How many search results are enough...and what can we do about it?

Speaker: Dr Andrew Booth, University of Sheffield

Date: Thursday 12 July

Outline

- Why no benchmarks for the “comprehensive search?
- What does the evidence say?
- Introducing some new evidence
- Implications of findings to date
- The Way Forward

The Expectation:

- Systematic reviews of interventions **require a thorough, objective and reproducible search of a range of sources to identify as many relevant studies as possible (within resource limits).** This is a major factor in distinguishing systematic reviews from traditional narrative reviews and helps to minimize bias and therefore assist in achieving reliable estimates of effects.
- **Time and budget restraints require the review author to balance the thoroughness of the search with efficiency in use of time and funds** and the best way of achieving this balance is to be aware of, and try to minimize, the biases such as publication bias and language bias that can result from restricting searches in different ways.

Cochrane Handbook Chapter 6.1.1.2 – Minimizing Bias

The Issue:

The University Of Sheffield.

- **There are no "community" norms or benchmarks on (1) what is a sensitive search (2) what is an acceptable number of results to screen (3) what is an acceptable number to retrieve per included study**
- In the absence of such "community norms" systematic reviews are likely to be subject to **unofficial, implicit organizational "norms"**
- This means that the sensitivity of a literature search is likely to be determined **as much by the centre conducting the literature search as by the topic or review purpose**
- Clearly this is **wrong!**

What do we already know? - 1

The University Of Sheffield.

Item	Number Screened	Number Included	Number Needed to Retrieve
94 systematic reviews (Sampson et al, 2011)	189334	5734	**33**
Nichol et al., 2004	10578	254	42
Gulmezoglu et al., 2004	64586	2443	26

Average no. of references per review = **2,014**

Systematic Reviews using Information Specialists (Original Data – Booth, 2017)

	No of Refs from Dbase Searches before De-Duplication	No by Supplementary Sources	No of References After Deduplication	No Reviewed at Full Text	No Included	Number Needed to Retrieve
1	12269	1154	13006	15	4	3252
2	1656	26	1348	63	38	35
3	126	0	114	29	3	38
4	1815	15	1830	68	43	43
5	506	0	506	21	11	46
6	6192	0	4484	18	8	561
7	1259	0	524	16	10	52
8	12137	3419	8619	169	20	431
9	33488	3681	29761	231	9	3306
10	16876	204	11046	836	12	921
11	6214	0	5220	526	101	52
12	8975	35	9010	538	259	35
13	1174	3	1141	19	3	380
14	716	4	674	80	8	84
15	15488	29	14852	413	87	171
16	8599	90	8385	230	159	53
17	241	172	312	60	11	28
18	8425	27	8343	233	11	758
19	107	0	83	7	5	17
20	2645	112	2757	252	26	106

From “Random” Sample (n = 20) on Google Scholar (All using Info Scientist)

The University Of Sheffield.

- Mean Number of References per Review = 6100 References (almost 3 times Sampson sample)
- Mean Number of References to Screen per Included Paper = 147 References (Almost 4.5 times Sampson Sample)

Methods for this Study

The University Of Sheffield.

- 5 most recently published Cochrane Public Health Reviews
- 5 most recently published reviews on NIHR Journals Library for each of the 9 NICE TAR teams:
  - Aberdeen HTA Group, University of Aberdeen
  - BMJ Technology Assessment Group (BMJ-TAG)
  - Centre for Reviews and Dissemination/Centre for Health Economics, University of York
  - Kleijnen Systematic Reviews Ltd
  - Liverpool Reviews and Implementation Group, University of Liverpool
  - PenTAG, Evidence Synthesis & Modelling for Health Improvement (ESMI), University of Exeter
  - School of Health and Related Research, University of Sheffield
  - Southampton Health Technology Assessments Centre, University of Southampton
  - Warwick Evidence, University of Warwick

(NB. Reviews could be but were not necessarily TAR reports)

TAR Teams

The University Of Sheffield.

- TAR teams commissioned on the basis of their expertise in reviewing complex literature and in general knowledge of health service research and management. Teams are characterised by:
- Strong institutional base, with established university or NHS links;
- Multi-disciplinary scientific staff skilled in systematic reviewing, health economics, economic modelling, qualitative research and statistics;
- Dedicated senior staff to supervise, take responsibility for and quality assure each report;
- Access to a network of experts in public health, health services research and, ideally, social policy, psychology and sociology;
- Established links with the NHS and familiarity with service issues.

Data Extracted

The University Of Sheffield.

- Total Number of References Retrieved (for main Effectiveness Search)
- Total Number of Papers Retrieved at Full Text
- Total Number of Included Papers (not Studies)
- Whether Information Specialist was involved in Construction or Execution of Search Strategies
- *NB. All Reports published between 2014-2018*

The Bottom Line

The University Of Sheffield.

