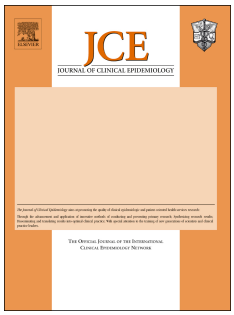
Journal Pre-proof



Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study.

Chris Cooper, Tristan Snowsill, Christine Worsley, Amanda Prowse, Alison O'Mara-Eves, Helen Greenwood, Emma Boulton, Amanda Strickson

PII: S0895-4356(19)31063-7

DOI: https://doi.org/10.1016/j.jclinepi.2020.03.013

Reference: JCE 10099

To appear in: *Journal of Clinical Epidemiology*

Received Date: 22 November 2019

Revised Date: 2 March 2020

Accepted Date: 19 March 2020

Please cite this article as: Cooper C, Snowsill T, Worsley C, Prowse A, O'Mara-Eves A, Greenwood H, Boulton E, Strickson A, Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study. *Journal of Clinical Epidemiology* (2020), doi: https://doi.org/10.1016/j.jclinepi.2020.03.013.

Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study.

Chris Cooper [1], Tristan Snowsill [2], Christine Worsley [3], Amanda Prowse [3], Alison O'Mara-Eves [4], Helen Greenwood [5], Emma Boulton [3], Amanda Strickson [3].

[1] Department of Clinical, Educational and Health Psychology, University College London, London WC1E 7HB, UK
[2] Health Economics Group, University of Exeter Medical School, University of Exeter, Exeter, Devon, United Kingdom
[3] Tolley Health Economics, Buxton, UK
[4] EPPi Centre, University College of London, London, UK
[5] Royal College of Psychiatrists, London, UK

Correspondence to:

Mr. Chris Cooper
University College London
1 – 19 Torrington Place
London WC1E 7HB
United Kingdom
Phone: not available
Fax: not available
E-mail: ucjucc4@ucl.ac.uk

Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study.

Abstract

Objective: To compare the effectiveness and efficiency of methods used to identify and export conference abstracts into a bibliographic management tool.

Study design and setting: Case study. The effectiveness and efficiency of methods to identify and export conference abstracts presented at the American Society of Hematology (ASH) conference 2016-2018 for a systematic review were evaluated.

A reference standard handsearch of conference proceedings was compared to: 1) contacting *Blood* (the journal who report ASH proceedings); 2) keyword searching; 3) searching Embase; 4) searching MEDLINE via EndNote; and 5) searching CPCI-S. Effectiveness was determined by the number of abstracts identified compared with the reference standard, while efficiency was a comparison between the resources required to identify and export conference abstracts compared to the reference standard.

Results: 604 potentially eligible and 15 confirmed eligible conference abstracts (abstracts included in the review) were identified by the handsearch. Comparator 2 was the only method to identify all abstracts and it was more efficient than the reference standard. Comparators 1, and 3-5 missed a number of eligible abstracts.

Conclusion: This study raises potentially concerning questions about searching for conferences' abstracts by methods other than directly searching the original conference proceedings. Efficiency of exporting would be improved if journals permitted bulk downloads.

Background

Searching for reports of studies presented at a conference is an acknowledged approach to study identification in systematic reviews (1-6). Guidance suggests that searching conferences may identify newly emerging studies, or updated findings of on-going studies, potentially ahead of journal publication (2, 3, 7, 8) and that identifying and including conference abstracts may help minimise the introduction of bias into systematic reviews (2, 4, 9). There is some evidence that searching conferences is an effective method of identifying studies which might be missed by other search methods and identifying studies that are reported at conferences but never published (4, 5, 10-13).

Handsearching has traditionally been the method used to search for reports of studies presented at conferences (6, 25, 26). Handsearching involves a manual, page-by-page, examination of the entire contents of relevant journals, conference proceedings and abstracts (2, 4, 7, 9, 14-16). There is evidence that handsearching is effective when compared to bibliographic database searching and that handsearching can identify studies (or study data) which may be missed by other search methods (4, 5, 7, 13-15, 17-24). Whilst handsearching is known to be an effective method of study identification, it is resource intensive (5).

When handsearching conference proceedings presented at the American Society of Hematology (ASH) conference (2016-2018) for a systematic review (25), 604 reports of potentially eligible abstracts were identified by a handsearch but there was no option to export all 604 records to a bibliographic management tool in one export. Instead, each of the 604 abstracts had to be individually identified and downloaded one-by-one. This added to the resources required to complete the handsearch of conference proceedings.

The inability to download all of the 604 potentially eligible abstracts at the same time, as is possible in bibliographic databases (where individual studies or a range of studies can be selected for export), motivated the question: what is the most efficient way to export abstracts identified by handsearching conference proceedings into a bibliographic management tool for further screening? The research team hypothesised potential alternative methods (henceforth comparators) which could lead to an efficient and successful export of abstracts into a bibliographic

management tool. This case study reports the evaluation of these comparators compared to the handsearch.

It is not a straight-forward evaluation to report. When the comparators were tested, it became apparent that, for some methods, the identification of abstracts could not be isolated from the task of exporting abstracts. As such, the research objectives became broader than the problem of exporting conference abstracts to include a focus on the effective identification of conference abstracts reported at ASH.

