Systematic Reviews: An Overview

Hannah R. Rothstein
Baruch College, New York, NY

Michael Borenstein
Biostat, Englewood, NJ
Outline

Systematic reviews
  – Rationale
  – Uses
  – A brief history
  – Stages of a systematic review
  – Criticisms and limitations
The rationale for systematic reviews
What is a systematic review?

- A summary of the research literature that uses
  - explicit methods to identify relevant studies
  - objective techniques to combine these studies
A systematic review aims to be

- Explicit (e.g. in its statement of objectives, materials and methods)
- Systematic (e.g. in its identification of literature)
- Transparent (e.g. in its criteria and decisions)
- Reproducible (e.g. in its methodology and conclusions)
- Unbiased
Meta-analysis is

- A systematic review that uses quantitative methods to summarize the results

OR

- The quantitative synthesis component of a systematic review
Why is systematic reviewing necessary?

- There may be so much literature on a topic that it is impossible to make sense of it.
- Educational research often produces results that appear to be contradictory, even when they are actually consistent.
- Educational research often produces results that are actually contradictory due to moderator variables (populations, treatments, settings, methods, designs), and we need to be able to understand how these influence the treatment effects.
Too much literature

Among the earliest meta-analyses were synthesis of:

- 345 studies of the effects of interpersonal expectations on behavior (Rosenthal & Rubin, 1978),
- 725 estimates of the relation between class size and academic achievement (Glass & Smith, 1979),
- 833 tests of the effectiveness of psychotherapy (Smith & Glass, 1977),
- 866 comparisons of the differential validity of employment tests for Black and White workers (Hunter, Schmidt & Hunter, 1979)
Apparent inconsistency

• See previous slide
Operation of Moderators

- Cooper et al. (1996) Summer vacation.
  - On average, students’ achievement scores were 1 month lower when they returned in the fall than when they left in the spring
  - However:
    - There were dramatic differences based on skill area (biggest losses were in math and spelling)
    - Reading loss was greater for disadvantaged children than for middle class children
    - Intelligence had no impact on summer learning loss
What is wrong with narrative reviews?

• Different experts may perform a review on the same question and come to different conclusions
  – Sometimes due to review of different sets of studies
  – Even when same studies are reviewed, the process of integrating them is subjective
What is wrong with narrative reviews?

• Narrative reviews become less efficient with more data
  – While a researcher can combine the results of a few studies in his or her head, it becomes increasingly difficult to do so as the number of studies increase.
  – In education and the social sciences, it is not unusual to have a review that includes 10 or more studies, sometimes many more than that.
What is wrong with narrative reviews?

- Narrative reviews have a hard time tackling variations in treatment effectiveness
  - As the number of studies grow, they often examine different populations
  - The size of the relationship of interest may vary in different populations (i.e., there are moderator effects)
  - The narrative reviewer, who has enough trouble summarizing studies when they are all done in similar situations, now has a much harder task
Why are systematic reviews better?

In addition to being more objective and more replicable

- They are *more* efficient with more information
- They can deal with variations in treatment effectiveness
Is class size related to student achievement?

• By 1978 there were hundreds of studies on this topic
Narrative review

• Thompson concluded that the relationship involved too many complex issues to be reduced to a single testable hypothesis, and that research findings were “necessarily inconclusive”.
Meta-analysis

• Smith and Glass (1978)
  – 80 studies
  – 700 comparisons of smaller and larger classes

• Results showed clearly that smaller classes are better for
  – student achievement,
  – classroom processes
  – teacher and student attitudes.
How can a systematic review be used?
Uses of Systematic Reviews

• Identifying areas to study
• Designing new studies
• Grant applications
• Informing
  – Practitioners
  – Policy Makers
  – Consumers
Identifying areas to study

- What works (Homework)
- What doesn’t work (Drug Abuse Resistance Education)
- What needs further study
  - Promising programs (Prison-based vocational education for adults)
  - Programs about which very little is known (Programs that prepare non-college bound students for the work force in the US)
Identifying areas to study *further*

**Homework**

- Overall doing homework improves achievement
- Relative to other education interventions, homework has an above average size effect
Identifying areas to study further

