Assessing Computer Searches Used for Systematic Reviews

Paul Fehrmann, Kent State University - Kent Campus
Assessing Computer Searches Used for Systematic Reviews

A paper/presentation for
APA Division 5 - Quantitative and Qualitative Methods

Annual Convention of the American Psychological Association
Toronto, Canada
August 8, 2015

Paul Fehrmann
University Libraries, Kent State University

Author Note

Paul Fehrmann is an Associate Professor, and Reference, Instruction and Collection Development Librarian with responsibilities for Anthropology, Psychology, Sociology, Religion and Philosophy

Correspondence concerning this paper, the CSRC, or this research may be addressed to
Paul Fehrmann University Libraries, Kent State University, Kent, OH 44242  pfehrman@kent.edu
Abstract

Systematic reviews (SR) are valued for increasing the rigor of and reader confidence in “reviews of the literature”. SR have been widely adopted in the health and social sciences. Computer searches are key for the information base of SR, and computer searches must be evaluated. The Computer Search Report Checklist (CSRC) is being developed to evaluate computer search reports.

This paper has 4 objectives:
2. To report comparisons to AMSTAR, PRISMA, PRESS, and others.
3. To report on the inter-rater agreement for CSRC items.
4. To report on using the CSRC to assess evidence for the comprehensiveness and reproducibility of computer searches in SR.

Methods to address paper objectives:
1. and 2. Major SR guidelines and assessment literature were evaluated recommendations used. Author experience was also used.
2. Item inter-rater agreement were evaluated with psychology SR and Campbell Library SR.
3. Psychology SR and Campbell Collaboration SR were evaluated with CSRC items.

Results:
1. A CSRC timeline is presented.
2. Guidance varies for assessing reproducibility and comprehensiveness.
3. Recent Kappas for 10 comprehensiveness items: 6 items moderate-perfect; 1 item poor; 3 items a constant value (so no Kappa).
4. Recent Kappas for 33 reproducibility items: 16 items moderate-perfect; 11 items fair-poor; 6 items a constant value (so no Kappa).
5. Article (2011) in Research Synthesis Methods: comprehensiveness of computer searches in computer search reports of psychology SR and Campbell Collaboration SR.

Key Conclusions:
1. Further testing can verify potential value of CSRC.
2. CSRC might be used as a supplement to existing SR assessment tool.
Assessing Computer Searches Used for Systematic Reviews

This paper is about assessing the computer searches that are used in systematic reviews. In the main, discussion briefly covers the development of the Computer Search Report Checklist (CSRC) and the use of data gathered with the CSRC. Comparisons are made to the key assessment resources AMSTAR, PRISMA, and PRESS.

Systematic reviews (SR) are valued for increasing the rigor of “reviews of the literature”. SR methods also increase reader confidence in the conclusions of reviews. SR are very important in the health and social sciences, and during the last ten to fifteen years the number of SR in psychology has increased considerably. The following PsycINFO search gave the results listed.

• Vendor EBSCO
• Date: July 21, 2015,
• Search terms: No search terms
• Limits: English language, Journal Article Document Type, Systematic review as Methodology
Number of hits: 2000 (17); 2005 (213); 2010 (1001); 2014 (1840).

Understandably, as with all research, the methods quality of SR themselves must be evaluated. A significant part of SR quality assessment involves the literature search, especially the reproducibility and comprehensiveness of the literature search. And, because they are a key part of overall literature search quality, the reproducibility of computer searches (Atkinson, Koenka, Sanchez, Moshontz, & Cooper, 2015; Mullins, Deluca, Crepaz, & Lyles, 2014) and the comprehensiveness of computer searches (Fehrmann & Thomas, 2011; McGowan, Sampson, & Lefebvre, 2010) both need to be evaluated.

The evaluation of computer searches involves assessing computer search reports, because such reports provide the data for evaluating computer searches. The conduct of the computer search is known through the report of the search (Schulz, Moher, & Altman, 2014)?

The Computer Search Report Checklist (CSRC) has been used and is being developed to evaluate computer search reports. Specific goals for the CSRC have been to help professional readers assess the reproducibility and comprehensiveness of computer searches which are used for SR. There are major tools that provide guidance for overall SR assessment; for example, PRISMA (Liberati et al., 2009) and AMSTAR (Shea et al., 2007). A third resource, PRESS, more specifically supports the evaluation of electronic searches (McGowan et al., 2010). However, these three do not provide the detailed guidance of the CSRC with respect to reproducibility or comprehensiveness of CS.


As argued elsewhere (Fehrmann & Thomas, 2011) it seems clear that major organizations who are leaders in the promotion and utilization of SR have been thorough as they publish guidelines for those who undertake or evaluate SR. Thus, the recommendations in those published guidelines are viewed as high quality recommendations that might be translated into checklist items for assessing computer search reports. Along with others, this approach mirrors steps used by the APA Working Group on JARS as they identified and translated content for checklist items (APA Publications Working Group on Journal Article Reporting Standards, 2008).
From 2007 to 2011, as reported previously (Fehrmann & Thomas, 2011), the following organizations/publications were used as resources for computer search report recommendations.

