Introduction to Systematic Reviewing and Meta-Analysis

Jeff Valentine
University of Louisville, USA

jeff.valentine@louisville.edu

Overview

• Introduction to systematic reviews
  – Definition and characteristics
  – Comparison to narrative reviews
  – Brief history of systematic reviewing and meta-analysis
• Cooper’s stages of the research review
  – Problem formulation
  – Data collection
  – Data evaluation
  – Data analysis and interpretation
  – Reporting the synthesis
Overarching goals in systematic reviewing

- **Limit bias** in literature reviews by using:
  - a systematic approach to data collection, analysis, and evaluation
  - both published and unpublished research (minimizing publication bias)
- **Enhance methodological rigor** by:
  - holding reviewers to high scientific standards
- **Employ better and more appropriate statistical tools** by:
  - avoiding decision rules with poor inferential characteristics
  - using statistics that are appropriate for the analysis of collections of studies

Example of confirmation bias

**Primary Research** (Burk, Henggeler, & Whelan, 1987)

- Data collected (36 subscales)
- Results reported (29 subscales)
- Data provided (19 subscales)

**Second Generation Reviews** (phrases)

- Born et al. (2003) "improved parent-child relations"
- Cortina (2000)
- Curtis et al. (2004) avg ES (multiple outcome measures) d=-1.32 (sd=-.65)
- Henggeler et al. (2002) "improved parent-child interactions"
- Henggeler & Shelton (2003) "successful...outcome"
- Hoagwood et al. (2001) "efficacy...has been...demonstrated"
- Katz (1999)
- Keenan & Weaver (1996) "effects have been replicated"
- Pushak (2002) "a promising treatment"
- Sorenson & Rowland (2002) "supported the promise"
- Swanson & Henggeler (2003)

**Legend**

- Red arrows: favors other group
- Green arrows: favor MST
- Red square: result not reported
- NS: no significant difference
What is a systematic review?

A *systematic review* is a summary of the research literature that
• uses explicit, replicable methods to identify relevant studies, and then
• uses objective techniques to analyze those studies.

The goal of a systematic review is to limit bias in the identification, evaluation and synthesis of the body of relevant studies that address a specific research question.

What is a systematic review?

Common labels include ...
• research synthesis,
• research review,
• systematic review,
• quantitative review, and
• meta-analysis.

Some scholars use the term “meta-analysis” to refer to only the quantitative summaries used in a systematic review, others use it more broadly.
Systematic review versus meta-analysis

• A systematic review need not include a meta-analysis.
• A meta-analysis need not be based on a systematic review (though often it should be!)

What is a systematic review?

Compared to traditional procedures, systematic reviews...
• use a more systematic approach to data collection,
• have more transparent and explicit procedures,  
  – thus have less potential for hidden assumptions or biases to drive results, and
• focus on the magnitude of effects rather than statistical significance
• result in conclusions that are less bound by context than conclusions that arise from individual studies
Traditional synthesis procedures typically ...

- rely on a convenience sample of studies,
- use decision rules that are not transparent,
  - e.g., rarely state *a priori* what constitutes “good” evidence,
- rely on statistical significance to gauge what a study “says”, and
- may use a “vote count” of results to draw conclusions about the literature as a whole.

What is wrong with narrative reviews?

- “rely on statistical significance to gauge what a study says”
  - problematic because statistical significance is a function of two things:
    - size of the study
    - magnitude of the effect
    - unimportant effects can be statistically significant, and important effects can be not statistically significant
  - more on this in the afternoon session
What is wrong with narrative reviews?

- “may use a ‘vote count’ of results to draw conclusions about the literature as a whole”
  - If a study found a statistically significant positive effect, a vote was tallied in that box
  - If a study found a statistically significant negative effect, a vote was tallied in that box
  - If a study failed to reject the null hypothesis, a vote was tallied in that box
The Trouble with Vote Counting

• Vote counts ignore Type II error
  – Given that most studies are conducted with low statistical power
    • Statistical power = the probability of correctly rejecting a false null hypothesis
    • Low power assures that the “ns” category will be the biggest
  – Also based on a misconception about the nature of study results (the sampling error problem)
    • More on these problems during the afternoon session

What is wrong with narrative reviews?

