Submitted on: 19.06.2015

Finding the best available evidence: how can we know?

Sarah Hayman
Flinders Filters, Flinders University, Adelaide, Australia.
sarah.hayman@flinders.edu.au

Jennifer Tieman
CareSearch and Flinders Filters, Flinders University, Adelaide, Australia.
jennifer.tieman@flinders.edu.au

Abstract:

The importance (in all scientific fields) of finding and using evidence is growing rapidly, with increased recognition that decisions should be based on sound evidence. Key to finding this evidence is effective searching. Alongside this imperative, the searching context is becoming more complex. The number of articles indexed is enormous and increasing. In the medical field, PubMed contains over 24 million citations with over 1 million entered in 2014. Effective searching requires an understanding of database mechanisms and the terminology (including associated thesauri) of each subject. Searchers need an understanding of the requirements of the end user: what is considered relevant and what are the levels of evidence?

We suggest that a scientific approach be taken to the searching process, to ensure that the best available evidence is found, and that search methodology is tested and validated. What methods can we employ to indicate what we might have missed in our search? Search results should be tested and results fed back into the search, to improve searching effectiveness and thereby outcomes for the end user.

Search filters are validated search strategies, created using known methodology, for a given bibliographic database. The relevant terminology and database mechanisms are built into a strategy that is created from, and tested against, subsets of a gold standard set of references. Results are screened by external reviewers with expert subject knowledge, to minimise bias. The search filter performance is rated for precision and sensitivity, to provide known effectiveness in a standard set that can be extrapolated to open search. Details of the methodology and the filter performance are published for transparency.

CareSearch and Flinders Filters have developed a number of subject-based search filters. This paper discusses the importance of evidence-based searching; how these search filters are developed and lessons for general searching in scientific literature.

Keywords: Search strategies; Retrieval performance measures; Search Filters; Searching.

1 THE IMPORTANCE OF EFFECTIVE SEARCHING

In all fields of human endeavour, there is an increasingly strong requirement to discover and acquire the best available evidence, on which to base decisions, protocols, guidelines, methodologies and research. Evidence-based practice has emerged strongly in the health sciences (medicine, nursing and allied health), and this paper is based on work undertaken in the health arena. However, evidence based approaches (such as the undertaking of systematic reviews) are now embedded in many other scientific and technological areas, with a strong presence particularly in environmental science, engineering and computer science, and indeed extending into related areas of policy, education, management and social sciences (examples of evidence-based approaches in other fields are found in Papaioannou et al., 2010, Ton et al., 2013 and Woodcock et al., 2014).

In health, decisions made about treatment of patients can have significantly different outcomes depending on the evidence on which those decisions are based. Adverse effects can result from wrong information in any field. Scientific development builds on research that has gone before and must be underpinned by accurate information. In an example from the field of software engineering, the *Guidelines for performing Systematic Literature Reviews in Software Engineering* state:

> "unless a literature review is thorough and fair, it is of little scientific value. This is the main rationale for undertaking systematic reviews. A systematic review synthesises existing work in a manner that is fair and seen to be fair. For example, systematic reviews must be undertaken in accordance with a predefined search strategy. The search strategy must allow the completeness of the search to be assessed. " (Kitchenham, 2007.)

As librarians understand well, the central point in the discovery of the best available evidence is the search.

Systematic reviews are considered the highest level of evidence in the evidence hierarchy (National Health and Medical Research Council, 2009) within evidence-based medicine. Other types of study include randomised controlled trials (RCT), cohort studies, case control studies, cross sectional surveys and case reports. Effective searching will find and distinguish between these study types. Systematic reviews will require all relevant studies to be found in relation to the research question. A missed piece of evidence can affect the outcome of the review and may ultimately have an adverse health effect.

Evidence based medicine has been described as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett et al., 1996). This requires the critical evaluation of the best available clinical research, reported in publications, and the application of this in conjunction with clinical expertise. Searching

for the best available published research is a key step in the process, whose crucial outcome is the delivery of high quality patient care.

Allied to evidence-based practice is the concept of knowledge translation, also reliant on the finding of evidence, but focussed strongly on the application of that evidence in practice. The Canadian Institute of Health Research's KT Clearinghouse addresses the bridging of the knowledge to action gap (2000-2014). In their Knowledge to Action cycle, the step of knowledge inquiry is central. This "knowledge inquiry" step is the search.

