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, Exeter, UK
[3] Tolley Health Economics, Buxton, UK
[4] EPPi Centre, University College of London, London, UK
[5] Royal College of Psychiatrists, London, UK

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

36 conference proceedings. Efficiency of exporting would be improved if journals

37 permitted bulk downloads.

38

39

40

41

42

43 **Background**

44 Searching for reports of studies presented at a conference is an acknowledged

45 approach to study identification in systematic reviews (1-6). Guidance suggests that

46 searching conferences may identify newly emerging studies, or updated findings of

47 on-going studies, potentially ahead of journal publication (2, 3, 7, 8) and that

48 identifying and including conference abstracts may help minimise the introduction of

49 bias into systematic reviews (2, 4, 9). There is some evidence that searching

50 conferences is an effective method of identifying studies which might be missed by

51 other search methods and identifying studies that are reported at conferences but

52 never published (4, 5, 10-13).

53

54 Handsearching has traditionally been the method used to search for reports of

55 studies presented at conferences (6, 25, 26). Handsearching involves a manual,

56 page-by-page, examination of the entire contents of relevant journals, conference

57 proceedings and abstracts (2, 4, 7, 9, 14-16). There is evidence that handsearching

58 is effective when compared to bibliographic database searching and that

59 handsearching can identify studies (or study data) which may be missed by other

60 search methods (4, 5, 7, 13-15, 17-24). Whilst handsearching is known to be an

61 effective method of study identification, it is resource intensive (5).

62

63 When handsearching conference proceedings presented at the American Society of

64 Hematology (ASH) conference (2016-2018) for a systematic review (25), 604 reports

65 of potentially eligible abstracts were identified by a handsearch but there was no

66 option to export all 604 records to a bibliographic management tool in one export.

67 Instead, each of the 604 abstracts had to be individually identified and downloaded

68 one-by-one. This added to the resources required to complete the handsearch of

69 conference proceedings.

70

71   The inability to download all of the 604 potentially eligible abstracts at the same time,

72   as is possible in bibliographic databases (where individual studies or a range of

73   studies can be selected for export), motivated the question: what is the most efficient

74   way to export abstracts identified by handsearching conference proceedings into a

75   bibliographic management tool for further screening? The research team

76   hypothesised potential alternative methods (henceforth comparators) which could

77   lead to an efficient and successful export of abstracts into a bibliographic

78   management tool. This case study reports the evaluation of these comparators

79   compared to the handsearch.

80

81   It is not a straight-forward evaluation to report. When the comparators were tested, it

82   became apparent that, for some methods, the identification of abstracts could not be

83   isolated from the task of exporting abstracts. As such, the research objectives

84   became broader than the problem of exporting conference abstracts to include a

85   focus on the effective identification of conference abstracts reported at ASH.

86

87   **Study objectives**

88   This case study aims to evaluate the effectiveness and the efficiency of methods to

89   identify and download eligible conference abstracts reported at ASH 2016-2018 for a

90   systematic review of intervention effectiveness. The research objectives of this case

91   study are:

92

93   1. to determine whether there is a more efficient method for downloading eligible

94   conference abstracts following a handsearch compared to the current technology

95   (i.e., individually downloading records);

96   2. to evaluate the effectiveness of comparator methods to identify the same

97   abstracts found by the reference standard handsearch across two stages of study

98   identification ('potentially eligible' and 'confirmed eligible'); and

99   3. to evaluate the efficiency of the various methods across two stages of study

100   identification ('potentially eligible' and 'confirmed eligible').

101

102   **Methods**

103

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

137

Syntax
Pevonedistat
MLN4924
Decitabine
Dacogen
Azacitidine
Vidaza

138

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

163 function on the journal website. Further detail on this method is presented in the

164 web-only material.

165

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

167 Embase was chosen over the bibliographic database MEDLINE due to its inclusion

168 of conference proceedings and material (27). The terms for the interventions, and

169 associated Emtree controlled indexing, were searched in Embase using the Ovid

170 interface. This search was limited by publication type to conferences in two ways:

171

172 First, controlled indexing and search fields were searched for abstracts indexed by

173 publication type (line 1 below) and the ASH conference was searched using relevant

174 field codes, namely: cf = conference information and cg = conference publication

175 (line 2 below).

176

177 1. exp conference paper/

178 2. ash.cf,cg.

179 3. 1 or 2

180

181 Secondly, the journal *Blood* was searched for using the journal field code (jn) and the

182 abstracts returned were combined with a search for conference.af. (af = all fields).

183

184 These two searches were combined using the Boolean connector "OR" so both

185 approaches to limiting by publication type were included. The full search syntax,

186 including a search narrative, is presented in web-only material (28),.

187

188 *Comparator 4: a search for the journal Blood was made in PubMed in EndNote*

189 The search terms in Figure 1 were searched using the online search function of

190 EndNote X8. The following search logic was applied:

191

192 Journal – contains – Blood; AND

193 Year – contains – 2016*; AND

194 All fields – contains – the intervention terms in Figure 1**.

195

196 * 2016 was searched first, then 2017 and finally 2018. ** the intervention terms were

197 searched one at a time.

198

199 Abstracts were visually inspected and manually de-duplicated. Study records which

200 reported conference proceedings were retained whilst other journal content (i.e.

201 abstracts not reported at the ASH conference) were deleted. The search strategy is

202 reported in web-only material.