- Cochrane PH Reviews – Mean Number of References per Review = 25,151 (Range = 7,804 - 50,270)
- Cochrane PH Reviews – Mean Number Needed to Retrieve = 860 (Range = 234- 1795) Between 2 – 15 hours sifting for each Included Paper
- TAR Teams – Mean Number of References per Review = 6,328 (Range = 73 – 102267)
- TAR Teams – Mean Number Needed to Retrieve = 269 (Range = 6 – 2646) [NB 1 Review = NO included studies from 3644 references]
- Between 3.5 mins – 22 hours sifting for each Included Paper

By Institution

The University Of Sheffield.

Institution	Mean Refs	Mean NNR	Institution	Mean Refs	Mean NNR
COCHRANE PH	25,151 (7,804 - 50,270)	860 (234 – 1,795)	EXETER	3,693 (724 – 10,753)	165 (16 - 392)
LIVERPOOL	24,039 (190 -102,267)	226 (29 - 846)	WARWICK	3,555 (73 – 13,627)	52 (6 – 121)
SHEFFIELD	7,545 (2,724 - 16,591)	652 (36 – 395)	YORK	3,368 (1,961-4,516)	81 (15 – 181)
KLEIJNEN	7,364 (3,524 – 9,870)	235 (98 – 431)	SOTON	1,913 (655 – 2,628)	189 (65 – 329)
BMJ	4,101 (1,428 – 6,079)	105 (8 - >3,644)	ABERDEEN	1,372 (658 – 2,952)	73 (15 - 173)

By Institution (Time Per Relevant Record)

The University Of Sheffield.

Institution	Mean NNR	MEAN TIME	Institution	Mean NNR	MEAN TIME
COCHRANE PH	860 (234 – 1,795)	7 Hrs	EXETER	165 (16 - 392)	1.5 Hrs
LIVERPOOL	226 (29 - 846)	2 Hrs	WARWICK	52 (6 – 121)	0.5 Hrs
SHEFFIELD	652 (36 – 395)	5.5 Hrs	YORK	81 (15 – 181)	0.6 Hrs
KLEIJNEN	235 (98 – 431)	2 Hrs	SOTON	189 (65 – 329)	1.5 Hrs
BMJ	105 (8 - >3,644)	1 Hr	ABERDEEN	73 (15 - 173)	0.6 Hrs

BMJ	Clinical effectiveness of interventions for treatment-resistant anxiety in older people; a systematic review	3644	109	0	>3644	X	
SHEF	A systematic review and economic evaluation of adalimumab and dexamethasone	10585	134	4	2646.25	✓	NB 4 articles but 3 studies
PH	Nutritional labelling for healthier food or non-alcoholic drink purchasing and consumption	50270	278	28	1795.357	X	
PH	Interventions to prevent injuries in construction workers	18271	143	17	1074.765	✓	
LIVE	Automated tests for cognitive impairment	13542	399	16	846.375	X	
PH	Welfare-to-work interventions and their effects on the mental and physical health of lone parents and their children	7804	165	12	650.3333	X	
PH	Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries	30453	183	56	543.8036	X	NB 56 papers but 21 studies
KLEI	**> 3 Hrs** ImmunoCAP ISAC and Microtest for multiplex allergen testing in people with difficult to manage allergic disease	8619	169	20	430.95	✓	20 publications from 15 studies
SHEF	The clinical effectiveness and cost-effectiveness of treat-to-target strategies in rheumatoid arthritis	16591	179	42	395.0238	✓	NB 42 articles but 16 studies
EXET	Genetic testing for Lynch syndrome in people with colorectal cancer	3921	77	10	392.1	✓	

