What Works *for* Children's Social Care

Cardiff University
Prifysgol Caerdydd

SEARCHING FOR RESEARCH STUDIES IN CHILDREN'S SOCIAL CARE: SOME TECHNIQUES AND TOOLS

January 2021

What Works *for* Children's Social Care

Acknowlegements

The authors gratefully acknowledge the advice and support of Ulugbek Nurmatov (School of Medicine, Cardiff University), Helen Morgan & Simone Willis (SURE, Cardiff University) during this research study and Jonathan Scourfield (Professor of Social Work, What Works for Children's Social Care, CASCADE, Cardiff University) for advising during the study and report write up.

Particular thanks are due to Zac Ghouze (EPPI-Centre, UCL Institute of Education, London) for superb support in relation to training the EPPI-reviewer classifier and managing outputs; and Vicky Clayton (Senior Researcher, What Works for Children's Social Care) for her detailed review of this report and very helpful suggestions.

Authors

Weightman A.L., Specialist Unit for Review Evidence (SURE), Cardiff University; **Searchfield L.**, Specialist Unit for Review Evidence (SURE), Cardiff University; **Lambert P.**, School of Chemistry, Cardiff University; **Thomas J.**, EPPI-Centre, University College London

Funding

Department for Education, England.

About What Works for Children's Social Care

What Works for Children's Social Care seeks better outcomes for children, young people and families by bringing the best available evidence to practitioners and other decision makers across the children's social care sector. We generate, collate and make accessible the best evidence for practitioners, policy makers and practice leaders to improve children's social care and the outcomes it generates for children and families.

About CASCADE

CASCADE is concerned with all aspects of community responses to social need in children and families, including family support services, children in need services, child protection, looked after children and adoption. It is the only centre of its kind in Wales and has strong links with policy and practice.

To find out more visit the Centre at: whatworks-csc.org.uk, or CASCADE at: sites.cardiff.ac.uk/cascade

If you'd like this publication in an alternative format such as Braille, large print or audio, please contact us at: info@whatworks-csc.org.uk

CONTENTS

ABSTRACT

Introduction

Producing an up-to-date summary of 'what is known' about a particular topic requires a detailed review of research studies and can be very time consuming, taking months or even years depending on the specific question(s) being asked. The aim of reviewers is to identify, by a literature search and screening of the papers identified, as many as possible of the relevant research studies in the shortest possible time. Some tools that might help with this process were explored.

Methods

Two very different reviews in children's social care were chosen as case studies to look at methods that might assist with the study identification process and introduce efficiencies into the traditional review methodology. The main context of the research was to look at the value of these methods in updating these literature reviews.

The case studies were:

- a review of systematic reviews in children's social care; and
- a systematic review exploring the impact of Intensive Family Prevention Services on out-of-home placement

The methods explored were:

Text analysis of relevant studies to identify search terms;

1. The choice of databases and other information sources to maximise the identification of relevant studies;
2. Citation analysis (the identification of newer papers that have referenced relevant papers) as a search technique;
3. Co-citation analysis (identifying citation networks where studies have two or more references in common) as a search technique;
4. Ranking of papers using machine learning to speed up the identification of relevant studies during title & abstract screening.

These were compared with traditional methods (where human researchers decide on search terms, search databases, and screen abstracts) for the two review questions.

Results

Based on the findings from the two case studies, *text analysis* of relevant research studies can help identify useful terms for searching and a wide variety of search terms are likely to be needed to develop a sensitive search that identifies the majority of relevant studies. A wide range of *databases and websites* also need to be searched. Increasing the sensitivity of the search also increases the number of records that need to be screened with an accompanying risk of screening fatigue and missing some of the relevant studies Reviewers will need to find a trade-off within their search strategy, using supplementary search techniques (such as reference list checking) to fill in the gaps. *Citation analysis* may be a useful supplementary search technique. *Machine learning* can assist with identifying the relevant studies within a very large set of search results, particularly if the classifier can be pre-trained with known relevant and irrelevant records. This could reduce the requirement for the trade-off described above.

Conclusions

Those seeking research information in the field of children's social care should consider using a wide range of search terms and information sources, including websites.

For systematic reviewers, where the identification of the vast majority of research studies for their topic area is paramount, other promising techniques discussed in this paper may be considered. Further research, using additional case studies, is recommended but there are clear implications relating to the development of good practice for reviewers in this field.

INTRODUCTION

A comprehensive research review using traditional methods, to maximise the identification of studies and minimise bias, requires a great deal of time. It can take from 6 weeks to 3.5 years (median 67.3 weeks) depending on the topic (Borah 2017). The literature search and screening of studies for relevance can represent as much as a quarter of this time requirement (predictor.org), not least because an extensive database search and additional search techniques, such as reference list and website checking, are needed to ensure identification of as many as possible of the relevant studies. Further, once the searching is completed, the titles and abstracts identified are commonly screened independently by two reviewers before agreeing on those studies to be looked at in full text to see if they are relevant to the review question and should be included in the review (Kugley 2017).

