Basic Steps and Procedures for a Campbell Systematic Review

Terri Pigott, Loyola University Chicago

KTDRR and Campbell Collaboration Research Evidence Training Session 2: March 20, 2019
Goals for the Presentation

- Provide an overview of the stages of a systematic review and meta-analysis.
- Provide overview of standards for producing a Campbell systematic review using MECCIR.
Campbell Collaboration MECCIR

- The Methodological Expectations of Campbell Collaboration Intervention Reviews (MECCIR) provide authors and users of Campbell reviews of intervention effects with clear and transparent expectations of review conduct and reporting.

- They are based on the original work of the Cochrane MECIR project team (Higgins et al., 2016)

- This presentation provides highlights from these standards for the conduct of a systematic review.

- Campbell’s MECCIR standards can be accessed here.
Problem Formulation

- What kinds of questions can be addressed in a systematic review?
- What is the scope of a systematic review?
- How does the review question shape the inclusion and exclusion criteria for studies included in the review?
Systematic reviews are a form of research

- Systematic reviews follow the same basic steps as any research study (Cooper, 2016) but with primary studies as the unit of analysis.

- Stages of a systematic review:
  1. Problem formulation
  2. Data collection
  3. Data evaluation
  4. Data analysis and interpretation
Stages of a Systematic Review

1. Problem Formulation
Systematic Review (SR) Questions: Effectiveness of interventions


This study is a systematic review of the effectiveness of academic interventions for elementary and middle school students with low-socioeconomic status.
SR Questions: Effectiveness of interventions (cont’d)

- Research questions for Dietrichson et al. (2017):
  - What are the effects of $x$ intervention on $y$ outcome for $z$ population?
  - Variations on this theme (e.g., differences in effects of interventions $x_1$ vs. $x_2$)
  - **Our example**: Effects of academic interventions for low-SES students
SR Questions: How do two groups compare?

- How does group X differ from group Y on some characteristic?

- **Example 1**: Gender differences in leadership, math performance, hours worked in the home, etc.
  

- **Example 2**: Differences between first- and second-generation immigrant students on academic achievement
  
SR Questions: Associations among constructs


- This systematic review and meta-analysis estimates the association between cyber-victimization and academic outcomes in students aged 12–17.
SR Questions: Associations among constructs (cont’d)

- Research questions in Gardella et al. (2017):
  - How does $x_1$ relate to $x_2$ for population $z$? (direction and strength of correlation)
  - Variations on this theme (e.g., differences in relation of $x_1$ and $x_2$ between populations $z_1$ and $z_2$)
  - Our example: Estimating the associations between peer cyber victimization (PCV) and educational outcomes among U.S. adolescents
Summary: Types of SR Questions

- **Intervention effects:** Does program A have larger effects than treatment as usual?
- **Comparison of groups:** Does group A differ from group B on some outcome?
- **Associations between two constructs:** What is the strength and direction of the association between two constructs?
- **Diagnostic and prognostic tests:** Which test is more accurate at diagnosis? Which test is better at prediction?
- **Prevalence:** What is the prevalence of some phenomenon?
Systematic reviews can vary in scope

- Specific, narrow questions
  - Useful for testing effects of specific treatments

- Broad, global questions
  - Useful for generating new knowledge
    - Identify common elements of effective programs.
    - Build better intervention theories to guide program development and evaluation design.
Examples of Narrow and Broad Questions

- In the case of an intervention, we might be interested in understanding the use of a particular intervention such as Second Step, an intervention for developing socio-emotional skills in elementary school children—a narrow question.

- A broader question: What are the impacts of interventions that are focused on developing socio-emotional skills in elementary school children?

