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:

• 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

<table>
<thead>
<tr>
<th>Item</th>
<th>Number Screened</th>
<th>Number Included</th>
<th>Number Needed to Retrieve</th>
</tr>
</thead>
<tbody>
<tr>
<td>94 systematic reviews (Sampson et al, 2011)</td>
<td>189334</td>
<td>5734</td>
<td>33</td>
</tr>
<tr>
<td>Nichol et al., 2004</td>
<td>10578</td>
<td>254</td>
<td>42</td>
</tr>
<tr>
<td>Gulmezoglu et al., 2004</td>
<td>64586</td>
<td>2443</td>
<td>26</td>
</tr>
</tbody>
</table>

**Average no. of references per review = 2,014**
<table>
<thead>
<tr>
<th></th>
<th>No of Refs</th>
<th>No by Supplementary Sources</th>
<th>No of References After Deduplication</th>
<th>No Reviewed at Full Text</th>
<th>No Included</th>
<th>Number Needed to Retrieve</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12269</td>
<td>1154</td>
<td>13006</td>
<td>15</td>
<td>4</td>
<td>3252</td>
</tr>
<tr>
<td>2</td>
<td>1656</td>
<td>26</td>
<td>1348</td>
<td>63</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>126</td>
<td>0</td>
<td>114</td>
<td>29</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>1815</td>
<td>15</td>
<td>1830</td>
<td>68</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>506</td>
<td>0</td>
<td>506</td>
<td>21</td>
<td>11</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>6192</td>
<td>0</td>
<td>4484</td>
<td>18</td>
<td>8</td>
<td>561</td>
</tr>
<tr>
<td>7</td>
<td>1259</td>
<td>0</td>
<td>524</td>
<td>16</td>
<td>10</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>12137</td>
<td>3419</td>
<td>8619</td>
<td>169</td>
<td>20</td>
<td>431</td>
</tr>
<tr>
<td>9</td>
<td>33488</td>
<td>3681</td>
<td>29761</td>
<td>231</td>
<td>9</td>
<td>3306</td>
</tr>
<tr>
<td>10</td>
<td>16876</td>
<td>204</td>
<td>11046</td>
<td>836</td>
<td>12</td>
<td>921</td>
</tr>
<tr>
<td>11</td>
<td>6214</td>
<td>0</td>
<td>5220</td>
<td>526</td>
<td>101</td>
<td>52</td>
</tr>
<tr>
<td>12</td>
<td>8975</td>
<td>35</td>
<td>9010</td>
<td>538</td>
<td>259</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>1174</td>
<td>3</td>
<td>1141</td>
<td>19</td>
<td>3</td>
<td>380</td>
</tr>
<tr>
<td>14</td>
<td>716</td>
<td>4</td>
<td>674</td>
<td>80</td>
<td>8</td>
<td>84</td>
</tr>
<tr>
<td>15</td>
<td>15488</td>
<td>29</td>
<td>14852</td>
<td>413</td>
<td>87</td>
<td>171</td>
</tr>
<tr>
<td>16</td>
<td>8599</td>
<td>90</td>
<td>8385</td>
<td>230</td>
<td>159</td>
<td>53</td>
</tr>
<tr>
<td>17</td>
<td>241</td>
<td>172</td>
<td>312</td>
<td>60</td>
<td>11</td>
<td>28</td>
</tr>
<tr>
<td>18</td>
<td>8425</td>
<td>27</td>
<td>8843</td>
<td>233</td>
<td>11</td>
<td>758</td>
</tr>
<tr>
<td>19</td>
<td>107</td>
<td>0</td>
<td>83</td>
<td>7</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>20</td>
<td>2645</td>
<td>112</td>
<td>2757</td>
<td>252</td>
<td>26</td>
<td>106</td>
</tr>
</tbody>
</table>
From “Random” Sample (n = 20) on Google Scholar (All using Info Scientist)