Within What Works for Children's Social Care, a number of tools have been explored to see if they increased the efficiency of searching for reviews and primary studies of relevance to children's social care. Search efficiency was considered, both in terms of maximising the ability of the search and screening processes to identify as much as possible of the relevant research on a specific topic and/or in reducing the time required to do this.

Using two very different reviews as case studies (1. a review of systematic reviews in children's social care to find reviews for inclusion in the What Works' Evidence Store[1] and 2. the impact of Intensive Family Prevention Services [IFPS] on out-of-home placement (Bezeczky 2019)), five techniques were looked at in some detail. Outcomes from these were then compared with the traditional methods used in each review.

1. **Text analysis** to identify the search terms to use in a database search to maximise the identification of relevant studies while minimising the return of irrelevant ones
2. **Choice of databases and other information sources** to maximise the identification of relevant studies
3. **Citation analysis** as an alternative or supplementary technique to the traditional database search
4. **Co-citation analysis** as an alternative or supplementary technique to the traditional database search
5. **Ranking of papers for relevance using machine learning** to speed up the identification of relevant studies during title & abstract screening

The aims of this paper are:

- To summarise the research methods and findings.
- To draw on the findings to provide guidance for researchers in children's social care.

The emphasis of the research was to assess how helpful such methods might be when updating an existing literature review but findings are also relevant to those conducting (or the conduct of) new reviews.

1 https://whatworks-csc.org.uk/wp-content/uploads/SYSTEMATIC-REVIEWS.pdf

1 TEXT ANALYSIS

The titles and abstracts of relevant papers can be 'mined' to identify specific key words and phrases to use in a search. This is not a routine requirement within the traditional systematic review search methodology (Kugley 2017), but can help in two ways (Stansfield 2017):

- **By identifying a comprehensive set of search terms** to increase the *sensitivity* of the search (the retrieval of as many as possible of the relevant papers), and

- **By identifying a very specific set of search terms** to help the searcher create a search with high *precision* that hones in on exactly the type of research study that they are looking for (by excluding a large proportion of the irrelevant papers).

A systematic search involves finding a middle ground between a search that has good sensitivity but is also precise enough to avoid a huge screening burden (and screening fatigue) which takes time and can lead to relevant studies being missed (Wang 2020).

1.1 Text analysis: Methods

Various tools are available that can be used to analyse a body of text and identify the words and phrases that appear most frequently. Using Voyant and VOSviewer, the analysis of the titles and abstracts of 90 records identified as potential systematic reviews in children's social care was carried out to try and identify an efficient set of search terms which identified both the topic area (children's social care) and the research type (systematic review). A more detailed analysis of 134 titles and abstracts was also undertaken in TerMine.

For this first case study the overall effect of cutting down the number of search terms to those found most frequently in the text analysis was assessed to see if precision could be increased without loss of sensitivity. This exploration was carried out in Medline where study PubMed Identifiers (PMIDs) can be used to quickly assess the percentage retrieval of the known relevant studies.

In the second case study, an analysis of 142 titles/abstracts of research looking at Intensive Family Prevention Services (IFPS) using TerMine was carried out to explore the terms used in the research literature to describe that intervention.

1.2 Text analysis: Findings

Text mining for individual words and phrases in the titles and abstracts of relevant papers helped to identify search terms for each of the reviews. However, identifying a small (precise) set of terms to identify all the research in a specific area was not feasible.

Based on the findings for systematic reviews in children's social care, any increase in sensitivity (the percentage of relevant studies identified) dramatically decreased precision (the exclusion of irrelevant studies) and the trade-off was not worthwhile. Of the 49 reviews identified in Medline via the traditional review method (a database search strategy and supplementary searching methods), the original database search strategy alone identified 86% of these studies (n=42) requiring the screening of 2,341 abstracts. The search was developed using the 40 most common phrases identified by this text mining exercise to describe children's social care as compared to the 72 phrases in the original search strategy This increased sensitivity by one additional study to 88% (n=43) but, contrary to expectation, the number of abstracts to be screened almost doubled to 4,433. Further increases in sensitivity resulted in large further losses of precision and a growing screening burden. The difficulty in identifying a set of precise but sensitive search terms appeared to be related to the large range of words used by researchers to describe both the topic (children's social care) and the research design (systematic review). The Termine analyses identified the following phrases in order of frequency:

- **Children's social care (top 20 phrases only)** – child welfare, child abuse, foster care, child(hood) maltreatment, child sexual abuse, child protection, out-of-home care, youth care, child neglect, foster parent, residential care, out-of-home placement, foster child, child sexual behavior[2] inventory, family reunification, adverse childhood experience, care placement, foster family, foster youth, child physical abuse

- **Systematic review (all phrases identified)** – systematic review, literature review, systematic literature review, rapid review, narrative review, integrative review, comprehensive review, scoping review, meta analysis, meta-analytic review, systematic critical review

For a specific intervention such as IFPS, a more precise set of search terms could be identified. However, an analysis of the 142 titles and abstracts identified as potentially relevant from the IFPS search showed that a range of phrases would need to be used to provide a comprehensive identification of relevant research papers.