Different experts may perform a narrative review on the same question and come to different conclusions.

• Sometimes this is because they review different sets of studies.
• Even when the same studies are reviewed, the process of integrating them in a narrative review is subjective.
• While systematic reviews also may reach different conclusions, it should be possible to discern why they did.
What is wrong with narrative reviews?

Narrative reviews have a hard time tackling variations in treatment effectiveness.

- As the number of studies grows, studies often examine different populations.
- The size of the relationship of interest may vary in different populations (i.e., there may be moderator effects).
- The narrative reviewer, who has enough trouble summarizing studies when they are all done in similar ways, now has a much harder task.

Why are systematic reviews better?

Systematic reviews are better than narrative reviews because ...

- they are more objective and more replicable,
- they are able to deal efficiently with large amounts of information,
- they can systematically examine variations in treatment effectiveness or relationship strength,
- they can examine differences due to study methods, types of participants, settings, etc.
Why are systematic reviews better?

• Research often produces results that appear contradictory, but that vary due to moderator variables (populations, treatments, settings, methods, designs); systematic reviews help us understand how these influence results.

• Studies with different strengths and weaknesses can be used to rule out each others’ alternative explanations.
  – Similar results from studies with different designs lead to greater confidence in conclusions.

A systematic review should contain ...

• a clearly defined, explicit question,
• a comprehensive and systematic search for studies,
• an explicit, reproducible strategy for screening and including studies,
• explicit, reproducible data extraction (coding),
• appropriate analysis and reporting of results,
• interpretations supported by data, and
• implications for future research, and if relevant, for policy or practice.
A C2 systematic review should …

- have all of these elements and
- be based on a predefined and pre-approved protocol that guides the work of the review team
  - peer reviewed by both substantive and methodological experts

Writing a protocol

A protocol ...
- forces you to read and understand the background of the problem of interest,
- makes you formulate a focused question,
- makes you plan your information retrieval strategy,
- makes you think through and describe inclusion/exclusion criteria clearly, and
- prompts you to think about the data you want to collect and how you will analyze them.
- helps protect against data-driven decisions
Writing a protocol

- Examples of protocols are available at: [http://www.campbellcollaboration.org/ECG/proto.asp](http://www.campbellcollaboration.org/ECG/proto.asp)
- A module on protocol development is also offered by C2

Brief History of Systematic Reviewing and Meta-Analysis
Scientists Have Been Thinking About How to Integrate Studies for a Long Time

- James Lind, English naval surgeon (18th Century):
  “...it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy ... by which the sources of these mistakes may be detected.”


  “…it sometimes happens that although few or [no statistical tests] can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are lower than would have been obtained by chance.” (p.99, emphasis added)


Resurgence in the 1970’s

- Explosion of research since the 1960’s
  - About 100 randomized experiments in medicine per year in the 1960’s
  - About 13,000 randomized experiments in medicine per year today
  - Right now in PsychInfo (an electronic database housing academic publications):
    - Over 21,000 references to self-esteem
    - Over 84,000 references to depression
    - Over 41,000 references to motivation
    - Over 20,000 references to aggression
Necessity is the Mother of (Re)-Invention

Plato


  - Over 700 estimates

  - Over 800 estimates

Current Status

- Voluminous literature on research synthesis and meta-analysis
  - Some confusion about SRs
- Scholars organized international, interdisciplinary groups to promote and expand interest in systematic reviewing:
  - Cochrane Collaboration (C1, named for Archibald Cochrane)
  - Campbell Collaboration (C2, Donald Campbell)
  - Society for Research Synthesis Methodology
Issues

• Conceptual issues in attempts to combine results of studies that vary (to some degree) in their methods, treatments, samples, outcome measures
  – Apples, oranges, and other fruits
  – Parallel problems in studies of individuals (there is no “average” person)
• Given variation in primary research
  – What should be included in a synthesis?
  – How should we synthesize results?