- The scope of the question will guide many aspects of the review.
The SR question guides decisions in the review

- The critical characteristics of studies included in the review are guided by the systematic review question
  - Participants
  - Types of interventions, if applicable
  - Study design
  - Outcomes
Problem formulation guides the study inclusion criteria: PICOS framework

- The **PICOS** framework:
  - **P**opulation/Participants (problems/conditions)
  - **I**nterventions (if applicable)
  - **C**omparison group (e.g., absolute vs. relative effects, counterfactual conditions), if applicable
  - **O**utcomes (primary and secondary outcomes, acceptable outcome measures)
  - **S**tudy Design (should be fit for purpose) or **S**ettings
Example: Dietrichson et al. (2017)

- **P**: Low-income students in elementary and middle school
- **I**: Interventions explicitly aimed to improve educational achievement that could be implemented by schools
- **C**: Non-treatment, control condition
- **O**: Standardized academic tests
- **S**: Randomized controlled trial or controlled quasi-experimental study
Summary: Problem Formulation

- As is true in any research study, the research question guides all other decisions made about methods in the study.
- Systematic review questions can involve questions about differences among groups, associations among constructs, prevalence.
- Systematic review questions vary in their scope:
  - Narrow questions about a particular intervention
  - Broad questions designed to understand a literature on a topic (IES Goal 1 questions)
- Systematic review is most productive when there is a large body of literature.
- The research question leads to PICOS—the inclusion criteria for studies eligible for the systematic review.
Stages of a Systematic Review

1. Problem Formulation
2. Data Collection
Data Collection: Searching the Literature

- **Goal:** Collect all primary studies examining the research question for the systematic review and that meet the inclusion criteria
- **Unit for the SR:** primary studies
- **Generalizing to a hypothetical population of studies on a given research question**
- **Realistic goal:** Use a search strategy that can support the argument that the SR has a representative sample of studies

**Source:** Kugley et al. (2017)
Data Collection: Locating Studies

- Need both sensitive (broad) and specific (focused) searches
  - Google Scholar sensitive but not specific
  - Databases like ERIC more specific
- Use of keyword searches of multiple electronic databases—and knowledge of the keywords used in that database
- Search of multiple sources of grey literature (websites, dissertations, special registers, reference lists of included studies and reviews, listservs, personal contacts)
- Hand-searching of selected journals
Databases to Search

- Important issues in disability research are multidisciplinary—researchers in education, health, medicine, psychology, sociology, and many other disciplines are interested in these questions.

- Systematic reviews in disability research need to search multiple databases to find all relevant studies.

- Appendix A in Campbell’s information retrieval guide (Kugley et al.) provides a list of potential databases to consider.
Search Strategies

- Generally, a search strategy to identify intervention studies will typically have three sets of terms: (1) the **condition of interest**, i.e., the population; (2) the **intervention(s)** evaluated; and (3) the **outcomes**.

- Each database may use different terms for the same phenomenon, so it is important to understand the concepts in each discipline.

- The formulation of search terms will include controlled vocabulary, keywords, Boolean operators and limiters.

- Consult Campbell’s information retrieval guide (Kugley et al., 2017) for more details [https://www.campbellcollaboration.org/library/searching-for-studies-information-retrieval-guide-campbell-reviews.html](https://www.campbellcollaboration.org/library/searching-for-studies-information-retrieval-guide-campbell-reviews.html)
Appendix A. ERIC Search Example

(Prevention AND “Sexual violence” AND Experiment*) OR
(Prevention AND “Sexual violence” AND Quasi-Experiment*) OR
(Prevention AND “Sexual coercion” AND Experiment*) OR
(Prevention AND “Sexual coercion” AND Quasi-Experiment*) OR
(Prevention AND “Peer support” AND Quasi-Experiment*) OR
(Prevention AND “Intimate Partner violence” AND Experiment*) OR
(Prevention AND “Intimate Partner violence” AND Quasi-Experiment*) OR
(Prevention AND “Bystander” AND Experiment*) OR
(Prevention AND “Bystander” AND Quasi-Experiment*) OR
(Prevention AND “Dating Violence” AND Experiment*) OR
(Prevention AND “Dating Violence” AND Quasi-Experiment*) OR
(Prevention AND “Physical Violence” AND Experiment*) OR
(Prevention AND “Physical Violence” AND Quasi-Experiment*) OR
(Prevention AND “Dating Aggression” AND Experiment*) OR
(Prevention AND “Dating Aggression” AND Quasi-Experiment*) OR