2 CHALLENGES FOR EFFECTIVE SEARCHING

Accompanying the imperative to search and find the best available evidence, to inform practice, research and policy, there is the challenge of searching effectively through an enormous and growing quantity of bibliographic references, and within databases of complex and differing structures.

The quantity of published and indexed articles is vast, even without considering the "grey literature" that must also be searched for a comprehensive search, such as one undertaken for a systematic review. The Scopus database contains 55 million records; Web of Science captures 65 million cited references annually; PubMed contains over 24 million records at the time of writing (May, 2015) and over 760,000 articles were indexed for Medline in 2014.

The technical challenges of searching are increasing, with a range of databases available in most scientific and technological fields, often using different thesauri and different search syntax. Effective searching requires an understanding of Boolean search techniques as well as knowledge of how they have been implemented in the particular search interface of each database. As McGowan and Sampson have written (2005), "expert searchers need to understand the specifics about data structure and functions of bibliographic and specialized databases, as well as the technical and methodological issues of searching."

Comparative studies have shown the need to search in more than one database to retrieve all relevant literature for a particular topic (e.g. Betrán at al., 2005, Michaleff at al., 2011 and Rollin et al, 2010). Cochrane systematic review guidelines suggest searching CENTRAL, Medline and Embase for all reviews, together with relevant subject-specific databases (Higgins and Green, 2011).

Another important aspect of undertaking a search for evidence is an understanding of the needs of the end user. If the search is for a systematic review, it must be as comprehensive as possible, because the aim of a systematic review is to uncover all relevant research, without bias, in order to appraise and synthesise the findings to answer the review question. This requires a highly sensitive search strategy (and typically an iterative search) with a degree of tolerance for irrelevant retrievals. If however the search is for a busy clinician who wants to discover good current evidence in a clinical situation, then the search needs a high degree of precision; if the searcher supplies a huge number of references, only some of which are relevant (so as not to miss any), it will be counter-productive for the user who will not want to look through them all and may therefore miss relevant items. Tailoring the search to the requirements of the user is an important aspect of effective searching, and requires a careful balance between sensitivity and precision.

3 APPROACHING SEARCHING SCIENTIFICALLY

How can we ensure the best possible searches? In this paper we would like to suggest some techniques that can be applied to searching, developed from the methodology employed by the team at CareSearch (www.caresearch.com.au) and its associated project Flinders Filters (http://www.flinders.edu.au/clinical-change/research/flinders-filters/) to develop subject-based search filters. These techniques are based on some key principles and can be tweaked to apply to highly sensitive or highly specific searches, as the need arises. These are likely to be principles and approaches already used by expert searchers; we hope that setting them out in this way will be useful and that elements of the approach can be used and adapted as necessary. It should be noted that while this approach is derived from the search filter development model used at CareSearch, it is by no means the full methodology, as detailed in our published papers on the various search filters (CareSearch, 2015). It is a highly abbreviated approach, based nevertheless on the same principles of transparency, thoroughness, iteration and minimisation of bias.

Literature searching, as a key underpinning element of evidence-based practice, must be able to be subjected to a scientific process of rigorous testing and falsifiability. Search strategies should be documented, transparent and reproducible. If a search is known to return 80% of relevant items, then that performance rating is an element of evidence to be taken into account. One type of validated and documented search strategy is the search filter (also known as a hedge).

