journals (Glanville et al. 2008). It logically follows that literature that has an inconsistent form of expression will be more difficult for (human, often post-graduate student) indexers to consistently categorise. To date, however, there are no studies that the author is aware of that tests these assertions.

Systematic reviews are intensive endeavours. The overall resource cost is highly influenced by the number of citation records retrieved via the search strategy (Sampson et al. 2006). The core aims of a search filter are to improve the efficiency and/or effectiveness of searching (Glanville et al. 2008). Different filters will err towards one or the other. Some favour efficiency, which relates to maximising the *precision* of the search, thereby maximising the number of relevant citation records identified whilst minimising the number of irrelevant citation records identified. Precision-maximising searches tend to save time in resources insofar that they reduce the time spent in screening citation records against a set of inclusion criteria, sometimes up to a week in a large review (Cohen et al. 2006). On the other hand, some filters favour effectiveness, which relates to maximising the identification of the number of relevant citations from all that exist in a database. This latter, sensitivity-maximising, search type is typical of systematic reviews that seek to perform a comprehensive search for as many relevant studies as possible, which is often the aspiration in systematic reviews looking to conduct a quantitative synthesis of data. Since precision and sensitivity are incompatible aims, some filters prioritise one at the expense of the other, whilst other filters seek to find an acceptable balance between the two.

Used extensively in the health sciences, many search filters have been designed for popular databases such as MEDLINE or PubMed and rigorously tested over time (Glanville et al. 2008). Use in the social sciences, and criminology in particular, is less common, albeit given the increasing frequency of systematic reviews produced in the fields of criminology, criminal justice and crime reduction (Bowers et al. 2014a; Pratt 2014) this can be anticipated to grow.

This study reports a modest empirical effort towards testing the effectiveness and efficiency of several search filter terms in the PsycINFO database, with the objective of identifying evidence syntheses with a measured crime reduction outcome. PsycINFO was chosen for this study because it has sophisticated functionality that facilitated the tests applied and was found by (Tompson and Belur 2016) to contain a high proportion of 'unique' records that met the inclusion criteria (i.e., that were not found in other

databases using a similar keyword search string). As Kugley et al. (2016) have stressed, filter terms used for searches in the social sciences may not be as universally applied as in other disciplines. For this reason, the productivity of the filter terms tested is benchmarked against a broader search strategy that combines natural language and controlled vocabulary to retrieve evidence syntheses with a crime reduction outcome.

It is somewhat unusual to present search filters on methods and topic in the same paper; they are customarily reported separately for different audiences (for an exception see Eady et al. 2008). Given the current attention focused on amassing the evidence base in crime reduction,[3] often with the aim of integration into evidence 'hubs', it was assumed that combining both in this paper would be illuminating for the many teams of international researchers working towards this goal.

The intention is that the evidence generated in this paper will inform researchers looking to use the PsycINFO database in systematic reviews in the fields of crime prevention, criminal justice and allied fields and, hopefully, save them time and effort. This paper proceeds as follows: first, I will outline the process originally used to generate a search strategy to identify evidence syntheses in crime reduction. Next, I will describe how a reference data set was constructed and how this was used to test different search filters in the PsycINFO database. The results will then be discussed in the context of their implications for researchers undertaking different styles of systematic review in this field.

Methods

According to Glanville et al. (2008), the development of a search filter starts with the identification of a gold standard reference data set (test data) then moves on to search term selection, strategy development, testing and validation, before being compared to the performance of other filters. Since this study reports on already available filter index terms in the Ovid interface of the PsycINFO electronic bibliographic database, this section deviates slightly from this lifecycle.

The quasi-gold standard (reference) data set

This work was undertaken as a component of the research programme that underpinned the UK What Works Centre for Crime Reduction. The first task of this was to systematically assemble existing evidence syntheses with a crime reduction outcome (for the protocol that documents this process see Bowers et al. 2014b). The search strategy used to identify evidence syntheses is detailed in (Tompson and Belur 2016) but is worth briefly recounting here. Synonyms were generated for three core terms: 'crime', 'reduction' and 'evidence syntheses'. This process involved checking lists of crime types from several countries, exploring the thesauri in bibliographic databases and consulting with topic and information scientist experts. The search terms resulting from this stage were individually tested for precision and sensitivity—what White et al.

---

[3] Both academic attention—see Weisburd et al.'s (2017) 'review of reviews,' and policy attention, —see Washington State Institute for Public Policy, Crime Solutions®, Evidence Mapping To Advance Justice Practice (EMTAP) and De Waard (2018) for examples.