- Cochrane Collaboration Handbook
- Campbell Collaboration Manual
- EPPI-Centre Methods For Conducting Systematic Review
- Systematic reviews: CRD's Guidance For Undertaking Reviews In Health Care
- Working group for the Meta-analysis of Observational Studies in Epidemiology
- Human Genome Epidemiology Network
- The CONSORT Group
- The PRISMA Statement

During 2007–2008, the then current versions of guidelines from these organizations were examined and a list created of non-redundant recommendations. Other items were also developed by the author, based on personal experience in an academic position with responsibilities for database searching in the social sciences. Some of those added items were developed to cover what seemed to be relevant emerging computer search functions.

Given the repeated emphasis on reproducibility and comprehensiveness in the SR methods guidelines literature, an emphasis in CSRC has been on content that pertained to reproducibility or comprehensiveness of searching. The initial CSRC had 47 items. Table 1. (from APA poster) is a general picture of CSRC development.

Table 1
**CSRC Timeline, Development, And Ongoing Work**

<table>
<thead>
<tr>
<th>CSRC Timeline, Development, Ongoing Work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2007 - 08</strong></td>
</tr>
<tr>
<td>- Major SR guidelines were reviewed</td>
</tr>
<tr>
<td>- Non-redundant recommendations translated = 47 items</td>
</tr>
<tr>
<td><strong>2008 - 09</strong></td>
</tr>
<tr>
<td>- Piloted test items</td>
</tr>
<tr>
<td>- Kappas for 36 items = moderate to perfect agreement</td>
</tr>
<tr>
<td><strong>2009 - 11</strong></td>
</tr>
<tr>
<td>- Added two items to original 47 giving 49 items</td>
</tr>
<tr>
<td>- Evaluated psychology SR with comprehensiveness items</td>
</tr>
<tr>
<td>- Article in Research Synthesis Methods</td>
</tr>
<tr>
<td><strong>2012 - 15</strong></td>
</tr>
<tr>
<td>- CSRC reproducibility items used with psychology and Campbell SR</td>
</tr>
<tr>
<td>- Campbell Collaboration Poster: Chicago, 2013</td>
</tr>
<tr>
<td>- Pursuing publication on reproducibility</td>
</tr>
<tr>
<td>- Web site for collaboration and work on additional items</td>
</tr>
</tbody>
</table>

In the work on comprehensiveness and reproducibility noted in the rows 3 and 4, Kappa calculations were used as data on inter-rater agreement for the items. A copy of the 49-item CSRC is available at this site: libguides.library.kent.edu/apacsric
Comparisons to AMSTAR, PRISMA, PRESS and others

As a comparison to CSRC, Table 2. (from APA poster), presents results of the current authors' assessment of contents found in the resources listed.

Table 2
Comparing AMSTAR, PRISMA, PRESS and CSRC

<table>
<thead>
<tr>
<th>Resources</th>
<th>Stated purpose of resource</th>
<th>Stated basis 2 for content of elements or items 4</th>
<th>Approx. number 5 of elements/items for assessing CsR or CsC</th>
</tr>
</thead>
</table>
<pre><code>                  |                                           | Csc = 4/11                                               |
                  |                                           | Csc = 4/27                                               |
</code></pre>
| PRESS (2010)       | To assess CS methods for CS sensitivity and CS precision | Systematic review of literature, survey of experts, and post-survey consult with experts | CsR = 1/30  
                      |                                           | Csc = 4/30                                               |
| CSRC (2008)        | Assess CS reporting         | Consulted SR manuals and guidelines and professional experience | CsR = 36/69  
                      |                                           | Csc = 32/49                                               |

As noted earlier, AMSTAR, PRISMA, and PRESS in column one are key resources for evaluating SR. The dates are publication dates for publications examined for this study that contained the checklists. The first column also notes the CSRC. For the first three resources the second column lists brief statements taken to be the purpose of the resources listed. These statements were derived from the checklist publication description of resource development. Similarly, for the first three resources, the third column information lists methods used for development of the corresponding checklist. The fourth column shows this authors’ assessment of the number of elements or items (in each of the resources) that guide the assessment of computer search reproducibility or computer search comprehensiveness. Comments about elements and items are given next.

The SR literature speaks of recommended search elements that should be reported to support assessment of CS. Search elements are search steps used in SR. This reported element information can be used to guide SR assessment. In particular, the information can be used for the assessment of computer searches. The SR literature also speaks of checklist items that can be used to guide assessment of SR. In the current paper, as shown in Table 2 column three, the term element is used to stand for a recommended CS characteristic that should be reported or evident as used in SR. The term item is used to refer to an explicitly worded statement or question that represents or addresses all or part of an recommended CS element.
Information in the fourth column of Table 2, for CSRC indicates the following. The CSc (comprehensiveness) data is based on a set of items developed and used in 2010. CSr (reproducibility) information is based on CSRC item versions developed and used in 2013.