(Intervention AND “Sexual violence” AND Experiment*) OR
(Intervention AND “Sexual violence” AND Quasi-Experiment*) OR
(Intervention AND “Sexual coercion” AND Experiment*) OR
(Intervention AND “Sexual coercion” AND Quasi-Experiment*) OR
(Intervention AND “Peer support” AND Quasi-Experiment*) OR
(Intervention AND “Intimate Partner violence” AND Experiment*) OR
(Intervention AND “Intimate Partner violence” AND Quasi-Experiment*) OR
(Intervention AND “Bystander” AND Experiment*) OR
(Intervention AND “Bystander” AND Quasi-Experiment*) OR
(Intervention AND “Dating Violence” AND Experiment*) OR
(Intervention AND “Dating Violence” AND Quasi-Experiment*) OR
(Intervention AND “Physical Violence” AND Experiment*) OR
(Intervention AND “Physical Violence” AND Quasi-Experiment*) OR

Example of a search strategy:
De La Rue et al. (2014)
Data Collection: Searching the Literature

Best Advice: Get a librarian

This photo by unknown author is licensed under CC BY-SA.
Data Collection: Issues to Consider

- Need methods to document the process from start to finish
  - All information sources searched; dates covered in search and date of search
  - Full electronic search strategy including limits and keywords for replication of search
- Searching may be an iterative process while defining key concepts and discovering appropriate search terms
- Use of software to manage the process (Endnote, RefWorks, Zotero)
Zotero example: https://www.zotero.org/
Zotero example (continued)
Data Collection: Summary

- Importance of the information retrieval/search process
  - Not a “one-shot” deal—usually involves several iterations of searches
  - Requires expertise in the planning and implementation of searches
  - Should consult with a librarian/search specialist

- Use of bibliographic management software to store, manage, and organize results

- Documentation of all steps so that search can be replicated
Stages of a Systematic Review

1. Problem Formulation
2. Data Collection
3. Data Evaluation
Data Evaluation: Screening and Coding

Once a search is “completed,” the next steps are

• To identify the studies that meet inclusion criteria
• To code information from included studies to use in the systematic review and meta-analysis
Data Evaluation: Screening of Titles and Abstracts

- Screening for relevant studies usually takes place in two waves.
  - Screening based on reading titles and abstracts
  - Used to exclude obviously irrelevant articles, such as opinion pieces or nonempirical studies
  - Titles and abstracts are notoriously unreliable, so first screening tends to exclude nonempirical studies
  - Best practice to double-code using two trained raters working independently
  - Machine learning strategies for screening (Abstrackr)
<table>
<thead>
<tr>
<th><strong>Abstrackr</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Abstrackr Logo" /></td>
</tr>
</tbody>
</table>

Software for (semi-automated?) abstract screening for systematic reviews. At present, *abstrackr* is a free, open-source tool for facilitating the citation screening process. Upload your abstracts, invite reviewers, and get to screening!

We are currently integrating machine learning technologies to semi-automate the screening process. Already, *abstrackr* will prioritize the screening of those articles most likely relevant to the review at hand. In the near-future, it will screen out irrelevant citations for you, automatically.

The tool.

The source code.
## Typical Numbers of Studies Located in Initial Searches

<table>
<thead>
<tr>
<th>Review</th>
<th>Initial Number of Studies</th>
<th>Number After First Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietrichson et al. (2017)</td>
<td>11,807</td>
<td>1,137</td>
</tr>
<tr>
<td>Duong et al. (2016)</td>
<td>1,478</td>
<td>238</td>
</tr>
<tr>
<td>Gardella et al. (2017)</td>
<td>6,378</td>
<td>508</td>
</tr>
<tr>
<td>Littell et al. (2005)</td>
<td>5,290</td>
<td>266</td>
</tr>
<tr>
<td>Voyer et al. (2014)</td>
<td>17,307</td>
<td>Unclear</td>
</tr>
</tbody>
</table>
Data Evaluation: Full-Text Eligibility Screening