3.1 Understanding and using Search Filters

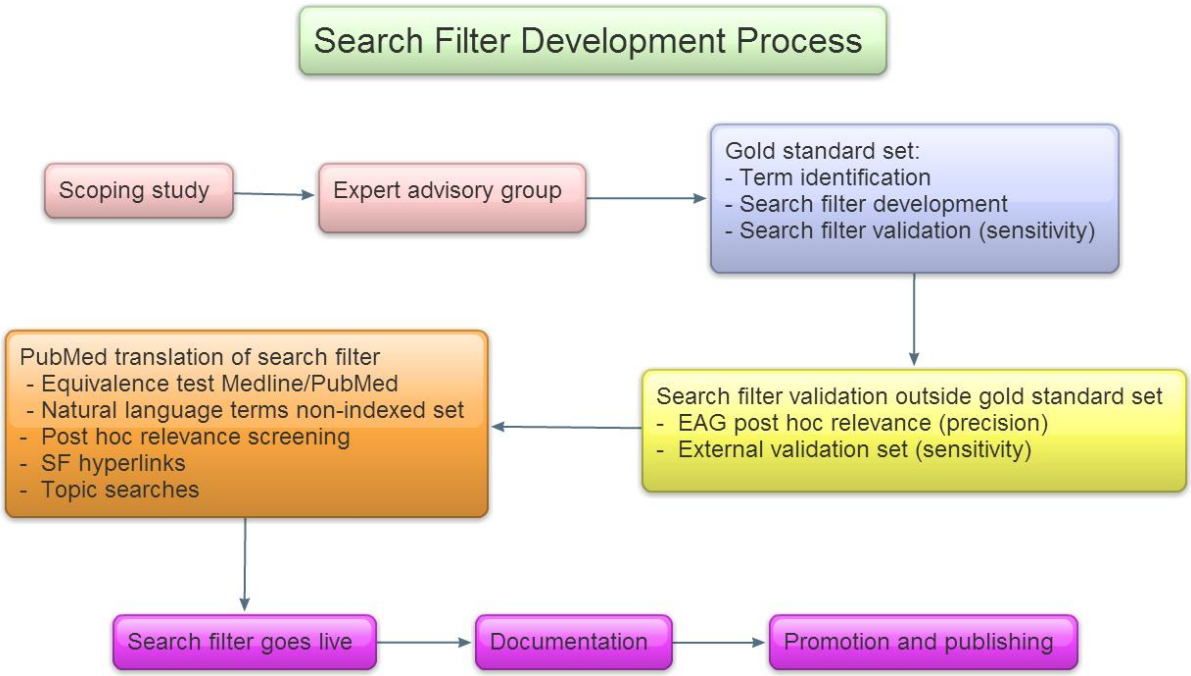
Search filters are validated search strategies, created using a known methodology, to search a given bibliographic database. The relevant terminology and database mechanisms are built into a strategy that is created from, and tested against, subsets of a gold standard set of references. Results are screened by external reviewers with expert subject knowledge, aiming to minimise bias throughout the process. The performance of the search filter is rated for sensitivity and precision, so that it has a known effectiveness in a given standard set that can be extrapolated to an open search. Details of the methodology and the filter performance are published to maintain transparency.

Many search filters are methodological, designed to look for particular study types, such as systematic reviews, diagnostic studies, economic evaluations or outcome studies. The InterTASC Information Specialists' Sub-Group Search Filter Resource is an excellent source of information about methodological search filters (https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home).The search filters developed by CareSearch and its associated project Flinders Filters are topical (subject-based) search filters in palliative care and other subject areas (including lung cancer, heart failure, residential aged care, dementia, bereavement, primary health care and Australian Indigenous health). Our papers detailing the development of the search filters are published and listed on the Flinders Filters website at http://www.flinders.edu.au/clinical-change/research/flinders-filters/publications.cfm.

Search filters are designed for a particular purpose and it is important to understand their function and performance if you are using them for a search. They will not always give you a comprehensive set of results but may be a very good starting point for a search. The benefit of the search filters created and deployed by CareSearch is their ability to be used by health

professionals in clinical practice and research; the technical searching expertise and testing is embedded in the final strategy and published as a link directly into the PubMed database. Busy clinicians who may not have the time or expertise to craft an effective search strategy can click on this link to go straight to a real-time search of known reliability.

Our search filters are created in the following steps:



We would like to focus on these key points in this process:

- Expert Advisory Group (EAG)
- Gold Standard Set
- Term identification
- Validation

The **EAG** is essential to ensure the clinical usefulness of the search filter and to minimise any bias that we (as searching experts but not necessarily subject experts) might bring to the development process. EAG members advise on the scope of the filter, candidate search terms and possible sources of a representative gold standard set; they also test filter retrievals for relevance as part of the filter validation process.