Components of a Good SR

• Protocol (plan) for SR developed in advance
• Systematic attempts to minimize bias at each step in the review process
• Explicit inclusion/exclusion criteria
• Comprehensive search strategy
• Inter-rater agreement on key decisions
• Transparent methods (as much as possible)
• Document decisions
• Meta-analysis when possible
Standards for SRs

- There are good frameworks and standards for systematic reviews
  - Cochrane Collaboration Handbook
  - Campbell Collaboration policy briefs
    - Both available online
    - Note that there is some variation between and within groups (both policy and practice)
  - *Practical Meta-Analysis* by M. Lipsey and D. Wilson

Standards (cont’d)

- QUOROM statement (Quality Of Reporting On Meta-analysis) will soon be updated
  - PRISMA (Principles of Reporting Information on Systematic reviews and Meta-Analysis)
  - Moher et al.
Steps in the SR process

SRs are a form of research
Follow basic steps in the research process (Cooper):
• Problem formulation
• Sampling
  – Studies are the sampling unit
  – Sampling frame = all relevant studies
  – Sample = studies available for analysis
• Data collection
  – Data derived (extracted) from studies
• Analysis
  – Qualitative (descriptive, study quality assessment)
  – Quantitative (vote counts/box scores, effect sizes, meta-analysis)
• Reporting

Stages of a research synthesis
The seminal article outlining five stages of a research review is by Harris Cooper:

- This material is also covered in Cooper’s books on research synthesis.

Stages of a research synthesis (Cooper, 1982)

- Problem formulation
  - Clarifying your questions and writing a protocol

- Data collection
  - Literature search
  - Information-gathering from studies

- Data evaluation
  - Criteria for including and excluding studies
  - Assessing study quality
Stages of a research synthesis (Cooper, 1982)

- Data analysis and interpretation
  - Integrating the effects from collected studies
  - Interpreting analysis results

- Report preparation

Stages of a research synthesis

- Most of the work involved in conducting a review is not spent in statistical analysis.

- The scientific contribution of the final product is dependent on all stages of the review and not just the statistical analysis stage.
Time required for a research synthesis

• Allen and Olkin documented the time spent on tasks in 37 meta-analyses conducted by a private company specializing in meta-analysis.
  – Median time spent was 1110 hours.
  – Range was 216-2518 hours.
• Then they graphed total number of citations initially retrieved against total time required to do the meta-analysis.

---

**Figure.** Citations Retrieved for a Meta-analysis and Total Hours Required to Complete the Meta-analysis

Total Time = 721 + 0.243x - 0.000123x^2

From Allen & Olkin (1999)
Time required for a research synthesis

\[ \text{Total Time} = 721 + 0.243x - 0.0000123x^2 \]

- The intercept of 721 reflects the fact that some constant time commitments are required for any meta-analysis.
- We add about 15 minutes for each citation retrieved (.24 hours), though slightly less time is needed as more and more citations are found.

Characteristics of each stage

Each stage of a synthesis can be better understood by considering the following questions:

- What research question is asked at this stage?
- What is the primary function of the stage in the review?
- What procedural differences create variation in review conclusions?
- What are the potential threats to the validity of review conclusions?
An example of the stages of a review

I will illustrate each of the stages by considering a meta-analysis of the treatment effects of stress-management programs:


Stage 1

Problem Formulation
Problem formulation

• Research question asked
  – What research evidence will be relevant to the problem or hypothesis of interest in the synthesis?
  – *In our example, the authors sought studies of the effects of occupational stress management interventions on individual and organizational outcomes.*

Problem formulation

• Primary function served in the synthesis
  – Define the (a) variables and (b) relationships of interest so that relevant and irrelevant studies can be distinguished.
  – *Treatment had to be an occupational stress management intervention.*
  – *Outcomes could have been either individual level (psychological or somatic) or organizational outcomes.*
What kinds of research questions can be asked?