([2001](#)) call 'univariate analysis' or could be considered as a stepwise procedure—against a known list of crime reduction evidence syntheses harvested from previous research (Petrosino [1997](#); Weisburd et al. [2016](#); Wells [2009](#)). Terms that, in combination, correctly identified many of the known records and that did not excessively inflate the number of results in the search were retained in the search string of natural language. These were then combined with the bespoke-controlled vocabulary (index terms) available in each of the 14 electronic bibliographic databases. Other features of the search strategy were forward and backward citation searching of candidate studies and consulting with experts for absent studies in the final sample. For the search period of January 1975–October 2017, 383 studies met the inclusion criteria which were that (a) they should be an evidence synthesis (systematic review and/or meta-analysis) and that (b) they should have a measurable crime reduction outcome.

These 383 studies are the starting point for the reference data set in the study that follows. Rather than referring to these studies as the 'gold standard', the common name for reference data in search filter testing, I prefer to use the term 'quasi-gold standard' (Jenkins [2004](#); White et al. [2001](#)), in recognition that it is impossible to identify the true number of studies meeting the inclusion criteria. However, since the search tactics employed in the original project were comprehensive, it is unlikely that there are large numbers of relevant studies missing in the reference data set.

Test search strategy

A relative recall[4] test strategy was employed to test the effectiveness and efficiency of the filter terms. Introduced by Sampson et al. ([2006](#)), relative recall is the proportion of studies (records), retrieved by a search, of a quasi-gold standard that has been generated through composite search tactics. Since the quasi-gold standard in this paper was generated from a comprehensive search strategy, involving many search tactics (e.g., multiple database searches, forward and backward citation searches, consultation with topic experts), this fulfils the brief of relative recall.

The 383 studies of evidence syntheses of crime reduction topics were first searched for in the Ovid interface of the PsycINFO database in March 2018.[5] Two hundred and fifty-five of these studies were discoverable—meaning that they had been indexed in the database by this date—and this corpus of studies formed the quasi-gold standard. The remaining 128 studies were largely published in academic journals not indexed by the database,[6] with around 30% coming from the grey literature (such as Campbell Collaboration reviews), which is generally less visible in electronic bibliographic databases. Furthermore, some recent studies may have not yet been indexed in PsycINFO, for there is often a delay between the publication date and the indexing date.

---

[4] This is different from absolute recall, which is when the gold standard is a true gold standard in that it represents every relevant study that exists in the database. This is notoriously difficult to determine unless one extensively hand searches to produce the gold standard (see Sampson et al. [2006](#)).

[5] It is important to stress here that the 383 studies had been screened and coded by October 2017, whereas the searches undertaken in this paper were performed in March 2018.

[6] For example, Journals not indexed in PsycINFO include: Criminology & Public Policy, Criminal Justice Policy Review, Justice Quarterly, Epidemiologic Reviews, PLoS ONE, BMC Public Health, Emergency Medicine Australasia, Health Technology Assessment, Annals of Medicine, and Health and Justice.

A sensitive approach was used to generate the filter terms. Each search field in PsycINFO was reviewed for index terms related to crime reduction (e.g., crime prevention, criminal rehabilitation and penology) and evidence synthesis (e.g., meta-analysis and systematic review). Five fields were found to contain this information: the methods (md), classification code (cc), classification word (cw), key concepts (id) and subject heading (sh). The term 'crime prevention' was also mapped to the subject headings in the advanced search window, which found related terms in the database thesauri and allowed the use of the 'explode' function to find more general index terms (for example, in this search 'crime' was a parent term in the hierarchy used by the database thesauri). The resulting index terms were used in searches to test their effectiveness and efficiency for identifying evidence syntheses on crime reduction topics.

The search string that was originally used to generate the quasi-gold standard (see Tompson and Belur 2016), which used natural language keywords combined with Boolean operators, was run first. The database filter index terms were then applied in different permutations to the first search, before being cross-referenced with the quasi-gold standard. For comparison, I also conducted searches solely with the database index terms (controlled vocabulary), in the absence of the original search string, to assess how well the filter terms retrieved records from PsycINFO on their own. These results were then cross-referenced against the quasi-gold standard. Sensitivity and precision statistics were calculated for each search run.