The **Gold Standard Set** is a set of references representing the scope of the subject to be retrieved by the search, and externally confirmed as relevant to the topic. This set is divided into three subsets so that term identification, creation and validation can all be done within different sets of data; again, the aim here is to minimise any potential bias from building and testing in the same set.

Term identification is the process of analysing the titles, abstracts and index terms of the references to identify textwords and controlled headings (usually MeSH terms) to be tested for their retrieval performance in the gold standard set.

Validation encompasses the testing of the search strategy within a subset of the gold standard set, the entire gold standard set and often an external validation set to arrive at a percentage figure for its retrieval performance. Its ability to retrieve items known to be relevant (e.g. within the gold standard set) gives a sensitivity percentage rating; the number of relevant records retrieved out of a total set retrieved by the search strategy gives the precision percentage rating (using relevance assessment by external reviewers).

3.2 Enhancing and testing your searches

These are the four approaches we believe can be adapted to general literature searching, to provide a scientific approach to searching.

Note that we have published a free online resource called Smart Searching (https://sites.google.com/site/smartsearchinglogical/home) which sets out in some detail the four steps or processes aligned to the four steps in the search filter development process described above. In this paper we will not have space to provide as much detail, nor the worked examples provided on that site. We recommend that, if you are interested in that greater level of detail, you visit the free Smart Searching website. We welcome feedback, comments and suggestion on that site so that we can keep improving it for users.

3.2.1 Subject Experts (adaptation of Expert Advisory Group)

The formal expert advisory group which plays a key role in the search filter development process can be equated to external advice that you are able to obtain from an expert in the subject for which you are searching. This person may simply be the researcher or clinician who has requested the search, or may be another colleague. Advice they provide can help reduce bias that you might bring to the search yourself and add a dimension to your search of external knowledge about the subject area. This knowledge can enable them to provide useful advice about appropriate scope for the search (e.g. dates when research in the subject changed significantly, or concepts that are uniquely associated with the topic), relevant terminology (e.g. synonyms in common use), key papers, journals, database, organisations, websites or authors in the field (they may even have a personal collection of papers they can share with you, as a potential sample reference set). They may also be able to undertake a relevance assessment of your search retrievals. If you have a set of items you have retrieved with your first search, you may be able to ask them to glance at it and say whether it is retrieving irrelevant items. This would enable you to adjust the search.

While both librarians and health professionals are busy and always working under time constraints, it can nevertheless be extremely valuable to get some suggestions to inform the development of your search strategy before your search and some feedback after your search – both can supply useful information about the effectiveness of your search that will allow you to analyse and tweak it. If it is not retrieving key papers that have been recommended in the field, why not? Check the index terms and text words and see if any have been missed. If it is retrieving a large number of items that are not relevant, why is this happening? Check the search terms that are retrieving the irrelevant items and see what happens if they are removed.

3.2.2 Sample set (adaptation of Gold Standard Set)

The creation of a formal gold standard set as employed in the development of a search filter is a major piece of work using an established methodology. Without going to those lengths

for a literature search, we nevertheless suggest that creating a sample set of references to guide your search can still be very useful. A sample set of references, known to be relevant to your search topic, gives you a test set for (1) identifying terms used in the literature for your topic and (2) testing the effectiveness of your search in retrieving references known to be relevant. Ideally, the relevance of this sample set is externally verified; otherwise you would be testing your search strategy against itself. Possible sources of a sample set are:

- a collection of papers provided by an expert in the subject
- a published database in the field
- references from key papers known to be relevant (included studies in systematic reviews are an excellent source as they have been assessed as relevant within the systematic review process)
- articles from relevant and authoritative journals in the field.

We advise that you confirm the relevance of the sources of the sample set of references with a subject expert, to minimise bias.

We recommend that you store the references in a reference management program (e.g. EndNote (http://endnote.com/) or Zotero (https://www.zotero.org/), for ease of analysis.

3.2.3 Term Identification

Analysing and testing candidate terms for your search strategy is a very useful technique for ensuring that you have a high performing search strategy that will capture a high proportion of relevant items and a low proportion of irrelevant ones.

In the full search filter development model, we undertake extensive research, analysis and testing of potential search terms for each subject. In general literature searching it is still possible to do some investigation and analysis to help you identify the best terns for your search.