Other papers are also relevant as resources for assessing SR, and for assessing CSr and CSc. Those papers could also be included in this kind of comparison, and be included in a set of sources for guidance on checklist items (Atkinson et al., 2015; Maggio, Tannery, & Kanter, 2011; Mullins et al., 2014; Niederstadt & Droste, 2010; Rader, Mann, Stansfield, Cooper, & Sampson, 2014; Yoshii, Plaut, McGraw, Anderson, & Wellik, 2009). A well-known site dedicated to reporting is the EQUATOR Network (Moher, Simera, Schulz, Hoey, & Altman, 2008).

Items from the 49 item version of the CSRC were used to collect data for testing inter-rater agreement. Select items from that version also were used to evaluate the evidence for comprehensiveness in computer searching in sets of SR published in psychology and in sets of SR from the Campbell Library (www.campbellcollaboration.org/lib). More recently a new set of the items pertaining to reproducibility have been tested and used on new sets of psychology SR and Campbell Library SR. Sample data for comprehensiveness and reproducibility are presented in the next section.

**Data on inter-rater agreement of CSRC items and data on frequency of reporting for recommended computer search information**

Table 3 is a screen shot of the poster at APA, and shows sample data from a study of inter-rater reliability. It contains sample data for 5 of the questions that focused on comprehensiveness information reported in the computer search report.

<table>
<thead>
<tr>
<th>Items Assessing Comprehensiveness Of Computer Searches</th>
</tr>
</thead>
</table>

| Computer Search Comprehensiveness Items | Item Frequencies | Inter-Rater Agreement (Kappa/| sig level) |
|-----------------------------------------|------------------|------------------------------|
| 1. Are more than one resource listed?   | 44/50            | .81 (.001)                   |
| 2. Does the report show or state that synonyms were used as search terms? | 47/50 | .34 (.015) |
| 3. Does the report indicate that truncation or wildcard searching was used? | 12/50 | .94 (.001) |
| 4. Does the report indicate that controlled vocabulary or thesaurus terms were used? | 12/50 | .73 (.001) |
| 5. Does the report state or show the use of resource or search function to find publications that have cited a known article (known author, etc.)? | 1/50 | 1.00 (.001) |

*Note. Key to superscripts.*
1. Evaluating reports of computer searches to see if recommended information is reported.
2. Used to assess a set of 50 Psychology SR in 2010.
3. Number of “Yes” scores out of total SR assessed, based on consensus after calculating Kappa.
As reported elsewhere (Fehrmann & Thomas, 2011), the frequency data were developed based on analyzing 50 SR randomly selected from 363 hits resulting from a PsycINFO search targeting SR. The Kappa numbers in this table were also calculated for that study.

More recently a similar study looked at reproducibility. Table 4. (from APA poster) shows sample data. Note the two items (or levels), a. and b., for each of the 5 elements/items.

Table 4
*Items Assessing Reproducibility Of Computer Searches*

<table>
<thead>
<tr>
<th>Computer Search Reproducibility Items 4.</th>
<th>Item Frequencies 3.</th>
<th>Inter-Rater Agreement (Kappa/ sig level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does the report state or show that it is providing a copy of the computer search strategy as a saved/downloaded/copy-pasted &quot;search history&quot; from the actual computer search session?</td>
<td>a. 1/25 1.00 (.000) b. 1/25 1.00 (.000)</td>
<td></td>
</tr>
<tr>
<td>a. For any resource b. For every resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Are names given for the computer search resources used?</td>
<td>a. 24/25 1.00 (.000) b. 24/25 .65 (.001)</td>
<td></td>
</tr>
<tr>
<td>a. For any resource b. For every resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Is a publication time frame listed?</td>
<td>a. 20/25 .39 (.014) b. 19/25 .39 (.014)</td>
<td></td>
</tr>
<tr>
<td>a. For any resource b. For every resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Does the report state (or show or make obvious) that all of the exact search terms (words or phrases) that were used are actually listed?</td>
<td>a. 22/25 .34 (.065) b. 22/25 .28 (.112)</td>
<td></td>
</tr>
<tr>
<td>a. For any resource b. For every resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Does the report explicitly state or show that it is listing how all of the search terms used were combined?</td>
<td>a. 18/25 .23 (.075) b. 18/25 .23 (.075)</td>
<td></td>
</tr>
<tr>
<td>a. For any resource b. For every resource</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Key to superscripts.
3. Number of “Yes” scores out of total SR assessed, based on consensus.

**Conclusion comments and recommendations**

Systematic reviews (SR) have become very significant in psychology, and the methods used for published SR need to be evaluated. Although tools that focus on evaluating SR do include guidance for evaluating computer searches, it does seem that additional details in a tool such as the Computer Search Report Checklist (CSRC) could have value for researchers, reviewers and consumers of SR. Further testing with the CSRC and/or its items would help to verify potential value. The CSRC might be used as a supplement to existing SR assessment tools such as AMSTAR, PRISMA and PRESS. Collaborations looking at the CSRC could lead to enhanced development.
1. External input on CSRC design and value should be gathered (possibly a survey).
2. Additional reliability and validity should be pursued.
3. Options to facilitate practical use should be developed (downloadable copies, online versions, etc.)
4. Comparative studies should be pursued, for example comparing CSRC to PRESS (a tool designed to support evaluation of electronic search strategies).
References