203

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

205 A search was undertaken in Conference Proceedings Citation Index- Science (CPCI-

206 S), Web of Science (Clarivate Analytics). The search terms in Figure 1 were

207 searched on the topic search field and search terms for ASH or: (American-Society-

208 of-Hematology) were searched on the conference search field. Searches were

209 refined to the years 2016, 2017 or 2018. The search strategy is reported in web-only

210 material.

211

212 **Analysis**

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

214

215 (stage 1) 'potentially eligible' abstracts were identified on the basis of title or

216 abstracts and the study record was retrieved for further inspection; and

217 (stage 2) 'confirmed eligible' abstracts were identified on the basis of screening the

218 abstract to confirm eligibility and inclusion in the systematic review.

219

220 For stage 1, the reference standard handsearch and comparator 2 (journal search

221 function, see below) were undertaken in the week commencing February 4th, 2018.

222 Abstracts were identified and individually (i.e. study-by-study) downloaded to

223 EndNote using the direct export function on the journal website. Google Chrome

224 (version 76.0.3809.132) was the web browser. Comparators 3-5 were undertaken on

225 June 20th, 2019. The search details are reported in web-only material.

226

227 For stage 2, the 604 abstracts identified in the reference standard were downloaded

228 to EndNote and were independently screened by two experienced reviewers (CW

229 and AP). A third experienced reviewer (AS) was available in the event of
230 disagreements.
231
232 **Outcome measurement**
233 The following outcomes were recorded for the reference standard and comparator
234 methods:
235
236 *Number of potentially eligible abstracts (stage 1)*
237 The reference standard identified 604 potentially eligible abstracts which were taken
238 forward for independent double-screening against predetermined inclusion criteria
239 (25). The number of abstracts identified by each of the comparator methods deemed
240 potentially eligible by the reference standard were recorded.
241
242 *Number of abstracts fulfilling inclusion criteria for the systematic review (stage 2)*
243 The number of abstracts identified from the reference standard as confirmed eligible
244 was 15. This represents the final point of comparison where the ability of the
245 comparators to identify these same 15 abstracts is compared.
246
247 *Time*
248 Time was recorded using the stopwatch function on an Apple iPhone 6s. Time was
249 recorded in minutes.
250
251 *Cost*
252 Cost was represented as GBP since this study was undertaken in the UK. An
253 approach similar to Shemilt *et al.* was followed to identify local unit costs (29). A mid-
254 point Grade 7 cost (spine point 40) was chosen, since this represents the median
255 pay of the grade of researcher who might usually undertake the work reported.
256 University College London salaries and on-costs (2018-2019) were used since this
257 represents the lead author's home institution and this was the year the case study
258 was undertaken. These costs included salary, direct salary costs (e.g. pension) and
259 university indirect costs. Similar to Shemilt *et al.* the costs included 'London
260 Weighting' which is an uplift provided to staff to cover additional costs of London.
261 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

296 (individually exporting abstracts). *Blood's* editorial team were contacted to enquire if
297 they could send the 604 potentially eligible records to the research team. All other
298 comparators could not isolate the export element of this objective from the search
299 element.
300
301 This approach assumed that the journal had superior access to the conference
302 abstracts than was available through the journal interface. For example, that the
303 study records and conference abstracts were available in a bibliographic
304 management tool housed on an internal server. The editorial team were contacted
305 twice to request data: first to make the request and second to chase for a response
306 to the initial e-mail. Contacting the journal took approximately five minutes and cost
307 approximately £2.65.
308
309 The journal could not provide any of the 604 conference abstracts. The editorial
310 team confirmed that they only had access to abstracts via the journal interface.
311 Given that no abstracts were acquired this is not a viable option for future
312 researchers. As such, there is currently no known way to expedite export of ASH
313 conference proceedings following a handsearch.
314
315 **Objective 2 – effectiveness of identifying conference abstracts**
316
317 The second objective was to evaluate the effectiveness of four comparators to
318 identify the same abstracts as the reference standard handsearch across two stages
319 of study identification. Stage 1: identification of potentially eligible abstracts through
320 searching and, stage 2: identification of confirmed eligible abstracts through
321 screening.
322
323 In Table 1, the results for stage 1 of the identification process – identifying the 604
324 potentially eligible abstracts – are presented. Only comparator 2 (journal search
325 function) recalled the same 604 abstracts as the reference standard, so it is the most
326 effective comparator, while the other comparators were less effective, identifying
327 fewer potentially eligible abstracts overall. Comparator 3 (Embase) and comparator 4
328 (EndNote) recorded modest differences in precision compared to the handsearch.
329 Comparator 3 (Embase) identified four duplicates and one study reported in another

330 journal, and comparator 4 (EndNote) identified 22 duplicate abstracts due to the

331 nature of search method.

332

333 Table 2 sets out differences between the reference standard and comparators as it

334 relates to the identification of the 15 confirmed eligible abstracts. The results for the

335 reference standard and comparator 2 (journal search function) are identical because

336 it was the exact same 604 references to be screened for inclusion in the review. No

337 additional search terms were identified by the handsearch, so no new search terms

338 were searched for using comparator 2 (journal search function).

339

340 The findings presented in Table 2 show that, for comparators 3-5 (Embase, EndNote

341 and CPCI-S), the differences in recall for stage 1 (Table 1) latterly impacted recall for

342 stage 2 (Table 2), since fewer potentially eligible abstracts were identified for

343 screening overall which included differing numbers of confirmed eligible abstracts.