• Questions about **associations**:
  – How does $x_1$ relate to $x_2$ for population $z$? Both direction and strength of relation are of interest.
  – Variations on this theme (e.g., differences in relation of $x_1$ and $x_2$ between populations $z_1$ and $z_2$).

• **Diagnostic/prognostic** questions:
  – Which test (A vs. B) is a better predictor of $y$?
  – Which test is a better predictor of $y$ for population $z_1$ vs. $z_2$?

What kinds of research questions can be asked?

• **Focus and scope-of-effectiveness** questions:
  – $X$ may represent an intervention (with narrow vs. broad definitions, etc.),
    • compared to what? (type of control group, counterfactual conditions)
  – $Y$ represents the outcome(s) (single/multiple, types of measures, length of follow-up period may vary, etc.)
  – $Z$ represents the populations or problems -- for which ones does the treatment work? (specific vs. general; inpatients vs. others, etc.)
Problem formulation

• **Procedural variation** that might produce differences in conclusions
  – Variation in the conceptual breadth and detail of definitions might lead to differences in the research operations (a) deemed relevant and/or (b) tested as moderating influences
  – *A key question was whether the effects varied by type of intervention or outcome.*
  – *Type of participant was of interest (e.g., office workers vs. paramedics); participants must have had no diagnosed psychiatric disorder.*

Checklist for problem formulation

1. Are the variables of interest given clear conceptual definitions?
2. Do the operations that empirically define each variable of interest correspond to the variable’s conceptual definition?
3. Is the problem stated so that the research designs and evidence needed to address it can be specified clearly?
4. Is the problem placed in a meaningful theoretical, historical, and/or practical context?
Stage 2

Data Collection

Data collection: Gathering studies

• Research question asked
  – What procedures should be used to gather studies?
    • Reference lists of past reviews and meta-analyses should be examined.
    • Multiple data bases should be searched.
    • Manual and electronic searches are needed.
    • “Grey literature” should be included.
    • Search methods must be documented.
    • Sufficient resources and appropriate staff are needed.
Electronic search strategies: Potential problems

While electronic searches are indispensable, they cannot be the only source of studies, because they are subject to a number of problems.

• Problem 1: Publication bias.
  – Publication bias represents the known difference in statistical significance of published vs. unpublished studies.
  – Hence, don’t rely only on electronic searches, which are more likely to locate published studies.

• Problem 2: English language bias.
  – This is the tendency for studies published in English language journals to have stronger findings than studies published in other languages.
  – Hence, don’t limit your search to only English language sources, or rely only on works published in English, unless your review will only be generalized to a limited domain (e.g., schools in the U.S.).
Electronic search strategies: Potential problems

- Problem 3: Inconsistent usage of terms across disciplines.
  - One solution to this problem is to become familiar with the thesaurus of terms for each database. For instance, ERIC, PsycINFO and MEDLINE do not use the same keywords.

- Problem 4: Inconsistent labeling of methodology.
  - This problem is more prevalent in the social sciences where abstracts are less structured and terms are less consistently used (e.g., experiment, randomized trial, RCT).

Data collection: Gathering information from studies

- Research question asked
  - What procedures should be used to extract information from each study report?
    - Coding should be organized, systematic and as objective as possible. Coding forms must be used to ensure consistency.
    - Multiple coders should be used, and the degree of agreement between them should be enhanced via training, then assessed during or after actual coding is complete.
Data collection: Gathering information from studies

- Data collection is critically important because the major source of bias in a review is systematic bias in the included studies.
- If the sample of studies retrieved is biased, the validity of the results of the meta-analytic review is threatened *no matter how systematic and thorough in other respects.*
- Similarly, the data collected (coded) from the gathered studies is the foundation for the inferences to be drawn.