• But despite decades of homework research
  – No tests of grade level, subject matter, or other influences could be conducted using results from *manipulated* homework studies (K=6).
  – Most factors that might influence the link between homework and student outcomes *have never been the focus of research*. 
An exemplary meta-analysis

Designing new studies

- Which treatment variant/population/treatment variant is most likely to yield an effect for what outcome
- Which factors need to be measured
- Estimation of the likely effect under clinically representative conditions
Designing new studies

- Whitehead and Becker (2000) used meta-analysis to test a model of post-divorce adjustment of young children.
- They found that the guiding model had not been fully studied in the literature.
  - E.g. Only five studies reported data by gender, so it was not possible to test whether the same model applied to boys and girls.
  - Their meta-analysis identified the “missing links” that need to be studied.
Estimation of the likely effect of a treatment under real life conditions

• Shadish et al. (2000)
  – Response to criticism that most randomized experiments of family therapy were conducted in settings that were unlikely to generalize to clinical practice
Grant Applications

• Justify need for new study
• Put the available data in context, show what new study can contribute
Informing policy makers, practitioners and consumers

• Policy Makers
  – Phil Davies, Government Social Research Unit, Cabinet Office, UK and the Magenta Book
  – Larry Sherman et al. and PREVENTING CRIME, A Report to the US
  – The US Department of Education and the What Works Clearinghouse
Informing policy makers, practitioners and consumers

- Practitioners and Consumers
  - Production of Evidence-based Practice Guidelines
  - U.S. Department of Education
  - Institute of Education Sciences
  - National Center for Education Evaluation and Regional Assistance
Informing policy makers, practitioners and consumers

• Practitioners and Consumers
  – Phil Abrami, Bob Bernard and Anne Wade at the Canadian Network for Knowledge Utilization at Concordia University

• “Focus on the conduct of high quality and timely systematic reviews... especially ones focusing on the effectiveness of interventions or “what works “

• Attempts to bridge gap between education research and classroom teaching
A brief history of systematic reviewing and meta-analysis
Statistical procedures for the integration of research have existed for 100 years

1904: K. Pearson. The effects of a vaccine against typhoid. (11 studies)

Modern synthesis methods develop starting in the mid-1970s

• Gene Glass
  – Coined the phrase “Meta-analysis”.
  – And defined Meta-analysis as “the statistical analysis of a large collection of analysis results from individual studies for purposes of integrating the findings”.
Meta-analysis is used nearly simultaneously in 3 different areas


...and flaws in traditional reviewing procedures are highlighted

The first textbooks on statistical procedures appeared in the 1980s…:

The first textbooks on statistical procedures appeared in the 1980s…:


A scientific paradigm for systematic reviewing emerged

Stages of a Systematic Review
parallel stages in primary research

- Problem Formulation
- Data Collection
- Data Evaluation
- Data Analysis
- Presentation of Results
The first texts on scientific research synthesis appeared shortly thereafter


Today

• The science of systematic reviews, and the use of systematic reviewing procedures continue to grow and develop
Approaches to Systematic Reviewing

Historical and contemporary
Historical Approaches

• Vote-counting
• Combining p values
Vote-counting

Often leads to incorrect conclusions

• Focuses on the statistical significance of the primary studies

• Often wrong, due to low power of primary studies

• Even when correct, doesn’t provide information about the size of effects or the consistency of effects across studies
Combining p values

Same problems as in primary studies
- effect may be large but not statistically significant
- effect may be trivial, yet statistically significant
Contemporary Meta-Analysis: Hedges-Olkin and similar approaches

- Focus on effect size
- Weight by precision
- Combined effect
- Heterogeneity
  - To qualify combined effect
  - To explain impact of moderator variables
- Assess impact of possible biases
Vote Counting vs. Hedges-Olkin

• Does School Spending Make a Difference?
• Eric Hanushek used the “vote counting method” to summarize the results of studies reported in 38 articles and books published from the 1960s-1980s and concluded that there was no systematic, positive relationship between school financial resources and student performance
Does School Spending Make a Difference?

• Larry Hedges and two of his students used meta-analysis to summarize the results of the same studies Hanushek examined.

• Hedges concluded that the relationship between resource inputs and student outcomes was consistent and positive. He suggested that these findings should be used to inform educational policy.
Where did this lead?