Sources for your terms will be:

- the thesaurus of your database (in the medical world Medical Subject Headings (MeSH) are the most used index terms but other databases use their own thesauri, such as Emtree for Embase, the IEEE Thesaurus and the ERIC Thesaurus)
- expert suggestions of relevant terms
- your analysis of your key references (your sample set).

You might wish to confirm with your expert that your candidate terms are correct and relevant.

In checking thesaurus terms, include broader, narrower and related. It is important to consider the narrower terms in the thesaurus hierarchy that are also relevant to your search topic (e.g. in MeSH, Huntington Disease is a term that sits under dementia in the tree hierarchy. Does your client want Huntington Disease included in his or her dementia search?). The date when a MeSH term was first used may also be an important consideration. You might need to consider whether earlier terms used should be candidate terms. Always look at the alternative non-preferred terms ("Used For" terms) in a thesaurus – in MeSH for example these are called "Entry terms".

You can analyse the frequency of textwords (natural language terms) in searchable fields in your sample reference set, typically the title and abstract fields. This will give you alternative candidate terms to test, ones you know to be associated with relevant references. There are tools available to do this textual analysis, such as Writewords word and phrase counter (http://www.writewords.org.uk/phrase_count.asp) and Concordance (http://www.concordancesoftware.co.uk/). The PubMed PubReMiner tool (http://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi) is an excellent resource for word frequency analysis in PubMed results.

Your searching experience and common sense may well suggest other terms to you that are likely to be associated with your search topic and worth considering as candidate terms for testing.

Use your own judgement about truncating or combining individual terms. (Remember, all strategies will be tested).

3.2.4 Testing (adaptation of Validation)

Build your search strategy by combining your candidate terms and testing as you go against your sample set to see how many are retrieved. You can test each term one by one for its recall ability or test the entire string. In a search filter we do not want redundant terms, and therefore test each term individually for its ability to add to the results; in general literature searching that may be less important, and you can test a whole string of terms together. Working with a reference storage program such as Endnote will assist in managing this process, while you run repeated searches in your database to test retrieval.

Testing the terms and their performance in a set of known relevance is important, as it can assist in identifying what is not retrieved and why; it can identify terms that add nothing to the search results; and it will facilitate adjustment of the search to improve results.

This type of test (assessing retrieval within a set of known relevant items), tests the sensitivity (or recall) of a search, that is, its ability to retrieve relevant items.

Another kind of testing is that undertaken to find out the precision of a search (that is, how many items of the items it has retrieved are relevant?) To assess this, you need external expert assessment of the relevance of number of relevant items your search has retrieved in a sample search in the open database.

For a comprehensive systematic review search, you want a search that is as sensitive as possible and would be less concerned with a high degree of precision; you want to find all relevant items and for that are willing to risk a large number of irrelevant retrievals and sift through the results. For searches for clinicians however, it is likely they will want most items retrieved to be relevant and not to have to wade through a large number of irrelevant items. It is possible and important to dial up or down the sensitivity depending on the requirements of the end user. As sensitivity increases, precision will decrease, and vice versa.

Testing is thus an iterative process that feeds back into the development of your search strategy, improving it each time, and resulting in an improved search that is less likely to miss key references.

4 CONCLUDING REMARKS: SOME RESULTS AND IMPLICATIONS OF THIS APPROACH

The search strategies created in this way will be known to have a high degree of recall, having been informed by external input from one or more subject experts and an external set of known relevant items. They are not just created by a librarian (however skilled in searching) but have additional rigour resulting from external validation and reduction of bias. The work required in applying this method need not be lost, but can be captured by publishing the searches and by saving them as alerts if the database allows (as many do).

Searches can be contributed to a public website for sharing expert search strategies (e.g. http://pubmedsearches.blogspot.com.au/), and the search strategy can be embedded as a link in a specialist subject library guide, such as are now appearing on many university websites and intranets. The search may also be able to be used as a basis for a search for the same topic in a different database, translating the terms for the relevant thesaurus and retesting the text words.

While the methods described here are not the full search filter development model, nevertheless this modified approach can provide some evidence about the effectiveness of a search, and the quality of its development process. Librarians may be able to use the results of their testing as an indication of searching expertise and a potential justification of why they should be included in a systematic review team (though this method needs to be used with caution when undertaking a comprehensive systematic review search). Other uses can be for a performance appraisal with a manager, or when applying for a new job.