Study objectives

This case study aims to evaluate the effectiveness and the efficiency of methods to identify and download eligible conference abstracts reported at ASH 2016-2018 for a systematic review of intervention effectiveness. The research objectives of this case study are:

1. to determine whether there is a more efficient method for downloading eligible conference abstracts following a handsearch compared to the current technology (i.e., individually downloading records);
2. to evaluate the effectiveness of comparator methods to identify the same abstracts found by the reference standard handsearch across two stages of study identification ('potentially eligible' and 'confirmed eligible'); and
3. to evaluate the efficiency of the various methods across two stages of study identification ('potentially eligible' and 'confirmed eligible').

Methods

Study design

A case study based on a systematic review is presented (25, 26). This case study was designed as a comparison between reference standard and comparators. The details of the reference standard and comparators are set out below alongside the methods of analysis.

Data

Data were conference proceedings reported at ASH 2016-2018 published in the supplement editions of the journal *Blood*. The editorial team at *Blood* confirmed that 17,759 conference abstracts were reported at ASH for this period. The reference standard handsearch identified 604 abstracts as potentially eligible for further screening and 15 abstracts were confirmed eligible for inclusion in the systematic review based on PICOS eligibility criteria and on the basis of independent double-screening. The 17,759 total eligible, 604 potentially eligible, and 15 confirmed eligible abstracts, represent data for this case study.

The reference standard

The reference standard is a method derived from recommended best practice guidance. A handsearch of the ASH conference proceedings was undertaken by one experienced reviewer (CC). The reviewer handsearched the supplement editions of the 2016-2018 ASH conference proceedings reported in the journal *Blood* and available from: http://www.bloodjournal.org/blood/search-results?f_ArticleTypeDisplayName=Meeting+Report

The reviewer handsearched on screen, page-by-page looking for any abstracts reporting the interventions reported in Figure 1, or any potential alternative references to these interventions, or possible mis-spellings (2, 4, 14, 15). Records of any additional search terms to those recorded in Figure 1 were kept and then a keyword search was undertaken using the search function on the journal website (see journal search function below for detail) to cross-check the handsearch in event of human error.

Figure 1: The search terms for this study

Syntax
Pevonedistat
MLN4924
Decitabine
Dacogen

Azacitidine
Vidaza

Comparators

Comparator 1: contacting the journal directly to request exports of the identified records

The editorial team of the journal *Blood* were contacted by e-mail to ask if they could download the 604 potentially eligible abstracts from their internal server. This is a very different comparator method compared to the other four in two ways. First, it does not include a search aspect and only taps into the 'download/export' aspect of study retrieval. Secondly, it is probable that this comparator method would have an all-or-nothing outcome: either the journal staff would send all 604 records, or they would not send any. Despite these differences, this comparator method was included because, if successful, the approach represents an efficient way to circumvent the individual download problem that was the original motivation for this work and thereby address objective 1. However, because it is fundamentally different to the other comparator methods, it was evaluated separately.

Comparator 2: the search function on the journal website

The journal *Blood* includes a search function where the supplement edition of a conference can be keyword searched. This keyword search was utilised in the reference standard, to ensure completeness of the handsearch in the event of human error, but it represented a way to identify the same 604 potentially eligible abstracts for export into a bibliographic screening tool.

The terms in Figure 1 were searched one-by-one and the abstracts that were identified were downloaded study-by-study to EndNote using the direct export function on the journal website. Further detail on this method is presented in the web-only material.

Comparator 3: identifying the specific journal in Embase and searching for abstracts

Embase was chosen over the bibliographic database MEDLINE due to its inclusion of conference proceedings and material (27). The terms for the interventions, and

associated Emtree controlled indexing, were searched in Embase using the Ovid interface. This search was limited by publication type to conferences in two ways:

First, controlled indexing and search fields were searched for abstracts indexed by publication type (line 1 below) and the ASH conference was searched using relevant field codes, namely: cf = conference information and cg = conference publication (line 2 below).

1. exp conference paper/
2. ash.cf,cg.
3. 1 or 2

Secondly, the journal *Blood* was searched for using the journal field code (jn) and the abstracts returned were combined with a search for conference.af. (af = all fields).

These two searches were combined using the Boolean connector “OR” so both approaches to limiting by publication type were included. The full search syntax, including a search narrative, is presented in web-only material (28),.

Comparator 4: a search for the journal Blood was made in PubMed in EndNote
The search terms in Figure 1 were searched using the online search function of EndNote X8. The following search logic was applied:

Journal – contains – Blood; AND
Year – contains – 2016*; AND
All fields – contains – the intervention terms in Figure 1**.

* 2016 was searched first, then 2017 and finally 2018. ** the intervention terms were searched one at a time.

Abstracts were visually inspected and manually de-duplicated. Study records which reported conference proceedings were retained whilst other journal content (i.e.

abstracts not reported at the ASH conference) were deleted. The search strategy is reported in web-only material.