Data collection: Gathering information from studies

- Key goals of coding are
  - to express study results in a standardized form,
  - to find predictive variables that may explain variation in study outcomes, and
  - to anticipate your critics and code what you will need to address possible critiques of your review.
Data collection: Gathering information from studies

• Data collection may also lead the reviewer to return to stage 1, to define the problem of interest more clearly.
  – The literature search may reveal extremely large numbers of studies, leading the reviewer to select more search terms and focus the question more narrowly.
  – Conversely a focused search may produce very few studies, leading the reviewer to broaden search terms, and the question itself.

Data collection: Gathering information from studies

• Begin by keeping a list of possible study features you might be interested in.
  – Eventually narrow the list: meta-analytic sample sizes tend to be small, with insufficient degrees of freedom to support analyses of many predictors.

• Begin preparing a draft coding manual. This will likely be modified as coding begins and uncertainties and discrepancies arise.
Data collection: Gathering information from studies

- Primary determinants of what information gets coded from studies include ...
  - theoretical concerns:
    - What are the important characteristics of the intervention? (or other independent variable)
      - Could include duration, who delivers treatment, etc.
    - What are the important outcomes?
      - Are they psychological, social, physical, etc.?

Data collection: Gathering information from studies

- Primary determinants include ...
  - What are important potential influences on study outcomes?
    - This includes moderators like setting, type of participant, etc.
  - inferential or methodological concerns, like
    - statistical differences (e.g., were results based on a t test versus ANCOVA),
    - research design variations, such as how samples were selected, and
    - differences in measures used.
Data collection: Gathering information from studies

• Consider a web-based coding system. These ...
  – save confusing multiple electronic files that coders periodically update,
  – allow coders to work at home, and
  – automatically save and update data.
• Regardless of the format of the coding system, be sure to have a regular backup routine.

• Details of coding will be covered in a separate training module (later today).

Data collection: Gathering information from studies

• Primary functions served in the synthesis
  – Develop criteria and methods for gathering studies, and gather them.
  – Create a coding frame for obtaining information from studies.
  – Train coders (includes development of manuals, trial coding etc.).
  – Code data from studies.
  – Assess the reliability of extracted information.
Data collection: Gathering information from studies

- Procedural variation that might produce differences in conclusions
  - Variation in procedures used to gather studies, and record information from studies, might lead to systematic differences in
    - what studies are collected, and
    - how those studies are represented in the research-synthesis data set.

Richardson and Rothstein example (1): Searching

The authors ...

- obtained all 45 studies used in the most recent review on the topic,
- conducted electronic searches of six databases: Academic Search Premier, British Library Direct, Dissertations Abstracts, ERIC, ProQuest ABI Inform Global, and PsycArticles,
- performed a “network search” by emailing colleagues knowledgeable in the field, and
Richardson and Rothstein example (2): Searching

- reviewed reference lists of all articles obtained,
- reviewed conference proceedings of APA/NIOSH Work, Stress and Health 2006, and
- searched private and government-sponsored websites devoted to stress research to locate additional unpublished literature:
  - The American Institute for Stress (www.stress.org), Canadian Institute of Stress (www.stresscanada.org), and the National Institute of Occupational Health and Safety (www.cdc.gov/niosh).

Stage 3

Data Evaluation
Data evaluation: Criteria for including and excluding studies

• Research question asked
  What retrieved research should be included or excluded from the synthesis based on
  – the suitability of the methods for studying the synthesis question and/or
  – problems in research implementation?

Data evaluation: Criteria for including and excluding studies

• Primary function served in the synthesis
  Identify and apply criteria to separate “comparable” research studies.