344 The number of missed confirmed eligible abstracts varied by comparator: seven

345 abstracts were missed in comparator 3 (the Embase search); all 15 abstracts were

346 missed in comparator 4 (the EndNote search); and six abstracts were missed in

347 comparator 5 (the CPCI-S search).

348

349 These findings indicate that, not only is there no way to expedite export of abstracts

350 presented at ASH (objective one), but also with the exception of comparator 2

351 (journal search function), all other comparators missed confirmed eligible abstracts.

352 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

353

354

355

356

357

358

359

360 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

361

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

397   that comparator 2 (journal search function) was as effective but more efficient should

398   be firmly situated in these limitations. The findings are not an argument to

399   discontinue handsearching in systematic reviews.

400

401   It is anticipated that the findings set out here are specific to the date that the

402   searching for comparators 3-5 were undertaken. Namely, as more content from ASH

403   is added to bibliographic databases, a greater number of eligible abstracts would be

404   identified. Changes in recall and precision in the comparators compared to the

405   handsearch over time are expected. It is worth noting that many conferences are not

406   published either separately on-line or in journals: work on how to identify such

407   studies may be particularly valuable'.

408

409   **Efficiency findings**

410   Comparator 2 (journal search function) was simple and easy to use but, without the

411   ability to select a range of abstracts (as is possible in bibliographic databases), the

412   interfaces are not 'user friendly' for systematic reviews where multiple abstracts are

413   likely to be downloaded. Most bibliographic database hosts have evolved to meet the

414   needs of systematic reviewers and most database hosts facilitate complicated

415   search strategies and the need to download a number of abstracts (34). Whilst the

416   focus in this case study was on the journal *Blood*, an informal look at other journals

417   which report conferences in supplement editions, suggests that the inability to

418   download a number of abstracts is a common issue. Whilst it is acknowledged that

419   journals and journal supplements serve a different purpose to bibliographic

420   databases, increasing the ease with which conferences can be searched (if not

421   handsearched) would be welcome, and the ability to select a number of abstracts for

422   downloading rather than individual abstracts, may contribute to improved efficiencies

423   in downloading conference abstracts and other material.

424

425   As it relates to efficiency, a question may be asked as to why it is necessary to

426   export potentially eligible abstracts for screening, when the screening could have

427   been undertaken during handsearching. The simple explanation in this case study

428   (which is common to other reviews undertaken by the authors) was data

429   management: so that a clear record of the studies/abstracts identified and processed

430   in the review was maintained, and the research team had access to the bibliographic

431   data from each study for review and citation. As is set out above, the efficiency

432   questions are to some extent unresolved, and other researchers may be less

433   interested in the downloading of abstracts reported at conferences, but the

434   practicable finding in recall between comparators is a key finding of this work.

435

436   **Is handsearching still valid? Yes.**

437   The finding that comparator 2 (journal search function) was as effective but more

438   efficient does not necessarily generalise to other conferences. Comparator 2 may,

439   however, provide some preliminary evidence that keyword searching might suit the

440   needs of rapid reviews, which may accept less certainty in the comprehensiveness

441   of their literature searching in exchange for more efficient searches (35). The risks of

442   keyword searching compared to handsearching requires further examination.

443

444   The claimed advantages of handsearching have been recently summarised in a

445   review of supplementary search methods (5). The advantages which relate to this

446   case study specifically, include: identifying abstracts which have not yet been

447   published or where there may be a delay between conference presentation and

448   publication (8); handsearching may identify data which may not be reported in the

449   abstract, for instance, where relevant data is reported in a figure or table, but not in

450   the abstract (5, 17); and handsearching (as defined by the Cochrane handbook (4))

451   would include searching letters and other content not necessarily available to

452   keyword searching (5, 14, 15, 19, 21).

453

454   The disadvantages of handsearching were also highlighted (5): namely, that

455   handsearching is a resource intensive method of study identification (14, 24) and

456   that handsearching may offer low precision (17, 21). This case study adds further

457   evidence to these findings,. Adams *et al.* also identified that handsearching missed

458   studies identified by bibliographic databases searching, which they associated with

459   handsearcher fatigue. As with all searching for systematic reviews, cross-over

460   between searches may mask the effect of the primacy of one search method over

461   another and a clear demonstration of 'true' effectiveness (6, 17).

462

463   Handsearching remains a valuable method of study identification in systematic

464   reviews. The findings do, however, underline that the resources required to

465   handsearch conferences may limit the practicable use of handsearching to

466   systematic reviews which require comprehensive literature searches, where

467   precision in the estimate from statistical meta-analysis is important, and

468   demonstrable confidence that 'all' studies have been identified is required.

469

470   **Conference abstract inclusion?**

471   The work reported is based on recommended best practice (2, 36). The findings of

472   this study support the importance of handsearching the ASH conference since 15

473   conference abstracts fulfilled inclusion criteria in the systematic review. These 15

474   abstracts represented 11.1% of includes. Studies reported at conferences represent

475   a challenge to the practice of undertaking a review (37). Whilst guidance

476   recommends searching conferences for a comprehensive literature search, guidance

477   and studies also urge caution when including conference abstracts since the

478   abstracts themselves rarely provide sufficient data to merit inclusion or permit quality

479   appraisal (2, 7, 9, 38, 39). Studies have also found differences between findings

480   presented at conferences and in peer-reviewed publications reported in journals

481   which raises concerns about the validity of their reporting and the use of this type of

482   study report in reviews (39-43).