Comparator 5: searching a conference proceedings database (CPCI-S)

A search was undertaken in Conference Proceedings Citation Index- Science (CPCI-S), Web of Science (Clarivate Analytics). The search terms in Figure 1 were searched on the topic search field and search terms for ASH or: (American-Society-of-Hematology) were searched on the conference search field. Searches were refined to the years 2016, 2017 or 2018. The search strategy is reported in web-only material.

Analysis

Outcomes were recorded at two stages in the study identification process:

(stage 1) 'potentially eligible' abstracts were identified on the basis of title or abstracts and the study record was retrieved for further inspection; and
(stage 2) 'confirmed eligible' abstracts were identified on the basis of screening the abstract to confirm eligibility and inclusion in the systematic review.

For stage 1, the reference standard handsearch and comparator 2 (journal search function, see below) were undertaken in the week commencing February 4th, 2018. Abstracts were identified and individually (i.e. study-by-study) downloaded to EndNote using the direct export function on the journal website. Google Chrome (version 76.0.3809.132) was the web browser. Comparators 3-5 were undertaken on June 20th, 2019. The search details are reported in web-only material.

For stage 2, the 604 abstracts identified in the reference standard were downloaded to EndNote and were independently screened by two experienced reviewers (CW and AP). A third experienced reviewer (AS) was available in the event of disagreements.

Outcome measurement

The following outcomes were recorded for the reference standard and comparator methods:

Number of potentially eligible abstracts (stage 1)

The reference standard identified 604 potentially eligible abstracts which were taken forward for independent double-screening against predetermined inclusion criteria (25). The number of abstracts identified by each of the comparator methods deemed potentially eligible by the reference standard were recorded.

Number of abstracts fulfilling inclusion criteria for the systematic review (stage 2)

The number of abstracts identified from the reference standard as confirmed eligible was 15. This represents the final point of comparison where the ability of the comparators to identify these same 15 abstracts is compared.

Time

Time was recorded using the stopwatch function on an Apple iPhone 6s. Time was recorded in minutes.

Cost

Cost was represented as GBP since this study was undertaken in the UK. An approach similar to Shemilt *et al.* was followed to identify local unit costs (29). A mid-point Grade 7 cost (spine point 40) was chosen, since this represents the median pay of the grade of researcher who might usually undertake the work reported. University College London salaries and on-costs (2018-2019) were used since this represents the lead author's home institution and this was the year the case study was undertaken. These costs included salary, direct salary costs (e.g. pension) and university indirect costs. Similar to Shemilt *et al*. the costs included ‘London Weighting’ which is an uplift provided to staff to cover additional costs of London. The hourly rate used was £31.38.

Evaluation metrics

Metrics were calculated at both stage 1 (handsearching of 'potentially eligible’ abstracts) and stage 2 (screening 'confirmed eligible' abstracts). What constitutes an effective, efficient or comprehensive literature search is uncertain (30-32). In this study, the following understandings are used (12, 30).

Effectiveness

Effectiveness was determined by comparison with the reference standard handsearch. Two by two tables were created (reported in web only material) and the following metrics were calculated to compare effectiveness:

- Recall (proportion of correctly identified abstracts);
- Precision (proportion of correctly identified abstracts out of all studies retrieved by the comparator); and
- F-Measure (a harmonic mean was used). The F1-measure is the harmonic mean of precision and recall; it has no specific weighting towards either, but will generally be closer to the lower of the two. It is the rate of true positives with respect to the arithmetic mean of TP+FP and TP+FN (the denominators for precision and recall respectively) (30, 33).

Efficiency

Efficiency was the comparison in resources between the reference standard handsearch and comparator methods, this was calculated as follows:

- Difference in time taken; and
- Difference in cost of time taken.

Findings

Objective 1 – efficiency of downloading the handsearch

The first study objective was to determine whether there is a more efficient method for exporting potentially eligible abstracts compared to the current technology (individually exporting abstracts). *Blood's* editorial team were contacted to enquire if they could send the 604 potentially eligible records to the research team. All other comparators could not isolate the export element of this objective from the search element.

This approach assumed that the journal had superior access to the conference abstracts than was available through the journal interface. For example, that the study records and conference abstracts were available in a bibliographic management tool housed on an internal server. The editorial team were contacted twice to request data: first to make the request and second to chase for a response to the initial e-mail. Contacting the journal took approximately five minutes and cost approximately £2.65.

The journal could not provide any of the 604 conference abstracts. The editorial team confirmed that they only had access to abstracts via the journal interface. Given that no abstracts were acquired this is not a viable option for future researchers. As such, there is currently no known way to expedite export of ASH conference proceedings following a handsearch.

Objective 2 – effectiveness of identifying conference abstracts

The second objective was to evaluate the effectiveness of four comparators to identify the same abstracts as the reference standard handsearch across two stages of study identification. Stage 1: identification of potentially eligible abstracts through searching and, stage 2: identification of confirmed eligible abstracts through screening.

In Table 1, the results for stage 1 of the identification process – identifying the 604 potentially eligible abstracts – are presented. Only comparator 2 (journal search function) recalled the same 604 abstracts as the reference standard, so it is the most effective comparator, while the other comparators were less effective, identifying fewer potentially eligible abstracts overall. Comparator 3 (Embase) and comparator 4 (EndNote) recorded modest differences in precision compared to the handsearch. Comparator 3 (Embase) identified four duplicates and one study reported in another journal, and comparator 4 (EndNote) identified 22 duplicate abstracts due to the nature of search method.