Sensitivity and precision

Whilst there are a number of metrics used to evaluate search filters (see Jenkins 2004), two of the most commonly calculated are sensitivity and precision (also known as recall). As mentioned previously, sensitivity relates to the number of relevant studies correctly retrieved in a search. Thus, the numerator to calculate sensitivity is the number of relevant studies retrieved by a search strategy, and the denominator is all of the relevant studies discoverable in the database (including those not retrieved). To maximise sensitivity, it is usually assumed that there will be many irrelevant studies identified by the search strategy, as effectiveness at identifying relevant studies is more important here than efficiency.

In contrast, precision relates to the efficiency or 'hit rate' of a search strategy. This is calculated with a numerator of the number of relevant studies retrieved in a search strategy, with the denominator being all of the studies (both relevant and irrelevant) retrieved by a search strategy. The formula for calculating sensitivity and precision is illustrated in Fig. 1.

Results

Seven search strategies were executed, as presented in Table 1. This begins with S_1, which was the original search syntax (reported in the Appendix) run on 8 March 2018, and which retrieved 4400 records. Searches 2–7 (S_2–S_7) included different permutations of the index terms harvested in PsycINFO—categorised into methods terms, topic terms and methods *and* topic terms. S_2–S_4 see these index terms applied as a 'filter' to

	Relevant studies indexed in PsycINFO	Irrelevant studies indexed in PsycINFO	
Studies retrieved by a search strategy	a	b	a+b
Studies not retrieved by a search strategy	c	d	

a+c

Sensitivity = a/(a+c)*100

Precision = a/(a+b)*100

Fig. 1 Formula for calculating sensitivity and precision

the original S_1 results. S_5–S_7 see these index terms used independently as a search syntax, to assess whether they are superior in identifying evidence syntheses with a crime reduction outcome when compared to the original keyword search (S_1).

When the seven search strategies were cross-referenced with the quasi-gold standard (Q-GS), they produced the data reported in Table 2. This shows that using the original search syntax alone (S_1) retrieved 171 of the possible 255 studies in the quasi-gold standard, which equates to a sensitivity of 67.1%[7] of the quasi-gold standard being identified. The precision for S_1 was 3.9, meaning that the 'hit rate' for relevant studies was around 4%. These statistics are reasonable given that the quasi-gold standard was identified over a dozen database searches, forward and backward citation analysis and with expert input (see Tompson and Belur 2016).

Using S_1 as a baseline, we can see from Table 2 that applying filter index terms for method in S_2 greatly increases the precision of the search, since the number of records retrieved in the search reduces by around four-fifths. This was at the expense of a little sensitivity though, with 34 fewer studies being identified in this search. S_3, which applied filter index terms for topic, resulted in better precision than S_1, but saw a marked drop in sensitivity, with 52 fewer studies from the quasi-gold standard being identified. S_4 combined both of these filter index terms and resulted in the best precision for all the search strategies at 44.1%. The penalty to the sensitivity of the search was marked though, with only 94 of a possible 171 studies being identified.

Moving on to the second set of search strategies—those acting as a filter for the whole PsycINFO database rather than the original search syntax—we see that S_5 produces the greatest sensitivity of all the searches—with 189 out of the 255 quasi-gold standard studies identified. It therefore identifies a greater volume of relevant studies than the original search (S_1). This maximised sensitivity is offset by the incredibly small precision rate though, as over 35,000 citation records would need to be screened to identify the relevant studies. S_6 offers no improvement for precision or sensitivity in comparison with S_5. S_7 results in the best precision for the filter terms alone, although the sensitivity equates to just less than half the sample of the quasi-gold standard being identified.

---

[7] All calculations of sensitivity use the full quasi-gold standard sample, so that comparisons can be made across the search strategies with and without the original search syntax.