The most common phrase used in the literature was 'intensive family preservation' but 'family preservation' was used almost as frequently, either alone or combined with service, services, program(me), program(me)s or intervention. More rarely, phrases such as 'family intervention project' were used as well as specific programme names such as Homebuilder(s) and Family First.

Based on these two case studies, text analysis can help identify additional search terms to increase sensitivity but the use of a large number of search terms is likely to lead to a reduction in precision and a much greater screening burden. Researchers in this field will have to adopt a trade-off; developing a search that includes the most common words and phrases used by authors, and seeking a balance of sensitivity and precision as described above[3]. Supplementary searching (such as checking reference lists, citation tracking [see section 3], website searching and author contacts) can be used to fill in gaps in identification.

Summary findings: Text analysis for search strategy improvement

1. The analysis of text words contained in titles and abstracts of relevant papers can help identify search terms for use in database searching.
2. There is variation in author language so an extensive set of search terms is likely to be needed to identify the majority of research on a particular topic.
3. Increasing sensitivity (recall of relevant studies) decreases precision (the ability of the search to exclude irrelevant studies) so the balance may need to be a trade-off between the two, using supplementary search methods to maximise the identification of relevant studies.

2 The American spelling only was identified by the software

3 The SURE based team aim to adopt a search strategy where 5-10% of the abstracts appear to have some relevance to the review topic and which identifies 80-90% of a set of relevant publications which were set aside at the outset and not used to develop the search strategy.

2 CHOICE OF DATABASES AND OTHER INFORMATION SOURCES

For any given review topic, it is rare that a single database or information source will include all the relevant studies. A sensitive search strategy, with a comprehensive set of key words and phrases, may identify the vast majority of studies in that particular database but other studies will be missed that were not indexed within that source. Thus, there is a need to include both a comprehensive set of search terms and a range of information sources to ensure a good coverage of a topic area. This should include both traditional academic sources such as journals and books, as well as grey literature[4] such as reports, working papers and government documents.

2.1 Information sources: Methods

For the two case studies explored above, the number of studies indexed in the large subscription databases Scopus and Web of Science were investigated since these have good coverage of the social science literature. The free-to-use database PubMed was included as a comparator. Microsoft Academic[5] was also explored. This is a huge and relatively new free-to-use database which allows direct download of publication data into bibliographic software.

2.2 Information sources: Findings

For the case study looking for systematic reviews in children's social care an extensive search of 19 databases and additional search methods identified 86 systematic reviews. The percentage of these reviews not included in the large subscription databases Scopus and Web of Science was quite high (Table 1.1). PubMed is a specialised biomedical database, and its coverage of social care is limited though it indexed seven reviews not included in Scopus or Web of Knowledge, whose coverage was largely similar. Microsoft Academic had good coverage for papers within this case study and contained every publication indexed in each of the other databases.

For the IFPS review, 33 studies were identified from a search of 12 databases and additional search methods (Bezeczky 2019). The number of research papers not included in the databases explored was even greater (Table 1.2). The two papers contained in PubMed were indexed in all three of the other databases. As with the review of reviews, Microsoft Academic contained every publication indexed in each of the other databases but 45% of the papers included in the final review were not contained in this source.

Table 1.1. Indexing of reviews in children's social care within major bibliometric databases

	Scopus	Web of Science	PubMed	Microsoft Academic
All systematic reviews of relevance to children's social care	73% [63/86]	72% [62/86]	53% [44/86]	93% [80/86]

4 Grey literature is materials and research produced by organizations outside of the traditional commercial or academic publishing and distribution channels [Wikipedia]

5 Microsoft Academic, is a rival to Google Scholar in terms of 'free to use' academic databases. In May 2020 it contained over 234 million publication records. This compares to 70 million records in Scopus, 76 million in the Web of Science core collection and 30 million in PubMed. Google Scholar does not publish its coverage but it was estimated as 389 million in 2018.

Table 1.2. Indexing of primary research studies of an IFPS intervention within major bibliometric databases

	Scopus	Web of Science	PubMed	Microsoft Academic
Research papers looking at the impact on IFPS on care entry	30% [10/33]	36% [12/33]	6% [2/33]	55% [18/33]

While it is clear that Microsoft Academic is superior in coverage to the very large and established subscription databases like Scopus and Web of Science its search functionality was limited at the time that this research was carried out (June - September 2020). The complex multi-line searching and Boolean logic (AND/OR/NOT) commonly used by systematic reviewers is not yet feasible other than via the website lens.org or EPPI-Reviewer which both provide a Boolean logic search platform for Microsoft Academic. The descriptive (meta) data on papers downloaded from Microsoft Academic for use in reference management software, via the use of the RIS file format, also lacked detail in some cases.