Table 2 sets out differences between the reference standard and comparators as it relates to the identification of the 15 confirmed eligible abstracts. The results for the reference standard and comparator 2 (journal search function) are identical because

it was the exact same 604 references to be screened for inclusion in the review. No additional search terms were identified by the handsearch, so no new search terms were searched for using comparator 2 (journal search function).

The findings presented in Table 2 show that, for comparators 3-5 (Embase, EndNote and CPCI-S), the differences in recall for stage 1 (Table 1) latterly impacted recall for stage 2 (Table 2), since fewer potentially eligible abstracts were identified for screening overall which included differing numbers of confirmed eligible abstracts. The number of missed confirmed eligible abstracts varied by comparator: seven abstracts were missed in comparator 3 (the Embase search); all 15 abstracts were missed in comparator 4 (the EndNote search); and six abstracts were missed in comparator 5 (the CPCI-S search).

These findings indicate that, not only is there no way to expedite export of abstracts presented at ASH (objective one), but also with the exception of comparator 2 (journal search function), all other comparators missed confirmed eligible abstracts.

344 Table 1: Identifying abstracts as potentially eligible for screening and downloading them (stage 1)

	Reference standard	**Comparators**			
	Handsearch	2. Journal search function	3. Embase	4. EndNote	5. Searching CPCI-S
Total number of abstracts	17,759	604	464	22	201 (of 17,759)
Total number of abstracts identified as potentially relevant	604	604	463	20	201
Recall (Sensitivity) %		100 (99.39, 100.00)	76.7 (73.07, 79.97)	3.31 (2.03, 5.07)	33.28 (29.53, 37.19)
Precision (Positive Predictive Value) %, (95% CI)		100 (99.2, 100) [a]	99.8 (98.8, 100.0)	90.9 (70.8, 98.9)	100 (99.2, 100) [a]
F-Measure (95% CI)		1.00[b]	0.87 (0.8447, 0.8889) [c]	0.06 (0.0368, 0.0878) [c]	0.49 (0.4576, 0.5425) [c]
Time taken for stage 1, minutes	689 (11 hours 48 minutes)	72	22	20	6
Cost, GBP £	365.17	38.16	11.66	10.60	3.18

345

346

347

348

349

350

351

352 Table 2: Identifying abstracts which fulfilled inclusion in the systematic review (stage 2)

	Reference standard	**Comparators**			
	Handsearch	2. Journal search function	3. Embase	4. EndNote	5. Searching CPCI-S
Total number of abstracts potentially relevant	604	604	468 (of 604)	20 (of 604)	201 (of 604)
Number of abstracts that fulfil inclusion criteria	15	15	8 (of 15)	0 (of 15)	9 (of 15)
	Number of abstracts that fulfil inclusion criteria based on 15 from reference standard				
Recall (Sensitivity) %		100 (78.20 to 100.00)	53.3 (26.6 to 78.7)	0 (0.00 to 21.80)	60 (32.29 to 83.66)
Precision (Positive Predictive Value) %		2.48 (1.40, 4.06)	1.71 (0.74, 3.34)	0	4.48 (2.07, 8.33)
F-Measure (95% CI)		0.0485 (0.0246, 0.0723) [a]	0.0331 (0.0106, 0.0555) [a]	0 (cannot be calculated using bootstrap)	0.0833 (0.0323, 0.1350) [a]
Time taken to screen at stage 2, minutes	420 (0.696 per abstract)	420 (0.696 per abstract)	324 (5 hours 24 minutes)	13	66 (1 hour six minutes)
Cost to screen, GBP £	219.66	219.66	177.82	6.76	34.32

Objective 3 – efficiency of identifying conference abstracts

The third objective was to evaluate the efficiency of the comparators compared to the reference standard handsearch. Table 1 demonstrates that comparator 2 (journal search function) was more efficient compared to the reference standard (72 *vs.* 689 minutes) and was accordingly cheaper to undertake overall.

Comparators 3-5 (Embase, EndNote, CPCI-S) were more efficient in both time and cost when compared to the reference standard, but they all missed confirmed eligible abstracts. In other words, the efficiency was not simply a function of increased precision - eligible abstracts were missed alongside the ineligible. Since the purpose of the comparators was to identify all 15 confirmed eligible abstracts identified by the handsearch, comparators 3-5 are deemed ineffective overall. The F-Measure illustrates the difference between comparators and the harmonised effectiveness and efficiency findings, further suggesting that comparator 2 (journal search function) was optimal when compared to the other comparators.

Discussion

This work was initially conceived to address the question: how does a researcher efficiently export potentially eligible conference abstracts identified by handsearching the ASH conference to a bibliographic management tool for screening? The aim was ultimately revised since the task of identifying abstracts in the comparators could not be separated from the act of exporting eligible abstracts. The variation in recall between the reference standard and comparators, and the finding that comparators 3-5 (Embase, EndNote, CPCI-S) missed eligible studies, is the main finding of this work. This raises some potentially concerning questions about searching for conference abstracts by methods which do not involve a direct search of conference proceedings (either by handsearch or keyword searches). We do not know the extent to which existing completed reviews may have missed conference abstracts if they used one of the (potentially sub-optimal) comparators.