Table 1 Search strategy abbreviations, syntax and description

Abbreviation	PsycINFO database syntax	Description
S_1	SEARCH SYNTAX	The original search syntax run (documented in the Appendix).
S_2	SEARCH SYNTAX AND ((meta analysis or metasynthesis or "systematic review").md. OR meta analysis).sh.)	Using the Boolean operator AND to filter for method field and subject heading field for methods.
S_3	SEARCH SYNTAX AND ((crime or crime prevention).sh. OR crime prevention.cc. OR crime.cw. OR (crime or criminology or crimonology).id. OR exp. Crime Prevention/ OR (criminal behavior & juvenile delinquency or rehabilitation or criminal rehabilitation & penology or criminal law & adjudication or mediation & conflict resolution or crime prevention or police & legal personnel).cc)	Using the Boolean operator AND to filter for topic (crime) synonyms in the classification code field, the classification word field, the key concepts field and subject heading field. Also, using the 'explode' thesauri function for the topic 'crime prevention'.
S_4	SEARCH SYNTAX AND (((meta analysis or metasynthesis or "systematic review").md. OR meta analysis.sh.)) AND ((crime or crime prevention).sh. OR crime prevention.cc. OR crime.cw. OR (crime or criminology or crimonology).id. OR exp. Crime Prevention/ OR (criminal behavior & juvenile delinquency or rehabilitation or criminal rehabilitation & penology or criminal law & adjudication or mediation & conflict resolution or crime prevention or police & legal personnel.cc)))	Using the Boolean operator AND to combine S_2 and S_3.
S_5	(meta analysis or metasynthesis or "systematic review").md. OR meta analysis.sh.	Using the method field and subject heading field for methods.
S_6	(crime or crime prevention).sh. OR crime prevention.cc. OR crime.cw. OR (crime or criminology or crimonology).id. OR exp. Crime Prevention/ OR (criminal behavior & juvenile delinquency or rehabilitation or criminal rehabilitation & penology or criminal law & adjudication or mediation & conflict resolution or crime prevention or police & legal personnel).cc	Using the classification code field, the classification word field, the key concepts field and subject heading field for topic. Also, using the 'explode' thesauri function for the topic 'crime prevention'.
S_7	((meta analysis or metasynthesis or "systematic review").md. OR meta analysis.sh.) AND ((crime or crime prevention).sh. OR crime prevention.cc. OR crime.cw. OR (crime or criminology or crimonology).id. OR exp. Crime Prevention/ OR (criminal behavior & juvenile delinquency or rehabilitation or criminal rehabilitation & penology or criminal law & adjudication or mediation & conflict resolution or crime prevention or police & legal personnel).cc)	Using the Boolean operator AND to combine S_5 and S_6.

Table 2 Sensitivity and precision calculations for all search strategies (Q-GS refers to quasi-gold standard)

Search strategy	Search strategy description	n Q-GS identified by search	n records retrieved in search	Precision	Sensitivity
S_1	Original search	171	4400	3.9	67.1
S_2	Original search filtered by methods terms	137	843	16.3	53.7
S_3	Original search filtered by topic terms	119	981	12.1	46.7
S_4	Original search filtered by methods and topic terms	94	213	44.1	36.9
S_5	PsycINFO filtered by methods terms	189	35,512	0.5	74.1
S_6	PsycINFO filtered by topic terms	168	94,408	0.2	65.9
S_7	PsycINFO filtered by methods and topic terms	121	917	13.2	47.5

Interestingly, S_5 and S_6 retrieved a somewhat different sample of the quasi-gold standard than the first set of search strategies. Figure 2 presents the overlap between S_5–S_7 and S_1. This shows that, when compared with S_1, 52 unique citations are returned by S_5 and 47 unique citations are returned by S_6. The results from S_7 are entirely subsumed in the results from S_1. This suggests that a sensitivity-maximising search could use the index terms featuring in searches 5–7 not as filters, but as a means of broadening the search strategy (with the Boolean operator OR) to return a greater number of relevant citations.[8]

From these results, it would appear that methodologic index terms are reasonably effective at identifying systematic reviews in both the results of the original search syntax and in PsycINFO more generally, with S_2 and S_5 reporting high rates of sensitivity. So, for this database at least, the concerns raised by Kugley et al. (2016) about methodological index terms being inconsistently applied appear not to manifest in a prohibitive way. It is likely that the tradition of transparent reporting throughout a systematic review prompts the explicit inclusion of methods in the abstract, where the methodological index terms are populated from.

Topic index terms are comparably weaker at identifying records on crime reduction topics (with less improvements in sensitivity seen), but perhaps this is to be expected given the diversity of expression across the many disciplines that the crime reduction literature spans (see Tompson and Belur 2016), some or all of which may be unfamiliar to the human indexers. It appears that S_2 was the best trade-off between sensitivity and precision, but the original search strategy used was also good at balancing the two aims.

It is difficult to contextualise the performance of these filter terms, since no comparable filters have been developed. Whilst filters *have* been designed for identifying systematic reviews in PsycINFO for other topics,[9] the author is unaware of any publications documenting their efficiency or effectiveness.

---

[8] The reason this was not attempted in the original search strategy is because it could not be applied consistently across the 14 databases.