The papers not indexed within the databases explored above were often (although not always) published in the grey literature. In the review of systematic reviews 14 of the 86 reviews were identified from the grey literature (16%) and figures for the IFPS review were 21 of 33 studies (64%).

Comparing these findings with other reviews of primary research within this field, 22 of 38 studies (58%) on 'signs of safety' (Sheehan et al, 2018) and 10 of 33 studies (30%) on shared decision making (Nurmatov et al, 2020) were from the grey literature.

In conclusion, an extensive search, including a number of databases and supplementary search methods, is needed for a sensitive search of the research literature in children's social care. Routine checking of reference lists in included studies, citation tracking (see section 3) and website searching can be valuable ways of increasing the identification of grey literature as well as other publications.

Some recommended databases and websites for searching for studies of relevance to children's social care are listed the Appendix.

Summary findings: Information sources

1. A comprehensive multi-database search and supplementary search methods is needed to identify, with high sensitivity, research studies in children's social care.
2. Many relevant studies within this field are published as reports or other grey literature

3 CITATION ANALYSIS

A supplementary search method adopted by many reviewers is citation analysis (Belter 2016, Janssens 2015, Sarol 2018) of relevant studies to seek other potentially relevant papers. This can be used to look both forwards in time, for newer papers that cite relevant papers in their reference lists; and also, backwards in time, by looking at the reference lists of the original set of relevant papers. In the context of this research, the capacity of citation tracking forwards only was considered to explore its value for updating an existing literature search. The databases Scopus, Web of Science, Microsoft Academic and Google Scholar all offer the ability to identify newer studies through their citation of older ones.

3.1 Citation analysis: Methods

An exploration of the efficiency of citation analysis using Scopus, Web of Science and Microsoft Academic was carried out for its ability to identify relevant studies within the two case studies, and the utility of each tool for this type of work. Google Scholar was excluded from the study since, although citations are associated with publications, these cannot be readily downloaded into reference management software for analysis.

In the first case study (the review of reviews), 67 systematic reviews were selected for summary within What Works' Evidence Store[6] from a search carried out in October 2018 using traditional methods. An update search for the Evidence Store in November 2019 resulted in the identification of a further 19 systematic reviews. Using citation tracking from the 67 reviews, we explored how many of the 19 newer systematic reviews could be identified because they had referenced one or more of the 67 reviews identified in the earlier search.

In the second study (the review of IFPS), citation tracking was carried out using the 17 research papers published up to 1992 to see how many of the 16 papers published from 1993 onwards could be identified because they had referenced one or more of the earlier research papers.

3.2 Citation analysis: Findings

For the review of reviews, results are summarised in Table 3.1.

From this first case study it appears that citation analysis in Microsoft Academic/Scopus/WoS can identify new research but with relatively low sensitivity because many of the newer reviews

Table 3.1. Citation tracking of systematic reviews in children's social care to identity newer reviews

Papers from update search	Included in Microsoft Academic	Included in Scopus	Included in Web of Science	Found in Microsoft Academic	Found in Scopus	Found in Web of Science
N=19	N=17	N=16	N=16	21%[7] [4/19] 24%[8] [4/17]	16% [3/19] 19% [3/16}	16% [3/19] 19% [3/16]

6 https://whatworks-csc.org.uk/wp-content/uploads/SYSTEMATIC-REVIEWS.pdf
7 % of studies found
8 % of studies found that were indexed in the database and could theoretically have been identified

were not referencing older ones. Microsoft Academic contained 17 of the 19 newer papers but only four of these were identified by citation tracking compared to Scopus/WoS where each database contained 16 of the 19 newer papers of which only 3 referenced one or more of the older reviews.

Since this review (of reviews) looked at all reviews in children's social care and covered a wide range of topic areas, it was not a surprising finding that the referencing of older reviews by newer ones was not comprehensive.

Thus the second case study looking at IFPS provided a more appropriate assessment of the potential utility of this technique. Results are summarised in Table 3.2.

As with case study 1, citation tracking/analysis in Microsoft Academic/Scopus/WoS identified some of the later research but with relatively low sensitivity. Microsoft Academic contains more of the relevant studies but citation tracking was of similar effectiveness to Scopus/WoS.