Generalisability of the findings

It is important to highlight the primary limitation of this work. The work presented here is the evaluation of one individual case study. The findings may not generalise to other searches in ASH, or other conferences, or in other disciplines. The finding

that comparator 2 (journal search function) was as effective but more efficient should be firmly situated in these limitations. The findings are not an argument to discontinue handsearching in systematic reviews.

It is anticipated that the findings set out here are specific to the date that the searching for comparators 3-5 were undertaken. Namely, as more content from ASH is added to bibliographic databases, a greater number of eligible abstracts would be identified. Changes in recall and precision in the comparators compared to the handsearch over time are expected. It is worth noting that many conferences are not published either separately on-line or in journals: work on how to identify such studies may be particularly valuable'.

Efficiency findings

Comparator 2 (journal search function) was simple and easy to use but, without the ability to select a range of abstracts (as is possible in bibliographic databases), the interfaces are not 'user friendly' for systematic reviews where multiple abstracts are likely to be downloaded. Most bibliographic database hosts have evolved to meet the needs of systematic reviewers and most database hosts facilitate complicated search strategies and the need to download a number of abstracts (34). Whilst the focus in this case study was on the journal *Blood*, an informal look at other journals which report conferences in supplement editions, suggests that the inability to download a number of abstracts is a common issue. Whilst it is acknowledged that journals and journal supplements serve a different purpose to bibliographic databases, increasing the ease with which conferences can be searched (if not handsearched) would be welcome, and the ability to select a number of abstracts for downloading rather than individual abstracts, may contribute to improved efficiencies in downloading conference abstracts and other material.

As it relates to efficiency, a question may be asked as to why it is necessary to export potentially eligible abstracts for screening, when the screening could have been undertaken during handsearching. The simple explanation in this case study (which is common to other reviews undertaken by the authors) was data management: so that a clear record of the studies/abstracts identified and processed in the review was maintained, and the research team had access to the bibliographic

data from each study for review and citation. As is set out above, the efficiency questions are to some extent unresolved, and other researchers may be less interested in the downloading of abstracts reported at conferences, but the practicable finding in recall between comparators is a key finding of this work.

Is handsearching still valid? Yes.

The finding that comparator 2 (journal search function) was as effective but more efficient does not necessarily generalise to other conferences. Comparator 2 may, however, provide some preliminary evidence that keyword searching might suit the needs of rapid reviews, which may accept less certainty in the comprehensiveness of their literature searching in exchange for more efficient searches (35). The risks of keyword searching compared to handsearching requires further examination.

The claimed advantages of handsearching have been recently summarised in a review of supplementary search methods (5). The advantages which relate to this case study specifically, include: identifying abstracts which have not yet been published or where there may be a delay between conference presentation and publication (8); handsearching may identify data which may not be reported in the abstract, for instance, where relevant data is reported in a figure or table, but not in the abstract (5, 17); and handsearching (as defined by the Cochrane handbook (4)) would include searching letters and other content not necessarily available to keyword searching (5, 14, 15, 19, 21).

The disadvantages of handsearching were also highlighted (5): namely, that handsearching is a resource intensive method of study identification (14, 24) and that handsearching may offer low precision (17, 21). This case study adds further evidence to these findings,. Adams *et al*. also identified that handsearching missed studies identified by bibliographic databases searching, which they associated with handsearcher fatigue. As with all searching for systematic reviews, cross-over between searches may mask the effect of the primacy of one search method over another and a clear demonstration of ‘true’ effectiveness (6, 17).

Handsearching remains a valuable method of study identification in systematic reviews. The findings do, however, underline that the resources required to

handsearch conferences may limit the practicable use of handsearching to systematic reviews which require comprehensive literature searches, where precision in the estimate from statistical meta-analysis is important, and demonstrable confidence that 'all' studies have been identified is required.

Conference abstract inclusion?

The work reported is based on recommended best practice (2, 36). The findings of this study support the importance of handsearching the ASH conference since 15 conference abstracts fulfilled inclusion criteria in the systematic review. These 15 abstracts represented 11.1% of includes. Studies reported at conferences represent a challenge to the practice of undertaking a review (37). Whilst guidance recommends searching conferences for a comprehensive literature search, guidance and studies also urge caution when including conference abstracts since the abstracts themselves rarely provide sufficient data to merit inclusion or permit quality appraisal (2, 7, 9, 38, 39). Studies have also found differences between findings presented at conferences and in peer-reviewed publications reported in journals which raises concerns about the validity of their reporting and the use of this type of study report in reviews (39-43).

Conference abstracts can, however, alert researchers to further unique studies, in particular those which may not otherwise be published, and highlight newly emerging data for studies which may or may not have already been identified. Whilst there are issues with the abstracts themselves, the need to identify studies reported at conference remains an important part of systematic reviews assessing the efficacy of clinical interventions.