- After initial screening of titles and abstracts, move on to screening full texts of potentially eligible studies.
- Reviewers develop list of inclusion and/or exclusion criteria based on SR research question following PICOS.
- Studies are screened with an accounting of ineligible studies and the reasons for their exclusion.
- Best practice: blind, double-coding.
- Recruitment and training of coders important.
You can find the PRISMA guidelines here:
http://www.prisma-statement.org/
Data Evaluation: Coding

- With complete set of eligible studies, next step is coding study content and effect sizes

- Develop a coding manual that includes:
  - Setting, study context, authors, publication date, and type
  - Methods and method quality
  - Program/Intervention and comparisons
  - Participants/Clients/Sample
  - Outcomes
  - Findings, effect sizes
Example: Coding Manual

Section C — Sample & Program Characteristics
Treatment Sample & Program Characteristics
C01. What was the SES of the students in the sample?
(Please indicate, with a 1, all that apply).
1. Low SES
2. Low-middle SES
3. Middle SES
4. Middle-upper SES
5. Upper SES
6. Only labeled as “mixed”
9. Can’t tell

Source: De La Rue et al. (2014). C01, p. 38; E04 and E05, p. 42.

Dependent measure (separate for each measure)
E04. What does the outcome measure?
   1- Teen dating violence knowledge
   2- Teen dating violence attitude
   3- Sexual violence knowledge
   4- Sexual violence attitude
   5- Rape construct (awareness, myths, etc.)
   6- Bystander awareness
   7- Other (specify)

E05. How was the outcome measure reported?
   1- Self-report
   2- Teacher report
   3- Administrator report
   4- Other (specify)
Data Evaluation: Coding Rationale

- Provide descriptive detail of studies included in systematic review
  - To understand the “landscape” of the evidence base
  - To identify gaps or issues in the evidence base
- Record information about study effects and methods to explain differences in effect size
Coding of Study Quality

- Variety of options for coding study methods
  - Cochrane risk of bias framework
  - GRADE system
  - Method quality checklists
  - Direct coding of methodological characteristics
- Used in meta-analysis to examine how results differ by study quality
## Cochrane Risk of Bias Framework 1.0 (for RCTs)

### Table 8.4.a: A common classification scheme for bias

<table>
<thead>
<tr>
<th>Type of bias</th>
<th>Description</th>
<th>Relevant domains in the Collaboration’s ‘Risk of bias’ tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection bias</td>
<td>Systematic differences between baseline characteristics of the groups that are compared.</td>
<td>• Sequence generation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Allocation concealment.</td>
</tr>
<tr>
<td>Performance bias</td>
<td>Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.</td>
<td>• Blinding of participants and personnel.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Other potential threats to validity.</td>
</tr>
<tr>
<td>Detection bias</td>
<td>Systematic differences between groups in how outcomes are determined.</td>
<td>• Blinding of outcome assessment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Other potential threats to validity.</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Systematic differences between groups in withdrawals from a study.</td>
<td>• Incomplete outcome data</td>
</tr>
<tr>
<td>Reporting bias</td>
<td>Systematic differences between reported and unreported findings.</td>
<td>• Selective outcome reporting (see also Chapter 10).</td>
</tr>
</tbody>
</table>

Table 8.4.a
Source: Higgins & Green, 2011
Newcastle-Ottawa Scale (NOS) for Non-randomized Studies

- Focuses on quality assessment for case-control and cohort studies
- Assesses selection of the experimental groups and the comparability of the groups
- For case-control studies, assesses exposure to treatment
- For cohort studies, assesses outcomes

Source: Wells et al. (n.d.)
http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
Coding: Issues to Consider

- Reviewers develop coding manuals for reviews
- The most time-intensive portion of a systematic review
- Best practice: Use of software to conduct and organize coding
  - Database software such as Access or FileMaker
  - Spreadsheets such as Excel or Google Sheets
- Attention to training of coders, double-coding, recording of reliability of coding decisions
Common Challenges in Data Evaluation Stage