483

484   Conference abstracts can, however, alert researchers to further unique studies, in

485   particular those which may not otherwise be published, and highlight newly emerging

486   data for studies which may or may not have already been identified. Whilst there are

487   issues with the abstracts themselves, the need to identify studies reported at

488   conference remains an important part of systematic reviews assessing the efficacy of

489   clinical interventions.

490

491   **Limitations**

492   The measure of effectiveness was ultimately the ability of the comparators to identify

493   the same 15 abstracts which eventually fulfilled inclusion into the systematic review.

494   The interpretation that it is necessary to identify all 15 abstracts may over-state the

495   contribution of these 15 (or individual) abstracts to the synthesis and overestimate

496   the impact of the findings in this study. As is set out above, conference abstracts

497   present a multitude of problems to the researcher, not least the paucity of data and

498   the inability to appraise study quality. Determining the value of the 15 confirmed

499   eligible abstracts as a way to interpret the findings (beyond the fact that they met

500   inclusion in the review) is difficult to empirically demonstrate. Where the abstracts

501   contribute data, repeating the various meta-analyses and including and excluding the

502   15 conference abstracts as a form of sensitivity analysis, would likely only marginally

503   alter the confidence intervals and not influence the overall estimate of effectiveness.

504   Any certainty as to the real value of these abstracts would therefore be speculative

505   beyond the fact that, in a review of intervention effectiveness, it is important to

506   identify all relevant studies and study data to minimise bias.

507

508   The handsearch of abstract books was undertaken by only one researcher. Milne

509   and Thorogood have suggested that independent double-handsearching could

510   minimise the risk of error (24) but the resources available for this study prohibited

511   this. It is acknowledged that two researchers independently handsearching abstracts

512   would have improved the rigour however, the handsearch was cross-checked with a

513   keyword search, and found the same abstracts.

514

515   Individual Cochrane groups undertake regular handsearching of conferences, the

516   results of which are loaded into group trials registers and Cochrane's Central

517   Register of Controlled Trials (CENTRAL). CENTRAL was searched to check if any of

518   these 15 abstracts were already indexed. Only four abstracts of the 15 were indexed

519   (44-47). The data file is reported in web-only material. This search was not included

520   as a comparator, but it is worth considering, since Cochrane groups are tasked with

521   handsearching journals to identify reports of studies. The findings of this case study

522   more generally might also indicate a subtle revision to MECIR conduct standard 28,

523   namely that databases of conference abstracts may not be a complete resource for

524   the identification of studies reported at conferences (48).

525

526   We considered the idea of including web-scraping as a comparator. The legal

527   position as to accessing data in this way and copyright generally were unclear. It

528   would seem an area for further study if the legal position can be clarified.

529

530   **Conclusion**

531   The findings of this case study suggest that, in the case of the ASH conference, the

532   efficiency of downloading abstracts could be improved if it were possible to identify

533 and export a range of potentially eligible abstracts. This finding appears relevant to
534 other journals which offer conference abstracts in supplement editions online.
535
536 The revised scope of this case study highlights the main finding. Four potential
537 comparators to a handsearch of conference abstracts for the ASH conference
538 missed substantial numbers of potentially eligible and confirmed eligible abstracts.
539 Further research is required to examine if this finding relates to other conferences or
540 research disciplines. This finding suggests that, for researchers undertaking
541 searches of the ASH conference, the only reliable method to identify eligible
542 abstracts was a search of the original supplement editions.
543
544 Only comparator 2 (journal search function) was as effective in identification and
545 recall as the reference standard handsearch, and it was more efficient. The other
546 four comparators, whilst more efficient than both the reference standard and
547 comparator 2, missed eligible abstracts so were deemed less effective.
548

599

600

601

602

603

604

605

606

607

608

609

610

611

612

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

641 9.   The Institute of Medicine. Finding What Works in Health Care: Standards for
642 Systematic Reviews. . Washington, DC: The National Academies 2011.
643 10.   Paez A. Gray literature: An important resource in systematic reviews. Journal of
644 Evidence-Based Medicine. 2017;10(3):233-40.
645 11.   Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews:
646 challenges and benefits. Research Synthesis Methods. 2014;5(3):221-34.
647 12.   Cooper C, Lovell R, Husk K, Booth A, Garside R. Supplementary search methods
648 were more effective and offered better value than bibliographic database searching: A case
649 study from public health and environmental enhancement. Research Synthesis Methods.
650 2018;9(2):195-223.
651 13.   Blumle A, Antes G. [Handsearching for randomized controlled clinical trials in
652 German medical journals]. Deutsche medizinische Wochenschrift. 2008;133(6):230-4.
653 14.   Armstrong R, Jackson N, Doyle J, Waters E, Howes F. It's in your hands: the value of
654 handsearching in conducting systematic reviews of public health interventions. Journal of
655 Public Health. 2005;27(4):388-91.
656 15.   Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of
657 handsearching versus MEDLINE searching to identify reports of randomized controlled
658 trials. Statistics in Medicine. 2002;21(11):1625-34.
659 16.   Young T, Hopewell S. Methods for obtaining unpublished data. The Cochrane
660 database of systematic reviews. 2011;9(11).
661 17.   Adams CE, Frederick K. An investigation of the adequacy of MEDLINE searches for
662 randomized controlled trials (RCTs) of the effects of mental health care. Psychological
663 Medicine. 1994;24(3):741-8.
664 18.   Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al.
665 Searching and synthesising 'grey literature' and 'grey information' in public health: critical
666 reflections on three case studies. Systematic Reviews. 2016;5(1):164.
667 19.   Croft AM, Vassallo DJ, Rowe M. Handsearching the Journal of the Royal Army
668 Medical Corps for trials. Journal of the Royal Army Medical Corps. 1999;145(2):86-8.
669 20.   Hay PJ, Adams CE, Lefebvre C. The efficiency of searches for randomized controlled
670 trials in the International Journal of Eating Disorders: a comparison of handsearching,
671 EMBASE and PsycLIT. Health Libraries Review. 1996;13(2):91-6.
672 21.   Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials
673 for systematic reviews. The Online journal of current clinical trials. 1993;33.
674 22.   Langham J, Thompson E, Rowan K. Identification of randomized controlled trials
675 from the emergency medicine literature: comparison of hand searching versus MEDLINE
676 searching. Annals of emergency medicine. 1999;34(1):25-34.
677 23.   Mattioli S, Farioli A, Cooke RMT, Baldasseroni A, Ruotsalainen J, Placidi D, et al.
678 Hidden effectiveness? Results of hand-searching Italian language journals for occupational
679 health interventions. Occupational and Environmental Medicine. 2012;69(7):522-4.
680 24.   Milne R, Thorogood M. Hand searching the Journal of Epidemiology and Community
681 Health as part of th Cochrane Collaboration. Journal of Epidemiology and Community
682 Health. 1996;50(2):178-81.
683 25.   Christine Worsley, Amanda Prowse, Amanda Strickson, Chris Cooper. Pevonedistat
684 hydrochloride (TAK-924/MLN4924) in combination with azacitidine, and comparators, for
685 the first-line treatment of patients with higher-risk myelodysplastic syndromes and higher-
686 risk chronic myelomonocytic leukaemia York: CRD; 2019 [Available from:
687 https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019124715.
688 26.   Heale R, Twycross A. What is a case study? Evidence Based Nursing. 2018;21(1):7-
689 8.