Limitations

The measure of effectiveness was ultimately the ability of the comparators to identify the same 15 abstracts which eventually fulfilled inclusion into the systematic review. The interpretation that it is necessary to identify all 15 abstracts may over-state the contribution of these 15 (or individual) abstracts to the synthesis and overestimate the impact of the findings in this study. As is set out above, conference abstracts present a multitude of problems to the researcher, not least the paucity of data and the inability to appraise study quality. Determining the value of the 15 confirmed

eligible abstracts as a way to interpret the findings (beyond the fact that they met inclusion in the review) is difficult to empirically demonstrate. Where the abstracts contribute data, repeating the various meta-analyses and including and excluding the 15 conference abstracts as a form of sensitivity analysis, would likely only marginally alter the confidence intervals and not influence the overall estimate of effectiveness. Any certainty as to the real value of these abstracts would therefore be speculative beyond the fact that, in a review of intervention effectiveness, it is important to identify all relevant studies and study data to minimise bias.

The handsearch of abstract books was undertaken by only one researcher. Milne and Thorogood have suggested that independent double-handsearching could minimise the risk of error (24) but the resources available for this study prohibited this. It is acknowledged that two researchers independently handsearching abstracts would have improved the rigour however, the handsearch was cross-checked with a keyword search, and found the same abstracts.

Individual Cochrane groups undertake regular handsearching of conferences, the results of which are loaded into group trials registers and Cochrane's Central Register of Controlled Trials (CENTRAL). CENTRAL was searched to check if any of these 15 abstracts were already indexed. Only four abstracts of the 15 were indexed (44-47). The data file is reported in web-only material. This search was not included as a comparator, but it is worth considering, since Cochrane groups are tasked with handsearching journals to identify reports of studies. The findings of this case study more generally might also indicate a subtle revision to MECIR conduct standard 28, namely that databases of conference abstracts may not be a complete resource for the identification of studies reported at conferences (48).

We considered the idea of including web-scraping as a comparator. The legal position as to accessing data in this way and copyright generally were unclear. It would seem an area for further study if the legal position can be clarified.

Conclusion

The findings of this case study suggest that, in the case of the ASH conference, the efficiency of downloading abstracts could be improved if it were possible to identify

and export a range of potentially eligible abstracts. This finding appears relevant to other journals which offer conference abstracts in supplement editions online.

The revised scope of this case study highlights the main finding. Four potential comparators to a handsearch of conference abstracts for the ASH conference missed substantial numbers of potentially eligible and confirmed eligible abstracts. Further research is required to examine if this finding relates to other conferences or research disciplines. This finding suggests that, for researchers undertaking searches of the ASH conference, the only reliable method to identify eligible abstracts was a search of the original supplement editions.

Only comparator 2 (journal search function) was as effective in identification and recall as the reference standard handsearch, and it was more efficient. The other four comparators, whilst more efficient than both the reference standard and comparator 2, missed eligible abstracts so were deemed less effective.

Acknowledgements

We acknowledge the input of Francine Wood, UCL subject librarian, and her suggestions as to potential comparators. We thank the editorial team at the journal *Blood* for their assistance with our emails. We acknowledge the feedback and comments on a draft of this study from Robin Featherstone, Cochrane Information Specialist. We thank Rachel Hibbs and Joanne Noble-Longster for their valuable comments and edits to the draft manuscript.

The work set out in this study was initially presented as a poster which was presented under the same title at ISPOR 2019 (Copenhagen) (49).

Funding

The systematic review related to this study was sponsored by Takeda Pharmaceuticals, Cambridge MA, USA. The case study presented here was undertaken without any specific sponsorship or funding.

Declaration of interest

CC, CW, AP and AS all received funding to undertake the original systematic review from Takeda Pharmaceuticals. CC has received funding for teaching from *Clarivate* Analytics in 2016. None of these declarations impact upon this article, since no funding was received to undertake the work reported here, but we acknowledge these items here for completeness.

Contribution of authors:

- **Chris Cooper:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- **Tristan Snowsill:** Conceptualization, Methodology, Validation, Formal Analysis, Data Curation, Writing – Review & Editing.
- **Christine Worsley:** Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Review & Editing;
- **Amanda Prowse:** Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Review & Editing;
- **Alison O'Mara-Eves**: Conceptualization, Methodology, Writing – Review & Editing;
- **Helen Greenwood:** Visualization, Writing – Review & Editing;
- **Joanne Noble-**Longster: Writing – Review & Editing;
- **Emma Boulton:** Review & Editing, Project Administration; and
- **Amanda Strickson:** Data Curation, Writing – Review & Editing, Supervision.

1. Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness R. In: Eden J, Levit L, Berg A, Morton S, editors. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC): National Academies Press (US)
Copyright 2011 by the National Academy of Sciences. All rights reserved.; 2011.
2. Centre for Reviews and Dissemination. Systematic reviews – CRD's guidance for undertaking reviews in healthcare. York: Centre for Reviews and Dissemination, University of York; 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
3. Collaboration for Environmental Evidence. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management: Environmental Evidence; 2013. Available from: http://www.environmentalevidence.org/wp-content/uploads/2017/01/Review-guidelines-version-4.2-final-update.pdf.
4. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. 2011 [cited Accessed 7th December 2017]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. The Cochrane Collaboration, [cited Accessed 7th December 2017]. Available from: http://handbook.cochrane.org/.
5. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. Systematic Reviews. 2017;6(1):234.
6. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Medical Research Methodology. 2018;18(1):85.
7. National Institute for Health Care Excellence (NICE). Developing NICE guidelines: the manual 2014. Available from: https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf.
8. Mathieu S, Baron G, Soubrier M, Ravaud P. Timing of publication of abstracts of randomized controlled trials presented in congresses: The example of the European League against Rheumatism meeting. Joint Bone Spine. 2018;85(1):109-14.