[9] See https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-identify-systematic-reviews. These sites were consulted in the generation of the original search syntax for the What Works in Crime Reduction project.
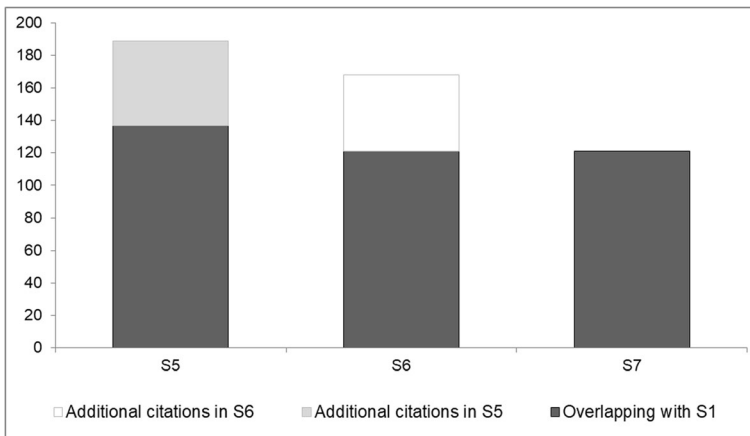
Fig. 2 Sensitivity and precision calculations for all search strategies (Q-GS refers to quasi-gold standard)

Discussion

This study reports an empirical test of the effectiveness and efficiency of various search filter terms in the PsycINFO database, against an original search strategy, with the objective of identifying evidence syntheses with a measured crime reduction outcome. Since the searching for studies phase of a systematic review, or meta-review, constitutes a large investment of effort, strategies to maximise the effectiveness and efficiency of a database search can yield a considerable saving of researcher time. Given that there are no reported empirical data on filters used in the field of crime prevention and allied fields such as criminal justice, this modest study represents an important advancement in knowledge in the area of evidence-based methods for systematic review methodology. The variety of search filter terms tested caters to different information requirements that researchers might have.

The results presented in Table 2 illustrate that there is typically a trade-off between sensitivity and precision. The strategy with the greatest sensitivity—that identified the greatest proportion of the quasi-gold standard (S_5)—also had the second smallest precision rate. In a real-world application of this search strategy, the hit rate (precision) may have been compromised further by screener fatigue when reviewing the 35,000 records generated by this (solely) controlled vocabulary search.[10] The most precise search strategy using the original search string was S_4 which filtered by both methods and topic index terms. This, similarly, resulted in a considerable loss of sensitivity.

It is not possible to comment on which strategy is the 'best', since different strategic approaches are appropriate to different styles of systematic review. As Moher et al. (2015) lucidly state, there is a 'family' of systematic approaches to reviewing and appraising evidence. Scoping, or 'mapping', activities can be done to gain an overview of a broad or emerging policy area (see Levac et al. 2010 for a methodological approach to this and Schucan Bird et al. 2016 for a Criminal Justice example). These

---

[10] The author thanks David Wilson for this injection of real-world experience into interpreting how different search strategies might play out in practice.

evidence 'maps' can be undertaken prior to doing a systematic review to identify the extent and nature of evidence on a given topic and to identify areas where gaps exist in the literature. Rapid evidence assessments (REA) employ systematic methods but typically truncate the comprehensiveness of a full systematic search. As the name suggests, these are typically undertaken to assemble evidence swiftly, often in a practice or policy environment, where some bias is acceptable if acknowledged. Systematic reviews are another member of the family, and these can vary on many dimensions, least not the breadth or depth of a topic (Gough et al. 2013). Finally, there are 'reviews of reviews', where summaries of evidence are produced that transcend multiple systematic reviews (e.g., see Caudy et al. 2016), commonly with the goal of informing policy or practice.

When resources or time are limited, search strategies that maximise precision are required. For example, for researchers completing an REA, or lone researchers completing a systematic review, S_4 might appeal, since a balance of sensitivity and precision is achieved, which minimises the time spent screening the citations against inclusion criteria. If time is less pressing in an REA, then a search without natural language search terms could be employed, such as that in S_7.

For researchers looking to complete a comprehensive systematic review, or review of reviews, that aims to minimise the possibility of a biased collection of studies, strategies that maximise sensitivity are called for. For instance, S_5 might be appropriate if an inexhaustible source of resources is available, or text-mining software[11] is used at the screening stage to reduce the resources needed to screen records (see O'Mara-Eves et al. 2015 for a systematic review on this topic). Failing that, S_1, which was the search syntax using a combination of natural language terms with controlled vocabulary (see the Appendix), might be favoured. The analysis of overlap between the searches using the original search strategy (which combined natural language with [non-filter] controlled vocabulary) and the searches using just the filter terms revealed that both sets of searches identified a different subset of the quasi-gold standard. This suggests that researchers striving for sensitivity-maximisation could apply the index terms used in this paper to *broaden* rather than limit their search, if commensurate resources are available to review the retrieved records.