Table 3.2. Citation tracking of primary research studies in Intensive Family Preservation Services (IFPS) to identify newer relevant research

Papers published ≥ 1993	Included in Microsoft Academic	Included in Scopus	Included in Web of Science	Found in Microsoft Academic	Found in Scopus	Found in Web of Science
N=16	N=11	N=7	N=8	13%[7] [2/16] 18%[8] [2/11]	13% [2/16] 29% [2/7]	13% [2/16] 25% [2/8]

The additional screening load is fairly small, suggesting that the routine use of citation analysis as a method of ensuring a highly sensitive search does not present a huge additional screening load for Scopus and Web of Science, where citations can be rapidly downloaded for a large dataset. In Microsoft Academic (MA) this takes longer since it needs to be done manually, paper by paper.

For the IFPS review, following deduplication, there were 211 citations in Microsoft Academic, 54 in Scopus and 51 in WoS. When all citations from the three sources were combined together, there were 191 records in total of which 150 were published from 1993 onwards.

From these 150 records, the majority were relevant to IFPS in general (e.g. review articles, qualitative studies and intervention studies) but without the controlled design and placement outcome that were inclusion criteria for the review. Three of the included studies were found – a number needed to screen of 50[9]. An analysis of the 150 abstracts did not identify further primary research studies that might have been included in the review. It is difficult to assess why the sensitivity of this method is relatively low for such a tightly defined topic area but, anecdotally, the primary studies did not always include extensive reference lists covering the existing research base.

Summary findings: Citation analysis

1. For the two case studies, forward citation tracking (analysis) can identify new relevant research, but with low sensitivity.
2. Citations can be readily and rapidly downloaded from Scopus and Web of Science. For Microsoft Academic this is a manual process, paper by paper, so is not as yet practical as a rapid search strategy.
3. The additional screening load resulting from citation tracking is not huge so it might be considered as a supplementary, but not a replacement, search strategy.

9 This compares well with the traditional search for the review, where 1948 abstracts were screened to find 33 relevant studies. A number needed to screen of 59.

4 CO-CITATION ANALYSIS

A developing search method with much theoretical promise is co-citation analysis. Typically, two to four relevant papers are used as seed papers to identify other papers that are closely linked though 'co-citations' where papers share at least 10% of references (Belter 2016, Janssens 2020). This method results in a web of forwards- and backwards-in-time searching. It should lead to a more precise identification of relevant research, in comparison with citation analysis, given the percentage of shared citations required. In the context of this research, its value was explored for the ability to identify newer publications for updating an existing literature search.

4.1 Co-citation analysis: Methods

There are relatively few tools offering co-citation analysis that are ready for use by librarians and information professionals without specialist IT knowledge. However, this is a growth area. Examples are the 'related records' function in Web of Science, that uses co-citations to relevance rank the records identified and CoCites (cocites.com) for use with PubMed.

Given their greater subject coverage of children's social care research, co-citation was explored in Microsoft Academic (MA), using a script developed by one of the authors (JT) and in Web of Science using the related records function.

4.2 Co-citation analysis: Findings

For each of the case studies, a sample of three publications (two journal articles and one grey literature report) were chosen as 'seed' papers. Three different seed papers were chosen, within each case study, for each of the methods explored.

For the review of systematic reviews, none of the reviews included from the update search (the target reviews) were identified by the co-citation script in Microsoft Academic. The related records function in Web of Science yielded over 40,000 records across the three seed papers. Ranking the related records by relevance (i.e. ranked according to the number of references shared with the parent record[10]) and checking the first 50 records from each seed paper identified three of the target papers.

For the IFPS review, only one of the papers included from the update search was identified using the co-citation script in Microsoft Academic. The related records function in Web of Science did not identify any of the target papers.

On the basis of the two case studies, co-citation analysis does not, as yet, show a great deal of promise within this research area. A citation network is a fragile system that works in some contexts and not others. It depends on the culture of citing and being cited within a particular research discipline and whether the seed papers are located firmly within the citation web. These are very preliminary findings however and further research is indicated to explore the effect of different seed papers and the percentage of citations that need to be shared. We would regard this as a development area for information searching in the field.

Summary findings: Co-citation analysis

1. Some co-citation tools are now available that can readily be used by information professionals without IT expertise
2. Very preliminary results suggest that co-citation analysis has little value for the case study topics explored but these minimal results should not guide practice and this is a potential development area

10 Related records in Web of Science are documents that cite at least one document cited by the parent record. These are ranked according to the number of references they share with the parent record. https://images-webofknowledge-com.abc.cardiff.ac.uk/images/help/WOS/hp_related_records.html

5 RANKING OF PAPERS FOR RELEVANCE USING MACHINE LEARNING FOR SCREENING

A comprehensive and sensitive search can yield a very large number of records that then need to be screened to identify studies for potential inclusion in a review.

A number of free and subscription tools exist to support the title and abstract screening stage of the systematic review process which include a classification/ranking procedure based on machine learning. These all essentially involve a computer algorithm that learns from the in/out decisions of the screener to analyse words and multi-word phrases to boost the not yet reviewed but 'likely to be relevant' abstracts to the top of the set (Gates 2019, Olofsson 2017, O'Mara-Eves 2015, Thomas 2017, Tsou 2020).