This evidence-based approach to our own practice, using critical reflection and external validation, is something librarians have not always considered doing in relation to our searching, and we believe it can be a very powerful tool.

Applying evidence-based techniques to the search itself will ensure a higher standard of searching and importantly a better outcome for the end users of searches for scientific evidence.

Acknowledgments

The authors gratefully acknowledge the support of the Australian Department of Health in funding the CareSearch Knowledge Network Project and Flinders University of South Australia for funding support for the Flinders Filters project. They would also like to acknowledge Health Librarians Australia and Medical Director (formerly Health Communications Network) for the ALIA HLA/HCN Health Informatics Innovation Award, 2012, which supported development of the Smart Searching website.

References

Betrán, A.P., Say, L., Gülmezoglu, A.M., Allen, T. and Hampson, L. 2005. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. BMC Medical Research Methodology, 5:6-6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC548692/ [Accessed May 8, 2015].

Canadian Institute of Health Research (CIHR). 2000-2014. KT Clearinghouse: The Knowledge-to-Action Cycle. Available from http://ktclearinghouse.ca/knowledgebase/knowledgetoaction [Accessed May 8, 2015].

CareSearch. 2015. CareSearch Peer Reviewed Journal Articles. Available from http://www.caresearch.com.au/caresearch/tabid/431/Default.aspx [Accessed May 15, 2015].

Higgins, J.P.T., Green, S, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0. The Cochrane Collaboration; updated March 2011. Available at http://handbook.cochrane.org/ [Accessed May 8, 2015].

Kitchenham, B. 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering, Version 2.3, EBSE Technical Report EBSE-2007-01. Keele University and University of Durham. Available from https://www.cs.auckland.ac.nz/~norsaremah/2007%20Guidelines%20for%20performing%20SLR%20in%20SE%20v2.3.pdf. [Accessed May 8, 2015].

Mcgowan, J. and Sampson, M. 2005. Systematic reviews need systematic searchers. Journal of the Medical Library Association. 93, 1: 74-80. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545125/ [Accessed May 8, 2015].

Michaleff, Z. A., Costa, L.O.P., Moseley, A.M., Maher, C.G., Elkins, M.R., Herbert, R.D. and Sherrington, C. 2011. CENTRAL, PEDro, PubMed, and EMBASE Are the Most Comprehensive Databases Indexing Randomized Controlled Trials of Physical Therapy Interventions. Physical Therapy. 91 (2): 190-197.

National Health and Medical Research Council. 2009. NHMRC Levels of Evidence and Grades for Recommendations for Developers of Clinical Practice Guidelines. Available at: https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf. [Accessed May 8, 2015].

Papaioannou, D., Sutton, A., Carroll, C., Booth, A. and Wong, R. 2010. Literature searching for social science systematic reviews: consideration of a range of search techniques. Health Information & Libraries Journal. 27(2):114-22.

Rollin, L., Darmoni, S., Caillard, J-F. and Gehanno, J-F. 2010. Searching for high-quality articles about intervention studies in occupational health-what is really missed when using only the Medline database? Scandinavian journal of work, environment & health.: 484-487.

Sackett, D., Rosenberg, W., Gray, M., Haynes, B. and Richardson, S. 1996. Evidence based medicine: what it is and what it isn't. BMJ, 312, 13 January. Available from: http://www.bmj.com/content/312/7023/71. [Accessed May 8, 2015]. doi: http://dx.doi.org/10.1136/bmj.312.7023.71

Ton, G., de Grip, K., Klerkx, L., Rau, M-L., Douma, M., Friis-Hansen, E., Triomphe, B., WatersBayer, A. and Wongtschowski, M. (2013) Effectiveness of innovation grants to smallholder agricultural producers: an explorative systematic review. EPPI-Centre, Social

Science Research Unit, Institute of Education, University of London. Available at: http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=4soRhylhV4A%3d&tabid=3401 [Accessed May 8, 2015].

Woodcock, P., Pullin, A.S. and Kaiser, M.J. 2014. Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: A methodology. Biological Conservation. 176(0):54-62.