9. The Institute of Medicine. Finding What Works in Health Care: Standards for Systematic Reviews. . Washington, DC: The National Academies 2011.
10. Paez A. Gray literature: An important resource in systematic reviews. Journal of Evidence-Based Medicine. 2017;10(3):233-40.
11. Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. Research Synthesis Methods. 2014;5(3):221-34.
12. Cooper C, Lovell R, Husk K, Booth A, Garside R. Supplementary search methods were more effective and offered better value than bibliographic database searching: A case study from public health and environmental enhancement. Research Synthesis Methods. 2018;9(2):195-223.
13. Blumle A, Antes G. [Handsearching for randomized controlled clinical trials in German medical journals]. Deutsche medizinische Wochenschrift. 2008;133(6):230-4.
14. Armstrong R, Jackson N, Doyle J, Waters E, Howes F. It's in your hands: the value of handsearching in conducting systematic reviews of public health interventions. Journal of Public Health. 2005;27(4):388-91.
15. Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. Statistics in Medicine. 2002;21(11):1625-34.
16. Young T, Hopewell S. Methods for obtaining unpublished data. The Cochrane database of systematic reviews. 2011;9(11).
17. Adams CE, Frederick K. An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. Psychological Medicine. 1994;24(3):741-8.
18. Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. Systematic Reviews. 2016;5(1):164.
19. Croft AM, Vassallo DJ, Rowe M. Handsearching the Journal of the Royal Army Medical Corps for trials. Journal of the Royal Army Medical Corps. 1999;145(2):86-8.
20. Hay PJ, Adams CE, Lefebvre C. The efficiency of searches for randomized controlled trials in the International Journal of Eating Disorders: a comparison of handsearching, EMBASE and PsycLIT. Health Libraries Review. 1996;13(2):91-6.
21. Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. The Online journal of current clinical trials. 1993;33.
22. Langham J, Thompson E, Rowan K. Identification of randomized controlled trials from the emergency medicine literature: comparison of hand searching versus MEDLINE searching. Annals of emergency medicine. 1999;34(1):25-34.
23. Mattioli S, Farioli A, Cooke RMT, Baldasseroni A, Ruotsalainen J, Placidi D, et al. Hidden effectiveness? Results of hand-searching Italian language journals for occupational health interventions. Occupational and Environmental Medicine. 2012;69(7):522-4.
24. Milne R, Thorogood M. Hand searching the Journal of Epidemiology and Community Health as part of th Cochrane Collaboration. Journal of Epidemiology and Community Health. 1996;50(2):178-81.
25. Christine Worsley, Amanda Prowse, Amanda Strickson, Chris Cooper. Pevonedistat hydrochloride (TAK-924/MLN4924) in combination with azacitidine, and comparators, for the first-line treatment of patients with higher-risk myelodysplastic syndromes and higher-risk chronic myelomonocytic leukaemia York: CRD; 2019 [Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019124715.
26. Heale R, Twycross A. What is a case study? Evidence Based Nursing. 2018;21(1):7-8.

27. Elsevier. Embase content 2019 [Available from: https://www.elsevier.com/solutions/embase-biomedical-research/embase-coverage-and-content.
28. Cooper C, Dawson S, Peters J, Varley-Campbell J, Cockcroft E, Hendon J, et al. Revisiting the need for a literature search narrative: A brief methodological note. Res Synth Methods. 2018;9(3):361-5.
29. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Research Synthesis Methods. 2014;5(1):31-49.
30. Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. J Clin Epidemiol. 2018;99:53-63.
31. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. International journal of technology assessment in health care. 2010;26(4):431-5.
32. Neyer S, Graduiertenakademie TN-I. Umfassende Literaturrecherchen–ein Kurzüberblick Comprehensive literature searches–an overview.
33. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval: Cambridge University Press; 2008.
34. Bethel A, Rogers M. A checklist to assess database-hosting platforms for designing and running searches for systematic reviews. Health Information & Libraries Journal. 2014;31(1):43-53.
35. Wagner G, Nussbaumer-Streit B, Greimel J, Ciapponi A, Gartlehner G. Trading certainty for speed - how much uncertainty are decisionmakers and guideline developers willing to accept when using rapid reviews: an international survey. BMC Medical Research Methodology. 2017;17(1):121.
36. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. Cochrane Handbook for Systematic Reviews of Interventions. (5.1.0) ed: The Cochrane Collaboration; 2011.
37. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. Systematic Reviews. 2019;8(1):264.
38. Saric L, Dosenovic S, Saldanha IJ, Kadic AJ, Puljak L. Conference abstracts describing systematic reviews on pain were selectively published, not reliable, and poorly reported. Journal of Clinical Epidemiology. 2019.
39. Hopewell S, Boutron I, Altman DG, Ravaud P. Deficiencies in the publication and reporting of the results of systematic reviews presented at scientific medical conferences. Journal of Clinical Epidemiology. 2015;68(12):1488-95.
40. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. The Cochrane database of systematic reviews. 2007(2):Mr000005.
41. Dundar Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR. Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies. Health technology assessment (Winchester, England). 2006;10(5):iii-iv, ix-145.
42. Saric L, Vucic K, Dragicevic K, Vrdoljak M, Jakus D, Vuka I, et al. Comparison of conference abstracts and full-text publications of randomized controlled trials presented at four consecutive World Congresses of Pain: Reporting quality and agreement of results. European Journal of Pain. 2019;23(1):107-16.
43. Weintraub WH. Are published manuscripts representative of the surgical meeting abstracts? An objective appraisal. Journal of Pediatric Surgery. 1987;22(1):11-3.