690 27.   Elsevier. Embase content 2019 [Available from:
691 https://www.elsevier.com/solutions/embase-biomedical-research/embase-coverage-and-
692 content.
693 28.   Cooper C, Dawson S, Peters J, Varley-Campbell J, Cockcroft E, Hendon J, et al.
694 Revisiting the need for a literature search narrative: A brief methodological note. Res Synth
695 Methods. 2018;9(3):361-5.
696 29.   Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, et al.
697 Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening
698 workload in extremely large scoping reviews. Research Synthesis Methods. 2014;5(1):31-49.
699 30.   Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review
700 identifies six metrics and one method for assessing literature search effectiveness but no
701 consensus on appropriate use. J Clin Epidemiol. 2018;99:53-63.
702 31.   Booth A. How much searching is enough? Comprehensive versus optimal retrieval
703 for technology assessments. International journal of technology assessment in health care.
704 2010;26(4):431-5.
705 32.   Neyer S, Graduiertenakademie TN-I. Umfassende Literaturrecherchen–ein
706 Kurzüberblick Comprehensive literature searches–an overview.
707 33.   Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval:
708 Cambridge University Press; 2008.
709 34.   Bethel A, Rogers M. A checklist to assess database-hosting platforms for designing
710 and running searches for systematic reviews. Health Information & Libraries Journal.
711 2014;31(1):43-53.
712 35.   Wagner G, Nussbaumer-Streit B, Greimel J, Ciapponi A, Gartlehner G. Trading
713 certainty for speed - how much uncertainty are decisionmakers and guideline developers
714 willing to accept when using rapid reviews: an international survey. BMC Medical Research
715 Methodology. 2017;17(1):121.
716 36.   Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. Cochrane
717 Handbook for Systematic Reviews of Interventions. (5.1.0) ed: The Cochrane Collaboration;
718 2011.
719 37.   Scherer RW, Saldanha IJ. How should systematic reviewers handle conference
720 abstracts? A view from the trenches. Systematic Reviews. 2019;8(1):264.
721 38.   Saric L, Dosenovic S, Saldanha IJ, Kadic AJ, Puljak L. Conference abstracts
722 describing systematic reviews on pain were selectively published, not reliable, and poorly
723 reported. Journal of Clinical Epidemiology. 2019.
724 39.   Hopewell S, Boutron I, Altman DG, Ravaud P. Deficiencies in the publication and
725 reporting of the results of systematic reviews presented at scientific medical conferences.
726 Journal of Clinical Epidemiology. 2015;68(12):1488-95.
727 40.   Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented
728 in abstracts. The Cochrane database of systematic reviews. 2007(2):Mr000005.
729 41.   Dundar Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR. Comparison of
730 conference abstracts and presentations with full-text articles in the health technology
731 assessments of rapidly evolving technologies. Health technology assessment (Winchester,
732 England). 2006;10(5):iii-iv, ix-145.
733 42.   Saric L, Vucic K, Dragicevic K, Vrdoljak M, Jakus D, Vuka I, et al. Comparison of
734 conference abstracts and full-text publications of randomized controlled trials presented at
735 four consecutive World Congresses of Pain: Reporting quality and agreement of results.
736 European Journal of Pain. 2019;23(1):107-16.
737 43.   Weintraub WH. Are published manuscripts representative of the surgical meeting
738 abstracts? An objective appraisal. Journal of Pediatric Surgery. 1987;22(1):11-3.