The software can be used in two ways: (1) The operator goes through the abstracts selecting those for include or exclude until such point as the classifier has enough information to begin classifying the remaining abstracts and boosting those more likely to be relevant, to the top of the list; (2) The classifier is pre-trained with a set of included and excluded abstracts identified via the traditional method where two independent researchers review each abstract and discuss any abstracts where there is disagreement (resorting to a third reviewer if needs be). It is then asked to rank a new set of abstracts based on a model developed at the training stage.

Early studies within this project to look at one of these tools, the Relrank machine learning algorithm in Rayyan suggested that this technique can help boost the vast majority of relevant studies in children's social care to the top of a list of titles/abstracts found from searching in this topic area; in particular if the tool was pre-trained as above.

From a literature review carried out to develop the methodology for this study (PL), four tools gained good reviews: Rayyan, Abstrackr, Robot Analyst and EPPI-reviewer. The first three tools are free at the point of use while EPPI-reviewer is a subscription service, but performed particularly well in the research studies examined.

Thus EPPI-reviewer was chosen for a detailed look at the ability of machine learning to assist with the ranking of studies identified from an updated search within the two case study reviews (the second use case considered above).

5.1 Machine learning for screening: Methods

For each of the case studies, sets of title and abstracts that had been included and excluded at title/abstract stage were used to train the classifier in EPPI-reviewer. The most basic of options for the classifier was chosen for both case studies (EXCLUDE on evidence)[11].

For the review of reviews case study, 4,302 abstracts in all were used to train the classifier of which 3,786 had been excluded and 516 included at title and abstract stage. The trained classifier was then run against the test set of 1,481 titles and abstracts from the review update search and the ranking of the final set of 29 included systematic reviews (identified via the traditional two-person manual screening method) was analysed.

11 A second classifier was also developed by grouping excluded studies into EXCLUDE on target group (not children's social care) and EXCLUDE on intervention (not systematic review) to provide more information for the classifier to learn from. This yielded slightly improved results. Mean score = 60 (standard deviation 18) (range 25-87).

For the IFPS case study a set of 1,784 records comprised the 'EXCLUDE on evidence' training set while 82 comprised the 'INCLUDE at title/ abstract' set. The test set comprised 1,740 records.

5.2 Machine learning for screening: Findings

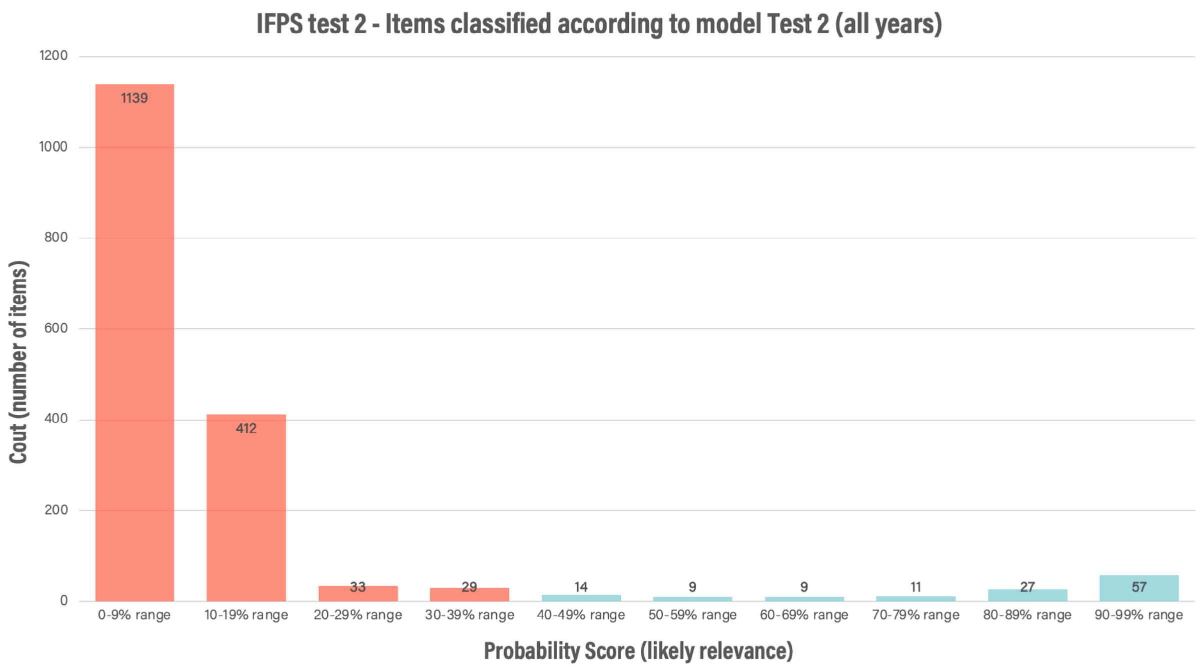
For the review of reviews study, the training of the classifier boosted the majority of the relevant abstracts from the test set into the top third of the screening set. The range of probability scores (which roughly equate to the probability of a paper being included in a review) for the 29 included studies was 23-86 with a mean score of 57 (standard deviation 18). 28 (97% of the studies) scored 32 or more (Figure 5.1). All the included studies fell within the range of probability scores identified by the blue (paler) bars in the Figure.