44. Zonghong Shao, Hui Liu, Hao Jiang, Hongyan Tong, Ruixiang Xiang, Linhua Yang, et al. FINAL Results of an Phase, Multicenter, Randomied, Controlled OPEN LEVEL Trial: Decitabine Therapy in Patients with Myelodysplastic Syndromes. ASH. 2018.
45. Finelli C, Clissa C, Follo MY, Parisi S, Fogli M, Mongiorgi S, et al. Comparison of Two Different Therapeutic Regimens with Azacitidine and Lenalidomide (Combined versus Sequential) in Higher-Risk Myelodysplastic Syndromes. Update of Long-Term Results of a Randomized Phase II Multicenter Study. Blood. 2018;132(Suppl 1):4365-.
46. Finelli C, Clissa C, Follo M, Stanzani M, Parisi S, Mongiorgi S, et al. Azacitidine and lenalidomide (combined vs sequential treatment) in higher-risk myelodysplastic syndromes. long-term results of a randomized phase II multicenter study. Blood Conference: 58th Annual Meeting of the American Society of Hematology, ASH. 2016;128(22).
47. Garcia-Manero G, Daver NG, Montalban-Bravo G, Jabbour EJ, DiNardo CD, Kornblau SM, et al. A phase II study evaluating the combination of nivolumab (Nivo) or Ipilimumab (Ipi) with azacitidine in pts with previously treated or untreated myelodysplastic syndromes (MDS). Blood Conference: 58th Annual Meeting of the American Society of Hematology, ASH. 2016;128(22).
48. Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews London2016 [Available from: http://methods.cochrane.org/mecir.
49. Cooper C, Worsley C, Prowse A, O'Mara-Eves A, Greenwood H, Boulton E, et al., editors. COMPARISON OF FIVE METHODS FOR DOWNLOADING CONFERENCE ABSTRACTS - A CASE STUDY. ISPOR Europe 2019, Copenhagen, Denmark; 2019: https://www.ispor.org/heor-resources/presentations-database/presentation ….

What's New

Key findings:

The effectiveness and efficiency of methods to identify and export conference abstracts presented at the American Society of Hematology (ASH) conference 2016-2018 for a systematic review were evaluated. Handsearching was the reference standard method which was compared to:

comparator 1: contacting the publisher to request abstracts;
comparator 2: keyword search of supplement editions;
comparator 3: searching Embase (ovid interface);
comparator 4: searching PubMed via Endnote X8;
comparator 5: searching Conference Proceedings Citation Index- Science (CPCI-S).

Only keyword searching (comparator 2) identified all eligible abstracts identified by the handsearch and it was more efficient than handsearching. All other comparators missed eligible studies.

No alternative methods to download conference abstracts in bulk – as opposed to abstract-by-abstract and individually – were identified.

What this adds to what is known:

The findings of this case-study may raise concerns about the coverage of the conference proceedings by the comparators set out above and their use to identify conference abstracts instead of handsearching. The comparators tested represent 'real world' options for use in systematic reviews, yet these findings suggest that many are sub-optimal.

There is a trade-off between effectiveness and efficiency that is likely to be inherent in decisions made about searching for conference proceedings, the extent of which has not been quantified and made explicit until now.

What are the implications:

Our findings are based on a single case-study and they may not generalise to other interventions reported at ASH, other conferences, or other topics. It is unlikely, however, that this is an isolated issue; further research might explore this.

Researchers should consider the potential risk of the trade-off between efficiency and effectiveness when designing their conference search strategy. Further research is also indicated on how conference abstracts are identified and included in databases, since this may affect recall (as identified in this study) and alter decision-making when deciding to handsearch.

The ability to bulk download eligible study abstracts from journal web-sites would also improve efficiency.

We do not know the extent to which existing completed reviews may have missed conference abstracts if they used one of the (potentially sub-optimal) comparators.

Declaration of interest

CC, CW, AP and AS all received funding to undertake the original systematic review from Takeda Pharmaceuticals. CC has received funding for teaching from *Clarivate* Analytics in 2016. None of these declarations impact upon this article, since no funding was received to undertake the work reported here, but we acknowledge these items here for completeness.

Funding

The systematic review related to this study was sponsored by Takeda Pharmaceuticals, Cambridge MA, USA. The case study presented here was undertaken without any specific sponsorship or funding.