739   44.   Zonghong Shao, Hui Liu, Hao Jiang, Hongyan Tong, Ruixiang Xiang, Linhua Yang,
740   et al. FINAL Results of an Phase, Multicenter, Randomied, Controlled OPEN LEVEL Trial:
741   Decitabine Therapy in Patients with Myelodysplastic Syndromes. ASH. 2018.
742   45.   Finelli C, Clissa C, Follo MY, Parisi S, Fogli M, Mongiorgi S, et al. Comparison of
743   Two Different Therapeutic Regimens with Azacitidine and Lenalidomide (Combined versus
744   Sequential) in Higher-Risk Myelodysplastic Syndromes. Update of Long-Term Results of a
745   Randomized Phase II Multicenter Study. Blood. 2018;132(Suppl 1):4365-.
746   46.   Finelli C, Clissa C, Follo M, Stanzani M, Parisi S, Mongiorgi S, et al. Azacitidine and
747   lenalidomide (combined vs sequential treatment) in higher-risk myelodysplastic syndromes.
748   long-term results of a randomized phase II multicenter study. Blood Conference: 58th Annual
749   Meeting of the American Society of Hematology, ASH. 2016;128(22).
750   47.   Garcia-Manero G, Daver NG, Montalban-Bravo G, Jabbour EJ, DiNardo CD,
751   Kornblau SM, et al. A phase II study evaluating the combination of nivolumab (Nivo) or
752   Ipilimumab (Ipi) with azacitidine in pts with previously treated or untreated myelodysplastic
753   syndromes (MDS). Blood Conference: 58th Annual Meeting of the American Society of
754   Hematology, ASH. 2016;128(22).
755   48.   Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological
756   Expectations of Cochrane Intervention Reviews London2016 [Available from:
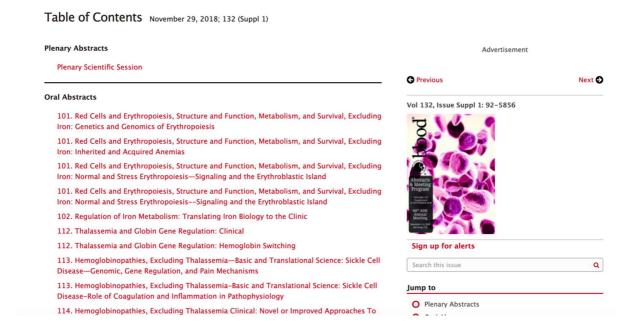757   http://methods.cochrane.org/mecir.
758   49.   Cooper C, Worsley C, Prowse A, O'Mara-Eves A, Greenwood H, Boulton E, et al.,
759   editors. COMPARISON OF FIVE METHODS FOR DOWNLOADING CONFERENCE
760   ABSTRACTS - A CASE STUDY. ISPOR Europe 2019, Copenhagen, Denmark; 2019:
761   https://www.ispor.org/heor-resources/presentations-database/presentation ….
762
763
764

Web Only Content

Reference standard

The handsearch was undertaken on Monday February 4th 2018 using the electronic supplement editions of the ASH conference reported in the journal Blood and available here: http://www.bloodjournal.org/page/ash-annual-meeting-abstracts
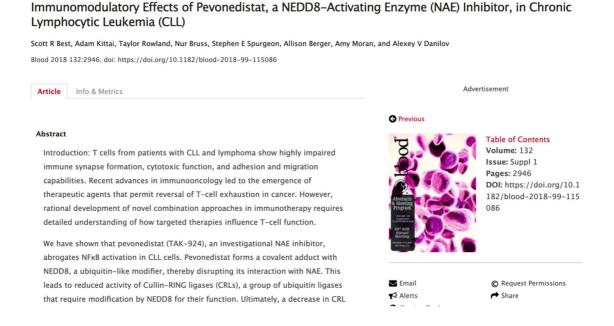
For each year (2016, 2017 and 2018) the table of contents was accessed. The table of contents lists the broad themes for the sessions (see below)

The hyper-link was followed which revealed the abstracts for each individual session. These were screened on title for the population or intervention.
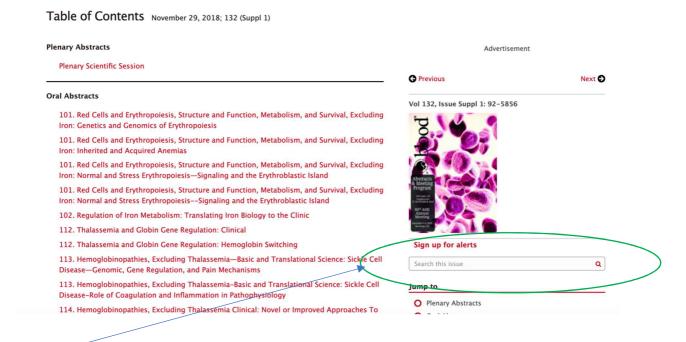
Any titles which looked relevant, or it was unclear, were screened at abstract.

Potentially eligible records were download to EndNote using the 'Citation Tools' function.

Comparator 2: journal search function

The area highlighted above in the green circle is journal search function. Keywords were searched here and then the approach for exporting data as reported for the reference standard (above) was followed.

Comparator 3: Embase

Database: Embase

Host: Ovid

Data parameters: 1974 to 2019 June 19

Date searched: June 20th 2019

Search purpose: the purpose of this search is to identify studies reported as conference abstracts at the conference ASH and which are available in the bibliographic database, Embase.