• 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

• 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

• 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

• 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

• 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

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

<table>
<thead>
<tr>
<th>Institution</th>
<th>Mean NNR</th>
<th>MEAN TIME</th>
<th>Institution</th>
<th>Mean NNR</th>
<th>MEAN TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCHRANE PH</td>
<td>860 (234 – 1,795)</td>
<td>7 Hrs</td>
<td>EXETER</td>
<td>165 (16 – 392)</td>
<td>1.5 Hrs</td>
</tr>
<tr>
<td>LIVERPOOL</td>
<td>226 (29 – 846)</td>
<td>2 Hrs</td>
<td>WARWICK</td>
<td>52 (6 – 121)</td>
<td>0.5 Hrs</td>
</tr>
<tr>
<td>SHEFFIELD</td>
<td>652 (36 – 395)</td>
<td>5.5 Hrs</td>
<td>YORK</td>
<td>81 (15 – 181)</td>
<td>0.6 Hrs</td>
</tr>
<tr>
<td>KLEIJNEN</td>
<td>235 (98 – 431)</td>
<td>2 Hrs</td>
<td>SOTON</td>
<td>189 (65 – 329)</td>
<td>1.5 Hrs</td>
</tr>
<tr>
<td>BMJ</td>
<td>105 (8 – &gt;3,644)</td>
<td>1 Hr</td>
<td>ABERDEEN</td>
<td>73 (15 – 173)</td>
<td>0.6 Hrs</td>
</tr>
<tr>
<td>Journal</td>
<td>Title</td>
<td>JCR 2019</td>
<td>cites</td>
<td>IF 2019</td>
<td>status</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>----------</td>
<td>-------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>BMJ</td>
<td>Clinical effectiveness of interventions for treatment-resistant anxiety in older people; a systematic review</td>
<td>3644</td>
<td>109</td>
<td>0&gt;3644</td>
<td>X</td>
</tr>
<tr>
<td>SHEF</td>
<td>A systematic review and economic evaluation of adalimumab and dexamethasone</td>
<td>10585</td>
<td>134</td>
<td>4</td>
<td>2646.25</td>
</tr>
<tr>
<td>PH</td>
<td>Nutritional labelling for healthier food or non-alcoholic drink purchasing and consumption</td>
<td>50270</td>
<td>278</td>
<td>28</td>
<td>1795.357</td>
</tr>
<tr>
<td>PH</td>
<td>Interventions to prevent injuries in construction workers</td>
<td>18271</td>
<td>143</td>
<td>17</td>
<td>1074.765</td>
</tr>
<tr>
<td>LIVE</td>
<td>Automated tests for cognitive impairment</td>
<td>13542</td>
<td>399</td>
<td>16</td>
<td>846.375</td>
</tr>
<tr>
<td>PH</td>
<td>Welfare-to-work interventions and their effects on the mental and physical health of lone parents and their children</td>
<td>7804</td>
<td>165</td>
<td>12</td>
<td>650.3333</td>
</tr>
<tr>
<td>PH</td>
<td>Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries</td>
<td>30453</td>
<td>183</td>
<td>56</td>
<td>543.8036</td>
</tr>
<tr>
<td>KLEI</td>
<td>ImmunoCAP ISAC and Microtest for multiplex allergen testing in people with difficult to manage allergic disease</td>
<td>8619</td>
<td>169</td>
<td>20</td>
<td>430.95</td>
</tr>
<tr>
<td>SHEF</td>
<td>The clinical effectiveness and cost-effectiveness of treat-to-target strategies in rheumatoid arthritis</td>
<td>16591</td>
<td>179</td>
<td>42</td>
<td>395.0238</td>
</tr>
<tr>
<td>EXET</td>
<td>Genetic testing for Lynch syndrome in people with colorectal cancer</td>
<td>3921</td>
<td>77</td>
<td>10</td>
<td>392.1</td>
</tr>
<tr>
<td>Study</td>
<td>Results</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour</td>
<td>The INTRABEAM® Photon Radiotherapy System for the adjuvant treatment of early breast cancer: a systematic review and economic evaluation</td>
<td>6 references but 1 study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td>655 44 6 109.1667</td>
<td>36 publications for 18 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>1204 82 13 92.6153</td>
<td>NB 13 articles from 12 RCTs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>3921 227 45 87.1333</td>
<td>NB 45 papers but 14 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alloplurinol for chronic kidney disease: a systematic review</td>
<td>1850 77 22 84.0909</td>
<td>NB 22 papers but 20 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplex tests to identify gastrointestinal bacteria, viruses and parasites in people with suspected infectious gastroenteritis: systematic review and economic analysis</td>
<td>2215 110 28 79.1071</td>
<td>23 studies in 28 articles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interventionsal management of hyperhidrosis: an evidence synthesis and value of information analysis</td>
<td>4057 435 57 71.1754</td>
<td>NB 57 records but 48 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual chroendoxyloscopy for real-time assessment of colorectal polyps during colonoscopy</td>
<td>2070 125 32 64.6875</td>
<td>32 papers for 30 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>2249 228 37 60.7837</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical and cost-effectiveness of interventions for the treatment of anogenital warts: systematic review and economic evaluation</td>
<td>4231 155 70 60.4428</td>
<td>70 publications describing 60 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adalimumab, etanercept and ustekinumab for treating plaque psoriasis in children and young people</td>
<td>2386 111 48 49.7083</td>
<td>NB 48 records from 9 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedation in intensive care</td>
<td>1182 83 24 49.25</td>
<td>NB 24 papers from 18 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic strategies for hereditary non-polyposis colorectal cancer</td>
<td>2036 224 43 47.3488</td>
<td>18 papers but 2 unpublished</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>724 135 16 45.25</td>
<td>NB 75 papers but 58 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sepsis: the LightCycler SeptiFast Test MGRADE®, SepsTest and IRIDICA BAC BSI assay</td>
<td>2892 177 66 43.8181</td>
<td>NB 34 reports for 30 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound joint examination for monitoring synovitis in rheumatoid arthritis</td>
<td>2724 154 75 36.32</td>
<td>8 articles representing 7 trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagenase clostridium histolyticum for the treatment of Dupuytren's contracture (ID621)</td>
<td>1222 187 34 35.9411</td>
<td>70 publications reporting 68 studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canagliflozin, dapagliflozin and empagliflozin monotherapy for treating type 2 diabetes</td>
<td>246 51 8 30.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The clinical and cost effectiveness of heated humidified high-flow nasal cannula vs usual care for preterm infants</td>
<td>290 49 10 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crohn's disease: Tests for therapeutic monitoring of TNF inhibitors (LISA-TRACKER ELISA kits, TNFa-Blocker ELISA kits, and Promonitor ELISA kits)</td>
<td>1616 257 70 23.0857</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the clinical and cost-effectiveness of conservative interventions for elbow tendinopathy?</td>
<td>1029 140 65 15.83077</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease modifying anti-rheumatic drugs</td>
<td>1961 182 130 15.0846</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoaguChek XS point-of-care blood coagulability testing system for those on long-term vitamin K antagonist therapy</td>
<td>658 120 45 14.6222</td>
<td>NB 45 papers from 26 RCTs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The use of fibrin sealant during non-emergency surgery: a systematic review of randomised controlled trials and observational studies</td>
<td>1022 67 45 109.9645</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Implications to Date

• 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

• 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

• “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