This small-scale study has some notable limitations. Pre-eminently, the results and the implications reported here are exclusive to both the PsycINFO database and the records sought—that is, evidence syntheses with a crime reduction outcome. They therefore cannot be extrapolated to other databases or topics. It is hoped that by providing empirical data on the performance of both methods and topic filter terms, researchers looking to identify one or the other, or even both, might be able to be guided when they design their search strategy in PsycINFO in alignment with the resources at their disposal. Furthermore, research teams with modest resources may begin by searching for systematic reviews on a specific crime reduction topic from which to harvest primary studies. A strategy such as this would only require a 'top up search' if the systematic reviews were of sufficient quality (as suggested by Eady et al. 2008)—meaning that the results presented here may be of use at that early stage of data collection.

---

[11] For example, available in EPPI Reviewer 4, see https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2967

Secondly, the empirical data derived from the testing of the search strategies is only as good as the quasi-gold standard. It cannot be assumed that the quasi-gold standard used in this study represents all relevant studies fulfilling the inclusion criteria of evidence syntheses with crime reduction outcomes. However, as mentioned previously, given the comprehensiveness of the original search strategy (see Tompson and Belur 2016), which searched many dozens of journals longitudinally, this limitation is hopefully mitigated. This last point has a bearing on generalisability, as search filters are often developed with gold standards that have been derived from one or two journals from particular years, which cannot then be generalised to other journals or other years. Whilst some medical journals may not have been adequately captured in the quasi-gold standard used in this study, the journals that comprised the quasi-gold standard did span the gamut of social sciences and related fields over several decades, meaning that the generalisability of the search filter results presented here is credible.

That said, external validity cannot be established for this study because it did not use an independent set of records (i.e., a separate quasi-gold standard) to test the filters. As Jenkins (2004) points out, establishing external validity involves completing a 'third-generation' type of filter development, whereby search filter terms are derived through frequency analysis of relevant records, or in a statistically objective way, before they are subject to testing with an independent gold standard. This third-generation filter development is ostensibly more robust than the second-generation approach taken by this study. However, as White et al. (2001) note, subjectivity encroaches on these third-generation studies since decisions need to be made on what terms to exclude from analysis and imposing an arbitrary 'cut-off' point on the frequencies.

The study presented here can be extended in several ways. First, further research on filter performance in PsycINFO would be advised to adopt a third-generation perspective in either the derivation of the filter terms or the validation phase, using a dataset that is independent of the original generation of the search filter. As is common in library studies, a retest of the search filter terms presented in this study at a future time, with an updated quasi-gold standard, would ascertain if precision and sensitivity scores remain stable. Variations of the method used in this paper to test search filters are commonly applied within information science (for example, see White et al. 2001; Sladek et al. 2006), and these could be used to generate additional empirical data on filter performance. In particular, other metrics such as specificity might be useful to calculate to further knowledge, in the acknowledgement that not all scholars are convinced of the merits of sensitivity and precision (Kagolovsky and Moehr 2003).

Importantly, the results of this study are confined to the broad topic of crime reduction and the PsycINFO database (but may have relevance for other databases accessed via the Ovid interface). Since literature on information retrieval in criminology is at a nascent stage, future studies might fruitfully look to test more precise filter term performance, for example searching for 'offender rehabilitation' studies or specific interventions, in a range of databases relevant to crime and criminal justice, which will be based on different index terms than those used in this study, albeit there may be some overlap. Conversely, a lot can be learnt about indexing conventions from examining those studies that are discoverable in a database but not retrieved by a search strategy.[12] As Wilson (2009) and Tompson and Belur (2016) note, other promising databases with sophisticated functionality are Criminal Justice Abstracts (via EBSCOHost), Criminal Justice Database in ProQuest, International Bibliography

of the Social Sciences (via ProQuest), SCOPUS, and Sociological Abstracts (via ProQuest). Given that, as Jenkins (2004: 157) in her review of search filter development found 'there is no great consensus in the approach to filter design', there is enormous scope for knowledge accumulation on this aspect of systematic review methodology in the field of criminology and criminal justice.