Search syntax	Search narrative
1 pevonedistat/ (428)	Lines 1-9 represent the principal search
2 Pevonedistat.ti,ab,kw,tn. (96)	terms for the interventions under review.
3 MLN4924.ti,ab,kw,tn. (416)	Search terms include controlled

4 decitabine/ (3561) 5 Decitabine.ti,ab,kw,tn. (3375) 6 Dacogen.ti,ab,kw,tn. (466) 7 azacitidine/ (12709) 8 Azacitidine.ti,ab,kw,tn. (3424) 9 Vidaza.ti,ab,kw,tn. (704) 10 1 or 2 or 3 or 4 or 5 6 or 7 or 8 or 9 (16884)	indexing using the Emtree language (lines 1, 4, and 7) and free-text search terms (lines 2, 3, 5, 6, 8, and 9). The free-text search terms are searched on: title (ti), abstract (ab), author generated key-word (kw) and drug trade name (tn). The searches for interventions are combined at line 10 using the Boolean connector OR. This means that all of the interventions are searched for.
11 exp conference paper/ (792271) 12 ash.cf,cg. (53694) 13 11 or 12 (845872) 14 10 and 13 (1760)	Line 11 is the controlled indexing term for studies reported at a conference. Line 12 focuses specifically on the ash conference by searching, 'conference information' (cf) or 'conference publication' (cg). Both of these ways of limiting to conferences are searched for at line 13 before they are combined with the intervention search terms set out above (lines 1-10).
15 blood.jn. (97533) 16 conference.af. (4263018) 17 15 and 16 (52448) 18 10 and 17 (1486) 19 14 or 18 (1761)	Line 15 searches for studies reported in the journal (jn) blood. This search is limited to studies reporting the term 'conference' in all available search fields (af).
20 limit 19 to yr="2016 - 2018" (468)	The search ((popn.) and (limit by publication type)) then limited to the years under review in this study.

Comparator 4: EndNote

Date of search: June 20th 2019

Search:

Journal contains Blood

AND

year 2016

AND

See Figure 1

	2016	2017	2018	
Pevonedistat	2	0	2	
MLN4924	0	0	2	
Decitabine	2	2 (4 in total but 2 were 2018 records)	7 (9 in total but 2 were 2018 records)	
Dacogen	2	2 (4 in total but 2 were 2018 records)	7 (9 in total but 2 were 2018 records)	
Azacitidine	1	5 (7 in total but 2 were 2018 records)	10 (12 in total but 2 were 2018 records)	
Vidaza	1	5 (7 in total but 2 were 2018 records)	10 (12 in total but 2 were 2018 records)	
Total	8	14	38	
- duplicates	4	9	27	
Total unique references	4	5	11	

Comparator 5: CPCI-S

Database: Conference Proceedings Citation Index- Science (CPCI-S)

Host: Web of Science (Clarivate Analytics)

Data parameters: 1990-Present

Date searched: June 20th 2019

1

1,291

TOPIC: ((Pevonedistat or "MLN4924" or Decitabine or Dacogen or Azacitidine or Vidaza))

Indexes=CPCI-S Timespan=All years

2

30,932

CONFERENCE: (ASH)

Indexes=CPCI-S Timespan=All years

3

81,815

CONFERENCE: (American-Society-of-Hematology)

Indexes=CPCI-S Timespan=All years

4

82,630

#3 OR #2

Indexes=CPCI-S Timespan=All years

5

696

#4 AND #1

Indexes=CPCI-S Timespan=All years

6

201

#4 AND #1

Refined by: PUBLICATION YEARS: (2018 OR 2017 OR 2016)

Indexes=CPCI-S Timespan=All years

2x2 tables for calculations

RS v IT 1 (journal search portal)

		Potential eligibility		
		In correct proceedings AND Contains intervention term(s)	Not in correct proceedings OR Does not contain intervention term(s)	Total
Searching journal portal	Study retrieved	604	0	604
	Study not retrieved	0		
	Total	604		
Recall (a.k.a. sensitivity): 604/604 = 100% (99.39%, 100%)				
Precision (a.k.a. positive predictive value): 604/604 = 100% (99.2%, 100%) [1-sided 97.5% CI]				
F1-measure: 1.00 (cannot calculate 95% CI in this instance)				

RS v IT 2 (Embase)

		Potential eligibility		
		In correct proceedings AND Contains intervention term(s)	Not in correct proceedings OR Does not contain intervention term(s)	Total
Embase	Study retrieved	463	1	464*
	Study not retrieved	136		
	Total	604		
*Excludes 4 duplicate entries				
Recall (a.k.a. sensitivity): 463/604 = 76.7% (73.1%, 80.0%)				
Precision (a.k.a. positive predictive value): 463/464 = 99.8% (98.8%, 100.0%)				
F1-measure: 0.867 (0.8447, 0.8889)				

RS v IT 3 (EndNote)

	Potential eligibility

		In correct proceedings AND Contains intervention term(s)	Not in correct proceedings OR Does not contain intervention term(s)	Total
EndNote	Study retrieved	20	22	42
	Study not retrieved	584		
	Total	604		

Recall (a.k.a. sensitivity): 20/604 = 3.31% (2.03%, 5.07%)

Precision (a.k.a. positive predictive value): 20/22 = 90.9% (70.8%, 98.9%)

F1-measure: 0.062 (0.0368, 0.0878)

RS v IT 4 (CPCI-S)

		Potential eligibility		
		In correct proceedings AND Contains intervention term(s)	Not in correct proceedings OR Does not contain intervention term(s)	Total
CPCI-S	Study retrieved	201	0	201
	Study not retrieved	403		
	Total	604		

Recall (a.k.a. sensitivity): 201/604 = 33.28% (29.53%, 37.19%)

Precision (a.k.a. positive predictive value): 201/201 = 100% (99.2%, 100%) [1-sided 97.5% CI]