- Underestimating the time needed to prepare and code studies
- Assuming that developing a coding protocol is not an iterative process (it is!)
- Coding drift—important to check that all coders are interpreting the items the same way
- Organizing and supervising coding process
  - Personnel issues
  - Appropriate software
Stages of a Systematic Review

1. Problem Formulation
2. Data Collection
3. Data Evaluation
4. Data Analysis and Interpretation
Data Analysis

Once eligible studies are coded, the analysis consists of two stages:

1. Description of the studies included, usually in narrative and table form
2. Meta-analysis of effect sizes, when possible
   - Examination of the average effect size and its confidence interval
   - Exploration of heterogeneity of study results

(Narrative discussion of study results if meta-analysis is not possible.)
Data Analysis: Effect Sizes

- In a meta-analysis, we express the results of each study using a quantitative index of effect size (ES).
- ES is a measure of the strength or magnitude of a relationship of interest.
- ESs have the advantage of being comparable (i.e., they estimate the same thing) across all of the studies and therefore can be summarized across studies in the meta-analysis.
- NOTE: Session 5 on “Effect Size and Meta-Analysis” is scheduled for June 19, 2019.
9.2.1 Teen Dating Violence Knowledge

Source: De La Rue et al. (2014), p. 89.

Notes: Study-level effect sizes and 95% confidence intervals; Positive effect sizes indicate knowledge improvement for the treatment group.
Data Analysis: Examining Heterogeneity

- As in any statistical analysis, the mean and its associated standard error may not be the best descriptive statistic for a distribution of data.
- In education, we expect that study results will vary.
- An important part of any meta-analysis is examining the amount of heterogeneity among effects.
Data Analysis: Exploring Heterogeneity

- With a sufficient number of studies, we can explore potential correlates of heterogeneity among effect sizes.
- With one categorical predictor, can use simple, one-way ANOVA models.
- With multiple predictors, can use meta-regression.
- Common predictor of heterogeneity: study quality.
Example: Sensitivity of Results to Study Quality

| Table 4.6: Moderator Analysis Using Method of Assignment to Condition for All Outcomes for Each Type of Effect |
|-------------------------------------------------|-------------------------------------------------|-------------------|------------------|
|                                                  | Immediate Post-test                             | Follow-Up         |
|                                                  | Random Assignment                              | Non-random Assignment | Random Assignment | Non-random Assignment |
| Teen Dating Violence Knowledge                   | .36** (.13, .59)                               | .09 (-.12, .30)    | .24 (-.45, .93)  | -.13 (-.72, .45)     |
| Teen Dating Violence Attitudes                   | .12** (.06, .18)                               | .19** (.11, .29)   | .13 (-.02, .27)  | -.09 (-.19, .38)     |
| Rape Myth Awareness                              | -.46** (-.78, -.15)                            | -.52 (-1.09, .05)  | NA               | NA                 |

Source: De La Rue et al. (2014), p. 45.
Common Challenges in Effect Size Computation and Synthesis

- Missing data in primary studies for computing effect sizes, particularly in older studies
- Appropriate effect sizes for complex designs in primary studies, i.e., clustered RCTs, regression models
- Combining effect sizes from different metrics
- Multiple effect sizes within studies: dependent effect sizes
- Often need to consult with a meta-analysis specialist
Summary: Data Analysis and Interpretation

- **Not** the most time-intensive part of the systematic review
- Exploration of heterogeneity of greatest importance
  - Expect variation in effects across studies
  - Careful planning for moderator analysis, both in coding stage and in analysis
- Care needed in interpretation of these exploratory results
- Rapid development of statistical methods may require consultation with a statistician
Narrative Synthesis

- Narrative synthesis methods may be needed when
  - There are few or no studies identified for the review.
  - The review question is focused on configuration rather than aggregation.
- Narrative and qualitative synthesis methods exist and also use transparent and replicable methods.
- Campbell is beginning to develop standards for these types of reviews.
References


References (Cont’d)


Thank you!

Please take a few minutes to respond to the brief Evaluation Survey:


Terri Pigott: tpigott@luc.edu