In conclusion, evidence synthesis based on a biased set of studies (due to improper searching) is liable to generating biased results. Therefore, database searches cannot be the sole source of data collection in a systematic review, particularly as much crime prevention literature is not discoverable in electronic bibliographic databases (Tompson and Belur 2016; Wilson 2009). That said, database searches comprise a sizeable component of a search strategy within a systematic review and incur a considerable resource cost. Filter terms, as tested in this paper, can be used judiciously to maximise the effectiveness and/or efficiency of a database search, which can aid researchers when managing their resources on a review.

Appendix

1. (abduct* or aggressi* or ("anti*social adj2 behavio*")or ("anti?social" adj2 "behavio*") or arrest* or arson* or assault* or blackmail or "bodily harm" or burglar* or "calls for service" or convict* or counterfeit* or crim* or delinquen* or ((dr*nk or dangerous) adj2 driv*) or DUI or explosi* or firearm or fraud or homicide or incarcerat* or incest or infanticide or kidnapping or (knife adj2 crim*) or (law adj2 breaking) or manslaughter or (money adj2 laundering) or murder or offen* or prostitute* or "public disorder" or rape or recidivis* or reconvict* or re-offen* or reoffen* or riot* or robber* or shoot* or shoplift* or terrori* or theft* or unlawful or vandalism or violen* or weapon or wounding or (drug adj2 misuse) or (drug adj2 abuse) or (drug adj2 market) or (drug adj2 deal*) or (drug adj2 traffick*) or (drug adj2 supply) or (drug adj2 possess*)).ab,ti.

   2. exp Crime/ or criminal behavior/ or antisocial behavior/ or juvenile delinquency/ or criminals/ or gangs/ or criminals/ or perpetrators/ or female criminals/ or male criminals/ or mentally ill offenders/ or crime/ or antisocial behavior/ or criminal behavior/ or juvenile delinquency/ or recidivism/ or sexual abuse/ or terrorism/ or violence/

   3. #1 OR #2

   4. (prevent* or deterr* or diver* or reduc* or control* or increas* or impact or (cost adj1 effective*) or (cost adj1 benefit) or displace* or (diffus* adj2 benefit)).ab,ti.

   5. exp. Crime Prevention/ or exp. criminal rehabilitation/ or "costs and cost analysis"/ or resource allocation/ or recidivism/ or spatial frequency/ or temporal frequency/

   6. #4 OR #5

   7. #3 AND #6

   8. (meta analy* OR meta?analy* OR meta*analy* OR (review adj5 evidence) OR (review adj5 literature) OR (review adj5 evaluation) OR (systematic* adj2 review) OR "comprehensive bibliography" OR "Integrat* Review" OR "comprehensive Review" OR (quantitative adj2 review) OR (qualitative adj2 review) OR (thematic adj2 analy*) OR meta*narrative OR (data adj2 extract*) OR (data adj2 integrat*) OR (narrative adj2

review) OR (realist adj2 review) OR (campbell adj3 review) OR (cochrane adj3 review)).ab,ti.

9. exp meta analysis/ or exp. response to intervention/ or exp. evidence based practice/

10. #8 OR #9

11. #7 AND #10

12. Limit #11 to (human and English language and year = "1975–2017")

13. Limit #12 to (abstract collection or bibliography or chapter or dissertation or encyclopedia entry or journal article or publication information or reviews).

References

Bowers, K., Tompson, L., & Johnson, S. (2014a). Implementing information science in policing: mapping the evidence base. Policing: *A Journal of Policy and Practice, 8*(4), 339-352.

Bowers, K. J., Johnson, S. D., Tilley, N., Tompson, L., & Belur, J. (2014b). Protocol for work package 1: Identifying existing systematic reviews of crime reduction. Available at: http://discovery.ucl.ac.uk/1462098/1/What%20works%20Work%20Package%201Protocol%20May2014.pdf. Accessed 29 May 2019.

Caudy, M. S. T., Tang, F. S., & L Watson, C. (2016). Evidence mapping to advance justice practice. In D. Weisburd, D. P. Farrington, & C. Gill (Eds.), *What works in crime prevention and rehabilitation* (pp. 261–290). New York: Springer.

Cohen, A. M., Hersh, W. R., Peterson, K., & Yen, P. Y. (2006). Reducing workload in systematic review preparation using automated citation classification. *Journal of the American Medical Informatics Association, 13*, 206–219.

De Waard, J. (2018). What Works?: A systematic overview of recently published meta evaluations / synthesis studies within the knowledge domains of Situational Crime Prevention, Policing, and Criminal Justice Interventions, 1997–2018 (update July 2018). https://doi.org/10.13140/RG.2.2.31254.86089.