• Procedural variation that might produce differences in conclusions
  Variation across reviews in criteria for study inclusion might lead to systematic differences in which studies are to be synthesized, and thus in conclusions that will be drawn.
Data evaluation: Criteria for including and excluding studies

• During protocol development the reviewer sets up inclusion rules. It may not be possible to evaluate all criteria during the data-collection stage if only titles and abstracts are available for scrutiny.

• The reviewer (or reviewers) examines each study, asking whether it meets the inclusion standards.

The next slides provide some examples of issues in the form of questions (but each review will have its own set of inclusion rules).

• What are acceptable operationalizations of your constructs? What measures are acceptable? Common issues include...
  • Are self-reports acceptable?
  • What evidence of instrument validity is required?
Data evaluation: Criteria for including and excluding studies

– What types of treatment are acceptable? Common issues include...
  • Is there a minimum “dose” for the treatment?
  • What characteristics operationally define the intervention? Are any of these absolutely essential? Are others less central to the construct definition?

– What kinds of participants are eligible?
  • Is there a restricted age range?
  • Operational definitions of participant types are needed (e.g., “at risk” students).

Data evaluation: Criteria for including and excluding studies

• Are sufficient data available to allow for computation of effect sizes?
  • Difficult to include in a meta-analysis if not
    • Can still include in systematic review
  – Effect sizes are dealt with in a separate training module (later this afternoon).

• Will the review restrict study designs?
  – Some reviews include only randomized controlled studies, others examine all sorts of studies. The key is to be explicit about this decision, which also relates to the issue of study quality.
Data evaluation: Criteria for including and excluding studies

- Will published and unpublished studies be included?
  - Publication status should not be a criterion for a good systematic review.
  - Some past reviews have included only studies published in peer-reviewed outlets, perhaps as a proxy for quality, but this can lead to publication bias (this will be covered in more detail in another module).

Data evaluation: Criteria for including and excluding studies

- How will study quality be handled?
  - Some researchers argue that “poor” studies should be excluded, claiming that a review based on a small number of high quality studies is more likely to be accurate than a larger review including lower quality studies.
  - Others argue that reviews should include a broad range of studies and then examine whether aspects of “quality” relate to study outcomes.
Data evaluation: Criteria for including and excluding studies

• How will study quality be handled?
  – The features of study quality that are thought to influence study results (i.e., effect size) can be coded and examined for moderating effects.
  – This avoids an unnecessary subjective decision to exclude studies based on an assessment of their quality, and also allows the reviewer to empirically examine whether various study features make a difference.

Data evaluation: Criteria for including and excluding studies

• How will study quality be handled?
  – In some fields it may be possible to identify true “fatal flaws” that call for a study to be excluded; the reviewer must be explicit about reporting such exclusion rules.
  – Because study quality is an important and contentious topic, the treatment of quality will be covered in more detail in the training module on coding study features.
Data evaluation: Criteria for including and excluding studies

• What moderator variables must be reported?
  – Are there some key sample or treatment features that are needed for a clear understanding of the problem?
  – For instance, must results be reported separately by gender or for certain age groups? Must age of the participant sample be reported?

• Tests for most moderating effects will be based on differences between studies. However often study characteristics are confounded, which can mean that several “competing” explanations may be found.
  – If some studies provide within-study comparisons (e.g., of gender differences) we can examine both within-study evidence and between-studies (review-level) evidence on the same question.
Data evaluation: Criteria for including and excluding studies

- What moderator variables must be reported?
  - Also, because meta-analyses are not structured or planned experiments, the information based on differences among studies is correlational.

Data evaluation: Criteria for including and excluding studies

Key questions to be asked at this stage include ...