Thus, all but one of the studies was within the top 28% of the results ranked by the classifier. If the operator screened only the first 28% of the abstracts, this would represent a saving of 72% of the screening burden. To include the outlier and identify 100% of the final includes would have required the screening to 44% of the total.

For the IFPS study the trained classifier boosted the majority of relevant abstracts into the top half of the screening set. The mean probability score was 83 (standard deviation 15) with a range of scores from 44-98 (Figure 5.2). All the included studies fell within the probability scores identified by the blue (paler) bars. This much better performance is not surprising given the more precise language used to describe this intervention (see section 1).

Figure 5.1. Probability scores (likely relevance) of systematic review records found by the search and ranked by the EPPI Reviewer classifier

Figure 5.2. Probability scores (likely relevance) of IFPS records found by the search and ranked by the EPPI Reviewer classifier



In conclusion, there is a clear indication that this may be a valuable technique when screening to update a review in children's social care.

Further studies of other topic areas would be needed to develop confirm this but, based on these two case studies, it is possible that papers with a probability score of, say, less than 20 (to allow for a large margin) might be routinely excluded from screening when a classifier had been pre-trained in the way described above. As can be seen from Figures 5.1 and 5.2 above, this would represent a considerable saving on the screening time required even with this very conservative cut off point; Approximately 53% of abstracts would need to be screened in the case of the review of reviews but less than 10% in the case of IFPS.

If shown to apply to other topics and for the case when the classifier learns as it goes along from reviewer decisions rather than being pre-trained, this is a very promising technique which could reduce the need for a sensitivity and precision trade-off on the part of the reviewer (see Section 1).

Summary findings: Ranking of papers using machine learning to speed up study identification from abstracts

1. Machine learning for screening shows promise for application to update searches for reviews, when the machine learning software is 'pre-trained' with included and excluded abstracts

RESEARCH GAPS

These case studies have identified the need for a comprehensive search strategy across a wide range of information sources when searching for research information in relation to children's social care. Results also indicate that machine learning is a promising technique for use in screening the records found from the search.

The findings are from two case studies only. Further research would be valuable, using other case studies, to see if these findings are generalizable to other review updates, and new review topics, in this research area.

Specific areas of work are also indicated:

- Additional work on the list of databases and websites identified to see if there are some sources that contain only those studies readily identified elsewhere and, thus, do not need to be routinely considered by reviewers.
- Co-citation analysis was only explored briefly and further research would be of value as more co-citation tools become available for use by information searchers without the need for specialist IT skills.
- The use of machine learning within the manual screening process (as described in the introduction to Section 5) as opposed to using a pre-trained classifier, which was the method explored in this research study.
- The potential value of machine learning in assisting with the screening of large search sets to reduce the need for the trade-off between a sensitive and precise search (as described in Section 1).

REFERENCES

Belter CW. (2016) Citation analysis as a literature search method for systematic reviews. JASIST Journal of the Association for Information Science and Technology 67(11): 2766-2777 https://doi.org/10.1002/asi.23605

Bezeczky Z, El-Banna A, Kemp A, Scourfield J, Forrester D, Nurmatov U. (2019) Intensive Family Preservation Services to prevent out-of-home placement of children: a systematic review and meta-analysis. London: What Works for Children's Social Care. https://whatworks-csc.org.uk/wp-content/uploads/WWCSC-Intensive-Family-Preservation-Services-to-prevent-out-of-home-placement-of-children-v2.pdf

Borah R, Brown AW, Capers PL, Kaiser KA. (2017) Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7:e012545. http://dx.doi/org/10.1136/bmjopen-2016-012545

Gates A, Johnson C, Hartling L. (2019) Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Systematic Reviews 8(1): 278 http://dx.doi/org/10.1186/s13643-019-1222-2

Janssens ACJW, Gwinn M. (2015) Novel citation-based search method for scientific literature: application to meta-analyses. BMC Medical Research Methodology 15: 84 https://doi-org.abc.cardiff.ac.uk/10.1186/s12874-015-0077-z

Janssens A, Cecile JW, Gwinn M, Elaine J et al. (2020) Novel citation-based search method for scientific literature: a validation study. BMC Medical Research Methodology 20(1): 25 https://doi.org/10.1186/s12874-020-0907-5