• Both the Hanushek-Hedges set of studies and the Glass and Smith studies led to new, more focused research—
Stages involved in undertaking a systematic review
Five Stages of A Systematic Review

1. Problem formulation
2. Data collection stage
3. Data evaluation stage
4. Analysis and interpretation stage
5. Public presentation stage
Five Stages of A Systematic Review

• Systematic Reviews/Research Synthesis are often viewed and taught as a set of statistical procedures.
• They are best viewed as as a multi-stage process with the statistical analysis as only one of the stages.
• Some systematic reviews have no statistics at all.
Five Stages

• 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.
What a systematic review should contain:

- A clearly defined, explicit question
- Comprehensive and systematic search for studies
- Explicit, reproducible strategy for screening and including studies (inclusion/exclusion criteria)
- Explanation of whether, why and how quality of primary studies was assessed
- Explicit, reproducible data extraction
- Appropriate analysis and reporting of results
- Interpretation supported by data
- Implications for future research, and if relevant, for policy and practice
Stage 1: Problem formulation

- Whether broad or narrow, the research question should be **focused**.
- How are these two questions different?
  - Does homework improve student achievement?
  - In high school students who are below average performers, does daily homework lead to better performance on unit tests, compared to no homework?
Stage 1: Problem Formulation

How are these questions different?
• Is sex education useful in preventing teenage pregnancy?
• For female high school students, does exposure to a sex education course at school reduce the likelihood of getting pregnant before age 18?
Issues to think about when formulating a question

• Theoretical Orientation
• Participants
• Intervention or Independent Variable
• Outcome or Dependent Variable(s)
• Acceptable research designs
• Comparisons
Write a protocol

- Forces you to read and understand the background
- Makes you formulate a focused question
- Makes you plan your information retrieval strategy
- Makes you think through and describe inclusion/exclusion criteria clearly
- Makes you think about the data you want to collect and the methods you will use to analyze them
Protocols

• Guidelines are available at:
  – http://www.campbellcollaboration.org/guidelines.asp

• Examples are available at:
  – http://www.campbellcollaboration.org/ECG/proto.asp
Stage 2: Data Collection

- Also known as the Information Retrieval stage.
- Critically important because the major source of bias in a systematic review is systematic bias in the included studies.
Stage 2: Data Collection

If the sample of studies retrieved is biased, the validity of the results of the meta-analytic review is threatened \textit{no matter how systematic and thorough in other respects}
Stage 2: Data Collection

So, what this stage is really about is conducting:

- An unbiased search for studies
- A thorough search for studies

How does one do that?
Stage 2: Data Collection

- Appropriate staff
- Multiple data bases
- Manual as well as electronic searches
- Inclusion of “grey literature”
- Documentation of what has been searched, and how
- Sufficient resources
Example

• For a review to determine the effectiveness of parental involvement in improving the academic performance of school age children in grades K-6.
1. From the research question, determine the main concepts to be searched (usually there are three):
   - Parental involvement
   - Academic performance
   - Kindergarten or Elementary students
2. Look up each concept in the thesaurus for this database.
   - **Descriptors**: Parental Involvement See: **Parent participation**
   - **Related descriptors**: Family involvement, Parent-school relationship, Parent role, Parents as teachers
   - **Related keywords**: parent* involvement, parent* effectiveness, parent* support, family support
   - **Descriptors**: Academic Performance See: **Academic achievement**
   - **Related descriptors**: Science achievement, Reading achievement, Writing achievement, Achievement gains
   - **Descriptors**: Elementary School Children See: **Elementary school students**
   - **Related descriptors**: Elementary education, Primary education, Kindergarten
Stage 2: Data Collection

• Importance of information retrieval process
  – Consulting with an Information Specialist is highly recommended
  – “Shoestring-budget information retrieval is likely to introduce bias, and should be avoided.” (Rothstein et al. 2004)

• Use of bibliographic management software
  – Store, manage and organize results

• Ability to replicate review
  – Documentation of entire process
Stage 3: Data Evaluation

• Now that you have collected potentially relevant studies, what do you do with them?
Inclusion and Exclusion Criteria

• Respondents/participants
• Key constructs/variables/operationalizations
• Study Types/research designs
• Cultural/national/linguistic restrictions
• Time frame
• Publication types
Exclusion on the basis of quality?

- Being too restrictive may restrict ability to generalize
- Being too inclusive may weaken the confidence that can be placed in the findings
- Methodological quality assessment is often subjective (e.g., attrition vs. randomization)
- Scales of methodological quality don’t correlate with each other
- Overall quality vs. specific features?
An Example

- The Effects of School-based Social Information Processing Interventions on Aggressive Behavior: A Campbell Collaboration Systematic Review
- *Education Review Group*
- Sandra Jo Wilson and Mark W. Lipsey
Example of Inclusion Criteria

Interventions

- The interventions must meet the three criteria that define social information processing programs. Although other treatment components may be included (e.g., behavioral social skills training, parenting skills training), the social information processing component must be the clear focus of the program.

- The definitional criteria are as follows:
  - 1. Programs involve training in one or more of the social information processing steps: (1) encoding situational and internal cues, (2) interpretation of cues, (3) selecting or clarifying a goal, (4) generating or accessing possible responses, (5) choosing a response, (6) and behavioral enactment.
  - 2. Programs emphasize cognitive skills or thinking processes rather than specific behavioral skills. By teaching generic thinking skills, the programs aim to improve information processing in myriad social situations.
  - 3. Programs involve the use of structured tasks and activities through which the cognitive skills are learned and applied to actual social situations.
Example of Inclusion Criteria

Settings and Subjects.

- The interventions must be delivered to school-aged children (K-12 or equivalent ages in international settings) in regular school settings during school hours. Special education classrooms and alternative schools are eligible school settings, although classrooms in residential facilities (e.g., psychiatric hospitals) are not. After-school programs are not eligible.

- Any qualifying school in any region or country is eligible.
Example of Inclusion Criteria

Outcomes.

- The study must report intervention effects for at least one outcome variable, measured on children, representing aggressive behavior, broadly defined to include violence, aggression, fighting, person crimes, disruptive behavior problems, acting out, conduct disorder, externalizing problems, and so forth.
Example of Inclusion Criteria

Study Design.

- Only studies using a control group design will be eligible. The intervention and control groups can be randomly or non-randomly assigned but, if nonrandom, must be matched or provide evidence of initial equivalence on key demographic variables and/or pretests.

- Control groups can represent placebo, wait-list, no treatment, or “treatment as usual” conditions.

- Studies without control or comparison groups are not eligible. This includes one group pretest/posttest studies and studies in which a treatment group is compared to another treatment group.
Stage 3: Data Evaluation
Dealing with Methodological Quality Issues

Assessment of Methodological Quality

- A range of characteristics related to methodological quality will be coded for each of the eligible studies.
- These characteristics include unit and method of assignment to experimental groups, analysis of pretreatment differences/similarities between groups, attrition from each group, and characteristics of the dependent measures including reporter, social desirability bias, treatment test overlap, and the like.
- These characteristics will be carried over into analysis of program effects.
- Methodological variables that have important relationships with effect size will be statistically controlled or (if that is not possible) used to identify studies to be excluded from the synthesis.
Results of the Collection/Evaluation Process

- Potentially relevant studies from literature identified through electronic and hand searches (n=300)
- Excluded on basis of abstract, e.g., not randomized or quasi-experiment with control group (n=250)
- Articles selected for full text review (n=50)
- Excluded after full text review (n=30)
- Eligible studies (n=20)
Stage 3: Data Evaluation

Coding

• Your coding form will need to contain
  – Identifying information
  – Effect size information
  – Characteristics of studies that you think may be related to study results

  – **Warning:** What you don’t code, you can’t examine later on!!
Stage 3: Data Evaluation

Levels of Coded Data

- Study Data
- Treatment/Comparison
- Sample
- Outcome
- Effect size
Stages 4 and 5: Meta-analysis and Presentation of Results

To be Presented after the Break

by

Michael Borenstein
Limitations of Systematic Reviews

- How do you capture qualitative distinctions between studies?
- Primary studies are reported poorly
- “Apples and oranges”
- Nearly all reviews include “flawed” studies to some extent (e.g., a randomized design with attrition)
- Publication bias is always an issue
- Analysis of between-study differences is correlational or “interocular”
Selected References