LIVE	A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour	102267	1268	930	109.9645	X	
SOUT	The INTRABEAM® Photon Radiotherapy System for the adjuvant treatment of early breast cancer: a systematic review and economic evaluation	655	44	6	109.1667	✓	6 references but 1 study
KLEI	Procalcitonin (PCT) testing to guide antibiotic therapy for the treatment of sepsis in intensive care settings and for suspected bacterial infection in emergency department settings.	3524	160	36	97.88889	✓	36 publications for 18 studies
ABER	The clinical effectiveness and cost-effectiveness of open mesh repairs in adults presenting with a clinically diagnosed primary unilateral inguinal hernia who are operated in an elective setting: systematic review and economic evaluation	1204	82	13	92.61538	✓	NB 13 articles from 12 RCTs
YORK	High-throughput, non-invasive prenatal testing for fetal rhesus D status in RhD-negative women not known to be sensitised to the RhD antigen: a systematic review and economic evaluation	3921	227	45	87.13333	✓	NB 45 papers but 14 studies
LIVE	Allopurinol for chronic kidney disease: a systematic review	1850	77	22	84.09091	X	NB 22 papers but 20 studies
WARW	Multiplex tests to identify gastrointestinal bacteria, viruses and parasites in people with suspected infectious gastroenteritis: systematic review and economic analysis	2215	110	28	79.10714	✓	23 studies in 28 articles
YORK	Interventional management of hyperhidrosis: an evidence synthesis and value of information analysis	4057	435	57	71.17544	✓	NB 57 records but 48 studies
SOUT	Virtual chromoendoscopy for real-time assessment of colorectal polyps during colonoscopy	2070	125	32	64.6875	✓	32 papers for 30 studies
LIVE	The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation	2249	228	37	60.78378	X	
BMJ	Clinical and cost-effectiveness of interventions for the treatment of anogenital warts: systematic review and economic evaluation	4231	155	70	60.44286	X	70 publicatiosn describing 60 studies
YORK	Adalimumab, etanercept and ustekinumab for treating plaque psoriasis in children and young people [ID854]	2386	111	48	49.70833	✓	NB 48 records from 9 studies
ABER	Sedation in intensive care	1182	83	24	49.25	✓	NB 24 papers from 18 studies
EXET	Diagnostic strategies for hereditary non-polyposis colorectal cancer	2036	224	43	47.34884	✓	
EXET	A systematic review and economic evaluation of intraoperative tests (RD-100i OSNA system and Metasin test) for detecting sentinel lymph node metastases in breast cancer	724	135	16	45.25	✓	18 papers but 2 unpublished
SHEF	Sepsis: the LightCycler SeptiFast Test MGRADE®, SepsiTest and IRIDICA BAC BSI assay	2892	177	66	43.81818	✓	
SHEF	Ultrasound joint examination for monitoring synovitis in rheumatoid arthritis	2724	154	75	36.32	✓	NB 75 papers but 58 studies
ABER	Collagenase clostridium histolyticum for the treatment of Dupuytren's contracture (ID621)	1222	187	34	35.94118	✓	NB 34 reports for 30 studies
WARW	Canagliflozin, dapagliflozin and empagliflozin monotherapy for treating type 2 diabetes	246	51	8	30.75	✓	8 articles representing 7 trials
LIVE	The clinical and cost effectiveness of heated humidified high-flow nasal cannula vs usual care for preterm infants	290	49	10	29	✓	
WARW	Crohn's disease: Tests for therapeutic monitoring of TNF inhibitors (LISA-TRACKER ELISA kits, TNFa-Blocker ELISA kits, and Promonitor ELISA kits)	1616	257	70	23.08571	✓	70 papers reporting 68 studies
EXET	What is the clinical and cost-effectiveness of conservative interventions for elbow tendinopathy?	1029	140	65	15.83077	✓	
YORK	Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease modifying anti-rheumatic drugs [ID579]	1961	182	130	15.08462	✓	
ABER	CoaguChek XS point-of-care blood coaguability testing system for those on long-term vitamin K antagonist therapy	658	120	45	14.62222	✓	NB 45 papers from 26 RCTs

< 1 Hr

Implications to Date

The University Of Sheffield.

- Little Evidence of “Community Norms” on Acceptable Result Sets or Numbers Needed to Read
- Considerable Variation in Means and Ranges
- Clear Difference in Expectations between Cochrane (Public Health) Reviews and NIHR HTA Products
- Some Evidence for “Institutional Norms” (although different Information Specialists involved; some without Info Specialists)
- Some Evidence that Information Specialists are associated with more Efficient Searches, No Info Specialist with less Efficient Searches

The Way Forward

The University Of Sheffield.

- Could Information Specialists Construct Search Strategies to deliver within Institutional/Community Norms (Total Result Set and Number Needed to Retrieve)?
- Should we move to Tiered Literature Searches to Deliver to Expectations?
- Would it be Helpful to develop Expectations by Discipline (e.g. Public Health, HS&DR, HTA) and/or by Purpose Cochrane Review/HTA?
- Could we make more use of the PRISMA routine data – for Benchmarking and Internal Audit?

Take Home Messages

The University Of Sheffield.

- “Typical” Systematic Review includes between 2000 and 6500 references (once duplicates removed)
- “Typical” Systematic Review delivers one relevant reference for every 33 - 240 references retrieved
- Translates from one every 10 minutes through to one every Two Hours of Sifting Time
- “Typical” Systematic Review requires looking at between 150 – 400 full texts (Mean = 215) to identify 57 Included studies
- You can document this for every search you conduct – build up the evidence base!

Above All

Transform the dialogue!

FROM:

- How **MANY** search results are enough?

TO:

- How **FEW** Search Results are enough?

References

- Gülmezoglu AM, Say L, Betrán AP, et al. WHO systematic review of maternal mortality and morbidity: methodological issues and challenges. BMC Medical Research Methodology 2004;4. doi:10.1186/1471-2288-4-16
- Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. Research Synthesis Methods; 2011 Jun;2(2):119–25. Available from: http://dx.doi.org/10.1002/jrsm.42