Kugley S, Wade A, Thomas J, Mahood Q, Klint Jørgensen A-M, Hammerstrøm K, Sathe N. (2017) Searching for studies: a guide to information retrieval for Campbell systematic reviews. Campbell Systematic Reviews 13(1): 1-73 http://dx.doi.org/10.4073/cmg.2016.1

Olofsson H, Brolund A, Hellberg C Silverstein R, Stenström K, Österberg M, Degerhamn J. (2017) Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. Research Synthesis Methods 8(3):275-280 http://doi-org.abc.cardiff.ac.uk/10.1002/jrsm.1237

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 http://doi-org.abc.cardiff.ac.uk/10.1186/2046-4053-4-5

Nurmatov BU, Foster C, Bezeczky Z, Owen J, El-Banna A, Mann M, Petrou S, Kemp A, Scourfield J, Forrester D, Turley R. (2020) Impact of shared decision-making family meetings on children's out-of-home care, family empowerment and satisfaction: A systematic review. London: What Works for Children's Social Care. https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_Family_Group_Conferencing_Report.pdf

Sarol MJ, Liu L, Schneider J. (2018) Testing a citation and text-based framework for retrieving publications for literature reviews. BIR 2018. 7th International Workshop on Bibliometric-enhanced Information Retrieval, Grenoble, France March 2018.

Sheehan, L., O'Donnell, C., Brand, S.L., Forrester, D., Addis, S., El-Banna, A., Kemp, A. and Nurmatov, U. (2018) Signs of Safety: Findings from a mixed-methods systematic review focussed on reducing the need for children to be in care. London: What Works for Children's Social Care. https://whatworks-csc.org.uk/wp content/uploads/Signs_of_Safety_a_mixed_methods_systematic_review.pdf

Stansfield C, O'Mara-Eves A, Thomas J. (2017) Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Research Synthesis Methods (May): 1–11 https://doi.org/10.1002/jrsm.1250

Thelwall, M. (2018) Microsoft Academic automatic document searches: accuracy for journal articles and suitability for citation analysis. Journal of Infometrics, 12(1): 1-7. http://dx.doi.org/10.1016/j.joi.2017.11.001

Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C et al, on behalf of the Living Systematic Review Network. (2017) Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology 91: 31-37 http://dx.doi.org/10.1016/j.jclinepi.2017.08.011

Tsou AY, Treadwell JR, Erinoff E et al. (2020) Machine learning for screening prioritization in systematic review: comparative performance of Abstrackr and EPPI-Reviewer. Systematic Reviews 9(1): 73 http://dx.doi.org/10.1186/s13643-020-01324-7

Wang Z, Neyfeh T, Tetzlaff J, O'Blenis P, Hassn Murad M. (2020) Error rates of human reviewers during abstract screening in systematic reviews. PLOS One 15(1): e0227742 https://doi.org/10.1371/journal.pone.0227742

APPENDIX

Some databases and websites to consider for the identification of research literature (including grey literature) in children's social care

Note: These are broadly ranked in order of their value based on the team's experience of searching across a wide range of topics within children's social care and expert guidance. The yield of relevant papers specific to individual topic areas and ability of each source to pick up 'unique' studies, not picked up by others, has not been analysed. Only with this additional work across a range of topic areas would it be possible to establish a ranking order and identify sources that might be excluded. Sources that are free at the point of use are listed in bold italics. URLs were checked on 18 August 2020.

Multi-topic databases
Web of Science
Scopus
Microsoft Academic[12]
Embase
Medline/*PubMed*
PsycINFO
CINAHL
Subject Specific databases
Social Policy and Practice
Child Development and Adolescent Studies
ASSIA
ERIC
Sociological Abstracts
International Bibliography of the Social Sciences (IBSS)
British Education Index

12 Some shortcomings for systematic reviewers. Complex Boolean logic search not feasible (other than via lens.org); results from citation tracking can only be downloaded on a paper by paper basis and some missing data when search results are downloaded into reference management software.

Grey literature (including websites)
Health Management Information Consortium (HMIC)
OpenGrey
Department for Education including Children's Social Care Innovations Programme
Child Welfare Information Gateway
REES Centre
National Institute for Health and Care Excellence - *Evidence Search* (includes Social Care Institute for Excellence [SCIE] search)
NSPCC Learning Library Catalogue
Joseph Rowntree Foundation
What Works for Children's Social Care
Nuffield Family Justice Observatory - Resources
Action for Children
Early Intervention Foundation
Barnardo's
Children's Commissioners (England, Northern Ireland, Scotland, Wales)
Coram
Care Leavers' Association
Children's Society
Anna Freud Centre[13]
Economic data
Econlit
NHS Economic Evaluation Database (NHS EED)
IDEAS/RePEc

13 Registration required and can only sign up if a member of a LA or organisation that is a partner of Research in Practice

info@whatworks-csc.org.uk
@whatworksCSC
whatworks-csc.org.uk