- If studies were excluded from the synthesis because of design and implementation considerations, were these considerations
  - explicitly and operationally defined, and
  - consistently applied to all studies?
- Were studies categorized so that important distinctions could be made among them regarding their research design and implementation?
Stage 4

Data Analysis and Interpretation

Data analysis: Integrating the results of studies

• Research questions asked about analysis
  – What procedures should be used to analyze and combine the research results?
  – This issue will be dealt with in detail in other training modules (later this afternoon).
• Primary function served in the synthesis
  – Identify and apply procedures for ...
    • combining results across studies and
    • testing for differences in results between studies.
Data analysis: Integrating the results of studies

• Procedural variation that might produce differences in conclusions
  – Variation in procedures used to analyze results of individual studies (e.g., how effect sizes are averaged, the types of models used for analysis) can lead to differences in cumulative results.

Data analysis: Checklist for analysis (1)

• Was an appropriate method used to combine and compare results across studies?
• If a meta-analysis was performed, was an appropriate effect-size metric used?
• If a meta-analysis was performed ...
  – were average effect sizes and confidence intervals reported, and
  – was an appropriate model used to represent the effects, and the error in effect sizes?
Data analysis: Checklist for analysis (2)

- If a meta-analysis was performed, was homogeneity of effect sizes tested?
- Were ...
  - study design and implementation features along with
  - other critical features of studies, including historical, theoretical, and practical variables,
...tested as potential moderators of study outcomes?

Data analysis: Interpretation of results

- Research question asked about interpretation
  What conclusions can be drawn about the cumulative state of the research evidence?

- Primary function served in the synthesis
  Summarize the cumulative research evidence with regard to its strength, generality, and limitations.
Data analysis: Interpretation of results

• Procedural variation that might produce differences in conclusions
  – Variation in
    • criteria for labeling results as “important” and
    • attention to details of studies
  – might lead to differences in interpretation of findings.

Data analysis: Interpretation checklist (1)

• Did any analyses test whether results were sensitive to statistical assumptions and, if so, were these analyses used to help interpret the evidence?
• Did the reviewer
  – discuss the extent of missing data in the evidence base and
  – examine its potential impact on the synthesis findings?
• Did the reviewer discuss the generality and limitations of the synthesis findings?
Data analysis: Interpretation checklist (2)

- Did the reviewer make the appropriate distinction between study-generated and review-generated evidence when interpreting the synthesis results?
- If a meta-analysis was performed, did the reviewer
  - compare the magnitude of effects from the review with other related effect sizes, and/or
  - present a practical interpretation of the significance of the effects?

Stage 5

Reporting the Synthesis
Report preparation

• Research question asked
  What information should be included in the report of the synthesis?

• Primary function served in the synthesis
  Identify and apply editorial guidelines and judgment to determine the aspects of methods and results that readers of the synthesis report need to know.

Report preparation

• Procedural variation that might produce differences in conclusions
  – Variation in reporting might
    • lead readers to place more or less trust in synthesis outcomes and
    • influence others' ability to replicate results.
Report preparation: Checklist

• Were the procedures and results of the research synthesis clearly and completely documented?

• Key topics for replicability are ...
  – search procedures and inclusion/exclusion rules,
  – coding and computation of effect sizes, and
  – analysis procedures.

• What is the strength of the cumulative evidence?
  – How good is the quality of evidence?
  – How large are the effects?
  – Are results consistent? If not, are differences explainable?

• Are biases and limitations of the review discussed?
  – Was the search comprehensive? Was publication bias discussed?
  – Was quality assessed?
  – Were analyses appropriate?
Overarching goals in systematic reviewing

• **Limit bias** in literature reviews by using …
  – a systematic approach to data collection, analysis, and evaluation
  – both published and unpublished research (minimizing publication bias)

• **Enhance methodological rigor** by…
  – holding reviewers to high scientific standards

• **Employ better and more appropriate statistical tools** by…
  – avoiding decision rules with poor inferential characteristics
  – using statistics that are appropriate for the analysis of collections of studies

---

**Introduction to Systematic Reviewing and Meta-Analysis**

Jeff Valentine
University of Louisville, USA

jeff.valentine@louisville.edu