F1-measure: 0.499 (0.4576, 0.5425)

RS v IT 5 (contacting the journal for exports of the identified records)

		Potential eligibility		
		In correct proceedings AND Contains intervention term(s)	Not in correct proceedings OR Does not contain intervention term(s)	Total
Contacting Blood	Study retrieved	0	0	0
	Study not retrieved	604		
	Total	604		

Recall (a.k.a. sensitivity): 0/604 = 0.00% (0.00%, 0.61%)

Precision (a.k.a. positive predictive value): 0/0 = 0

F1-measure: N/A

2nd phase

RS v IT1 (journal search portal)

		Confirmed eligibility		
		Eligible for inclusion in review	Not eligible for inclusion	Total
Searching journal portal	Study retrieved	15	589	604
	Study not retrieved	0	17155	17155
	Total	15	17744	17759
Recall: 15/15 = 100.00% (78.20%, 100.00%)				
Precision: 15/604 = 2.48% (1.40%, 4.06%)				
F1-measure: 0.0485 (0.0246, 0.0723)				

RS v IT2 (Embase)

		Confirmed eligibility		
		Eligible for inclusion in review	Not eligible for inclusion	Total
Embase	Study retrieved	8	455	463*
	Study not retrieved	7		
	Total	15		
*Excludes 4 duplicates and 1 study not from correct proceedings				
Recall: 8/15 = 53.5% (26.6%, 78.7%)				
Precision: 8/463 = 1.73% (0.75%, 3.38%)				
F1-measure: 0.0335 (0.0111, 0.0566)				

RS v IT3 (EndNote)

		Confirmed eligibility		
		igible for inclusion in review	Not eligible for inclusion	Total
EndNote	Study retrieved	0	20	20
	Study not retrieved	15		
	Total	15		
Recall: 0/15 = 0.00% (0.00%, 21.80%)				
Precision: 0/20 = 0% (0.00%, 16.84%)				

F1-measure: 0.00 (cannot be calculated using bootstrap)

RS v IT4 (CPCI-S)

		Confirmed eligibility		
		Eligible for inclusion in review	Not eligible for inclusion	Total
CPCI-S	Study retrieved	9	192	201
	Study not retrieved	6		
	Total	15		
Recall: 9/15 = 60.00% (32.29%, 83.66%)				
Precision:9/201 = 4.48% (2.07%, 8.33%)				
F1-measure: 0.0833 (0.0323, 0.1350)				

RS v IT5 (contacting the journal for exports of the identified records)

		Confirmed eligibility		
		Eligible for inclusion in review	Not eligible for inclusion	Total
Contacting Blood	Study retrieved	0	0	0
	Study not retrieved	15		
	Total	15		
Recall: 0/15 = 0.00% (0.00%, 21.80%)				
Precision: 0/0 = NA				
F1-measure: NA				

The 15 abstracts of studies fulfilling inclusion in this study and if they are indexed in CENTRAL

Study title	Included in CENTRAL?
FINAL results of an phase, multicenter, randomised, controlled OPEN LEVEL trial: decitabine therapy in patients with myelodysplastic syndromes (49)	yes
A Randomized Phase II Study of Azacitidine (AZA) Alone or with Lenalidomide (LEN), Valproic Acid (VPA) or Idarubicin (IDA) in Higher-Risk MDS: Gfm's 'pick a Winner' Trial (50)	no
Long-Term Experience with Hypomethylating Agents in Patients with Chronic Myelomonocytic Leukemia (51)	no
Preliminary Results from a Phase II Study of the Combination of Azacitidine and Pembrolizumab in Patients with Higher-Risk Myelodysplastic Syndrome (52)	no
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 (45)	yes
Double Immune Checkpoint Inhibitor Blockade with Nivolumab and Ipilimumab with or without Azacitidine in Patients with Myelodysplastic Syndrome (MDS)(53)	no
A Phase II Study of Nivolumab or Ipilimumab with or without Azacitidine for Patients with Myelodysplastic Syndrome (MDS) (54)	no
Planned Interim Analysis of a Phase 2 Study Evaluating the Combination of Pracinostat, a Histone Deacetylase Inhibitor (HDACi), and Azacitidine in Patients with High/Very High-Risk Myelodysplastic Syndrome (MDS) (55)	no
Azacitidine Use in the Real World Does Not Replicate AZA-001 Results in Higher Risk MDS/Low Blast Count AML: An Audit of 1101 Patients in the Cancer Care Ontario Registry (56)	no
Phase 2 Expansion Study of Oral Rigosertib Combined with Azacitidine (AZA) in Patients (Pts) with Higher-Risk (HR)	no

Myelodysplastic Syndromes (MDS): Efficacy and Safety Results in HMA Treatment Naïve & Relapsed (Rel)/Refractory (Ref) Patients (57)	
Azacitidine and lenalidomide (combined vs sequential treatment) in higher-risk myelodysplastic syndromes. long-term results of a randomized phase II multicenter study (46)	yes
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) (47)	yes
Myelodysplastic syndromes/myeloproliferative neoplasms treated with 5-azacytidine. Results from the hellenic 5-azacytidine registry (58)	no
Azacytidine failure revisited: An appraisal based on real life data from the MDS registry of the hellenic myelodysplastic syndrome study group (hMDS) (59)	no
The outcome of patients with high risk MDS achieving stable disease after treatment with 5-azacitidine. A retrospective analysis of the hellenic (Greek) MDS study group (60)	no