Eady, A. M., Wilczynski, N. L., & Haynes, R. B. (2008). PsycINFO search strategies identified methodologically sound therapy studies and review articles for use by clinicians and researchers. *Journal of Clinical Epidemiology, 61*(1), 34–40. https://doi.org/10.1016/j.jclinepi.2006.09.016.

Glanville, J., Bayliss, S., Booth, A., Dundar, Y., Fernandes, H., Fleeman, N. D., … Welch, K. (2008). So many filters, so little time: the development of a search filter appraisal checklist. *Journal of the Medical Library Association, 96*(4), 356–61. https://doi.org/10.3163/1536-5050.96.4.011

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2013). *An introduction to systematic reviews*. London: Sage Publications Ltd..

Jenkins, M. (2004). Evaluation of methodological search filters–a review. *Health Information and Libraries Journal, 21*(3), 148–163. https://doi.org/10.1111/j.1471-1842.2004.00511.x.

Kagolovsky, Y., & Moehr, J. R. (2003). Current status of the evaluation of information retrieval. *Journal of Medical Systems, 27*(5), 409–424.

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A. M. K., Hammerstrøm, K., & Sathe, N. (2016). *Searching for studies: a guide to information retrieval for Campbell Systematic Reviews*. Campbell Methods Guides 2016:1. https://doi.org/10.4073/cmg.2016.1.

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*, 69.

Moher, D., Stewart, L., & Shekelle, P. (2015). All in the family: systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. *Systematic Reviews., 4*, 183.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews, 4*(1), 5. https://doi.org/10.1186/2046-4053-4-5.

Petrosino, A. J. (1997). "What works?" revisited again: a meta-analysis of randomized field experiments in individual-level interventions. Newark: Rutgers University.

Pratt, T. C. (2014). Meta-analysis in criminal justice and criminology: what it is, when it's useful, and what to watch out for. In T. C. Pratt, *Advancing quantitative methods in criminology and criminal justice* (pp. 58–74). Routledge. https://doi.org/10.4324/9781315877907

Sampson, M., Zhang, L., Morrison, A., Barrowman, N. J., Clifford, T. J., Platt, R. W., … Moher, D. (2006). An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Medical Research Methodology, 6*, 33. doi https://doi.org/10.1186/1471-2288-6-33

Schucan Bird, K., Vigurs, C., Quy, K., & Gough, D. (2016). *Criminal justice interventions with perpetrators or victims of domestic violence: a systematic map of the empirical literature* (p. UCL). London.

Sladek, R., Tieman, J., Fazekas, B. S., Abernethy, A. P., & Currow, D. C. (2006). Development of a subject search filter to find information relevant to palliative care in the general medical literature. *Journal of the Medical Library Association, 94*(4), 394–401.

Thompson, J., Davis, J., & Mazerolle, L. (2014). A systematic method for search term selection in systematic reviews. *Research Synthesis Methods, 5*(2), 87–97. https://doi.org/10.1002/jrsm.1096.

Tompson, L., & Belur, J. (2016). Information retrieval in systematic reviews: a case study of the crime prevention literature. *Journal of Experimental Criminology 12*(2):187–207

Weisburd, D., Farrington, D., & Gill, C. (Eds.). (2016). *What works in crime prevention and rehabilitation: lessons from systematic reviews*. New York: Springer-Verlag.

Weisburd, D., Farrington, D. P., & Gill, C. (2017). What works in crime prevention and rehabilitation: An assessment of systematic reviews. *Criminology & Public Policy, 16*(2), 415–449

Wells, E. (2009). Uses of meta-analysis in criminal justice research: a quantitative review. *Justice Quarterly, 26*(2), 268–294.

White, V. J., Glanville, J. M., Lefebvre, C., & Sheldon, T. a. (2001). A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *Journal of Information Science, 27*(6), 357–370. https://doi.org/10.1177/016555150102700601.

Wilson, D. B. (2009). Missing a critical piece of the pie: simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology, 5*(4), 429–440. https://doi.org/10.1007/s11292-009-9085-5.

Lisa Tompson is an Associate Professor at UCL. Her research interests include evidence generation for practitioners, particularly using systematic review methods and spatio-temporal crime analysis. In recent years she has worked on the Commissioned Partnership Programme that underpinned the UK What Works Centre for Crime Reduction.

Affiliations

Lisa Tompson[1]

[1]   Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK