

Selecting the “Right” Program: Using Systematic Reviews to Identify Effective Programs

Sarah Avellar, Mathematica Policy Research,
Janet Horras, Iowa State Department of Health,
and
Lauren Supplee, Administration for Children and Families

Agenda

- Context of Evidence-Based Decision Making: Why does this matter?
- Standards for evaluating research: What is strong evidence?
- Aggregating the evidence
- Applying the results
- Using Needs Assessment to Match the Model to the Community
- Evidence-based Decision Making: The whole package
- Resources

1

Context of Evidence-Based Decision Making: Why does this matter?

“Wherever possible, we should design new initiatives to build rigorous data about what works and then act on evidence that emerges — expanding the approaches that work best, fine-tuning the ones that get mixed results, and shutting down those that are failing.”

- Peter Orzag, past director of the Office of Management and Budget, June 8, 2009

2

Context of Evidence-Based Decision Making: Why does this matter?

"Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future."

- Jeffrey Zients, acting director of the Office of Management and Budget, May 18, 2012

3

OMB Memorandum M-10-32 & M-12-14

- In July 2010 OMB announced a voluntary program to provide additional funds to those agencies that provide documentation that their FY12 budget request are evidence-based or subject to rigorous evaluation.
- In May 2012 OMB sent a memo directing all federal agencies to demonstrate the use of evidence throughout their FY14 budget proposals. This memo also emphasized the use of cost-effectiveness data in allocating resources and infusing evidence into grant-making.

4

What is Evidence-Based Policy or Decision Making?

- **Evidence-based decision making** is the explicit, conscientious and judicious consideration of the best available evidence in the provision of services.
- **Evidence-based policy** is public policy informed by rigorously established objective evidence. It is an extension of the idea of evidence-based medicine to all areas of public policy. (Wikipedia, January 31, 2011)

5

Where did this Concept Come From?

“Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.

The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.” (Sackett, Rosenberg, Gray, Haynes & Richardson, 1997, p 71)

6

Evidence-based Decision-making (Nutley et al 2008)

- Helps well-informed decisions by putting the best available evidence at the heart of policy development and implementation
- Includes questions not just of “what works” but also “why does it work? How do you get stronger impacts?”
- Using research in conjunction with monitoring data, expert knowledge and stakeholders

7

What does Evidence-based Decision Making Require?

- Ability to access research (original articles vs. systematic reviews)
- Understand the results (i.e. evidence)
- Use knowledge to apply evidence to a context

“Good research should be seen more as a stimulus to respectful dialogue and less often as a trump card to truncate any debate” (Nutley et al, 2008, p 300)

8

What is Evidence vs Knowledge?

- All of these definitions rely on the term “evidence” – What is evidence?
- Research is a process and the output is research findings (Davis & Nutley, 2008).
 - Research findings used to inform decision-making become “evidence”
 - Knowledge includes research findings but also other ways of knowing (Davis & Nutley, 2008).

9

Sources of Knowledge

- Experience and judgment (tacit knowledge)
- Expert opinion
- Anecdote
- Values
- Political pressures and public opinion
- Tradition and professional norms
- Research and evaluation

10

Sources Of Evidence and Knowledge

Sources of Evidence or Knowledge	Examples
From research	Most relevant, high-quality qualitative or quantitative evidence available, can be specifically impact of a program or a body of research on a particular topic
From frequency, causes, modifying factors in local community	Surveillance data and community health status reports, significance of issue in comparison to other community health concerns
From people about community and political preferences	Needs and interests of community members, support of public officials, current organizational climate
From governments and programs about resources	Available financial resources, human resources, materials

From National Collaborating Centre for Methods and Tools

11

Domains That Influence Evidence-based Decision-making



Graphic from Spring et al 2007 & 2008

12

A Challenge of Evidence-based Policy & Practice (EBP)

- In order to apply “what works” to practice, some barriers have been identified including (Meline & Paradiso, 2003; Nelson, Leffler, & Hansen, 2009):
 - Gathering all of the available research
 - Learning all of the necessary research methods to evaluate the evidence
 - Deciding if the evidence is relevant to the client population
 - Summarizing the information to be used.

13

Use of Evidence

- William T. Grant foundation funded a study of educators, policy-makers and others’ use research evidence (Nelson, Leffler, & Hansen, 2009)
- All participants said research evidence was one of many factors taken into consideration in making decisions.
 - Participants expressed cynicism about research evidence and it’s applicability to their context.
 - There was a feeling there is a gulf between research and practice and research rarely reflected the real world context.

14

Use of Evidence

- Participants in this same study stated that evidence was a key part of decision-making but that they also used their own data (Nelson, Leffler, & Hansen, 2009)
- The value to administrators of building data systems and utilizing continuous quality improvement in a program is that it broadens the context of evidence.
 - Program administrators can merge research findings with local data to build knowledge and inform decision-making

15

Building Evidence

Think back to when you were a student and you were asked to write an essay defending a point. Did you use all of the information you found? Did you use only the points that defended your argument? (Greenhalgh , 1997)

16

Sources of Bias in Synthesizing Research (Littell, 2008)

- Three categories:
 - Bias arising in the original studies
 - Dr. Avellar will speak about this source of bias
 - Bias in dissemination of the study results
 - “File drawer problem,” peer reviewers
 - Bias in the review process
 - E.g. availability, familiarity, confirmatory hypotheses

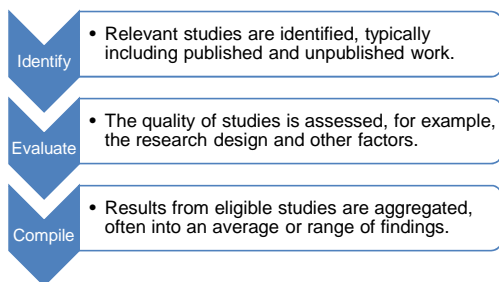
17

What is a Systematic Review?

- Literature review
- Uses systematic and transparent methods to identify and evaluate available literature
- Critiques the available literature
- Focuses on a specific question
- Synthesizes the results
 - Littell (2008) states, “syntheses are important because single studies, no matter how rigorous, have limited utility and generalizability”

18

The Review Identifies and Synthesizes Research



19

Why Conduct Evidence Reviews?

- May identify effective programs
- Efficient use of resources to use existing data, rather than conduct new studies
- Independent studies may provide information on the program in different contexts
- Assess the state of evidence for a particular intervention
- Identify gaps in research evidence

20

Critiquing a Systematic Review

- Did the systematic review include transparent intentions and methods?
- Did the systematic review captured all of the research?
 - Limitations of only using peer-review
- Was the selection of which articles to review appropriate?
- Was the assessment of study quality appropriate?
- Analyzing & reporting the results –
 - Were the results similar across studies for the same program?
 - Were the results similar within a domain across programs?
- Will the systematic review be updated to remain current and relevant?

21

ACTIVITY: Cookies!

Please come up to the front and get one cookie of each brand and a napkin.

Take one bite of the cookie.

Rate the cookie on a scale of 1-10.

22

Standards for Evaluating Research: What Makes for Strong Evidence?

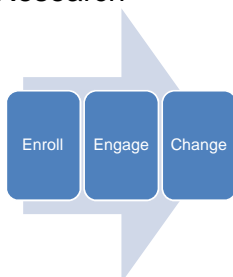
February 10, 2013

Association of Maternal and Child Health
Programs Skills-Building Training

Sarah Avellar

What We Want to Know From the Research

- Did the program cause the outcomes?
 - This is know as *internal validity*
- Otherwise, we do not know if the program or some other factor led to the outcomes



24

Other Factors That May Affect Outcomes

- Natural change over time
 - Through maturation and development, a child develops new skills
- Characteristics of respondents
 - Those who sign up for the program are committed to improving their children's health and make changes of their own
- Other changes that occur simultaneously to the program
 - A new health clinic, which offers a wide range of free or low-cost preventative services, opens in the area

25

Focus on Internal Validity

- Most systematic reviews focus on internal validity
- Standards are designed to assess how well the study can estimate unbiased impacts of the program
 - Bias is a systematic error in the estimated effect of the program

26

Important Factors in Assessing Internal Validity

- Research design
- Method for selecting groups
- Initial equivalence of groups
- Attrition or loss of sample members
- Confounding factors

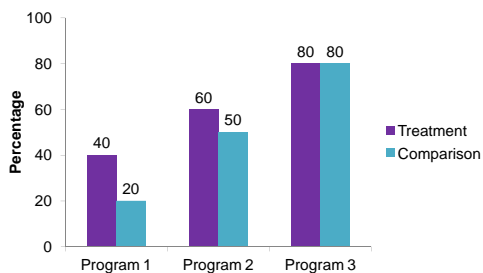
27

Research Design

- A pre-post design does not have strong internal validity because it cannot rule out other factors that may affect outcomes
- A comparison group is needed to represent what would have happened to participants without the treatment

28

Without Comparisons, Results May Be Misleading



Without a comparison, Program 3 might appear to be the most effective.

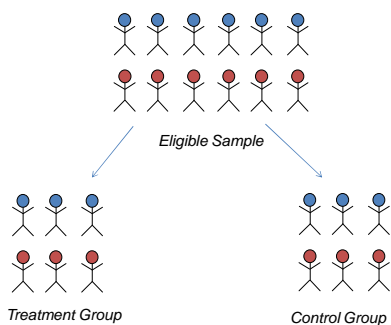
29

Methods for Selecting Groups

- To form treatment (or program) and comparison groups, different methods may be used:
 - By chance or randomly, known as *randomized controlled trials* (RCTs)
 - By purposive selection, known as *quasi-experimental designs* (QEDs)

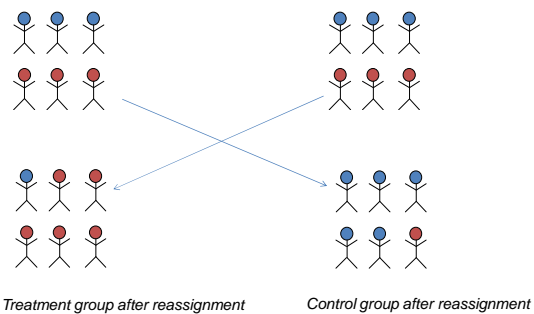
30

Randomization Produces Similar Groups



31

Reassignment Can Create Dissimilar Groups



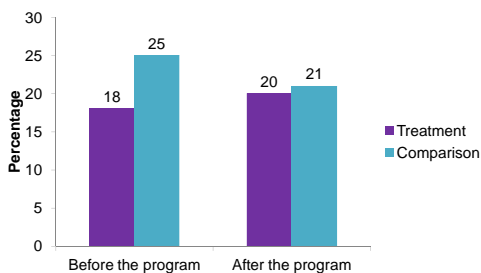
32

Purposive Selection May Create Differences

- An administrator assigns the neediest families to the treatment group
- The more motivated families choose to participate in the treatment group
- Staff decide that those who live within a mile of program facilities will be assigned to treatment, but the sociodemographics of the area change dramatically farther out

33

Without Baseline Equivalence, Results Are Unclear



Percentages went up in the treatment group and down in the control group, but interpretation is unclear because the groups were different at baseline.

34

What about Unobserved Characteristics?

- Unobserved characteristics were not measured in the study and may be difficult to measure
- With random assignment, groups will be similar, on average, for observed and unobserved characteristics
- With purposive selection, we can never be certain that the groups are similar on all characteristics

35

Attrition Can Affect Sample Composition

Even if groups are initially equivalent, the loss of respondents may create dissimilar groups.

36

Program Effects Cannot Be Isolated from Confounding Factors

There is a confound between the program and home visitor.

Program is implemented by multiple home visitors.

37

Other Factors Can be Harder to Quantify

- Measurement
 - Are the measures appropriate for the target population?
 - Is the length of follow-up acceptable?
- Analysis
 - Were the data properly analyzed?
- Do the results apply to those outside of the study?
 - This is known as external validity
 - There are few standards to assess this

38

Translating Research: Two Key Questions

- Are the results something I can believe?
 - This is an issue of internal validity and the focus of most systematic reviews

- Will it work for me?
 - This is an issue of whether the study results will generalize, known as *external validity*
 - We will discuss this issue in more detail in the later session

39

Summary

- In most evaluations, we are interested in whether the program caused the outcomes
 - A comparison group is critical
- A study's ability to isolate the impacts of the program is affected by numerous factors including:
 - Random versus purposive selection
 - Baseline equivalence
 - Attrition
 - Confounding factors

40

ACTIVITY: Cookies!

- Now use the standards and definitions we've generated to rate your cookies again.
- Take another bite and rate each cookie on each of the standards using 1-10 scale.

41

Applying the Results of a Systematic Review

February 10, 2013

Association of Maternal and Child Health Programs Skills-Building Training

Sarah Avellar

What Does a Systematic Review Tell You?

- Reviews often highlight interventions or services with some research evidence of effectiveness
 - This does not mean that the program will work in all situations
- In addition to information on effectiveness, a review may provide useful contextual information for selecting a program, such as:
 - Implementation requirements
 - Populations served

43

The Role of Context

- Broad questions to consider include:
 - How similar is it to my community/target population?
 - How much should I weigh the findings from this program over another one?
 - Does this community have the infrastructure necessary to implement the program?
- Ideally there is a dialogue between decision-makers and service providers or other stakeholders when selecting a program

44

"We cannot expect the result of RCTs and systematic reviews to be relevant to all patients and all settings (this is not what is meant by external validity) but they should at least be designed and reported in a way that allows clinicians to judge to whom they can reasonably be applied" (Rothwell, 2005, p. 83)

45

Deciding Whether a Program Model Is Effective

- Studies may show a mix of favorable and unfavorable results, as well as outcomes on which the program has no discernible effect
- Consumers of research evidence should consider many factors when determining whether a program is "right" for a given population and setting

46

Issues That May Affect External Validity of the Results (Rothwell, 2005)

- Setting of the study
- Selection of the participants
- Characteristics of the participants
- Relevance of the control condition
- Outcome measures
- Adequacy of the length of the follow up

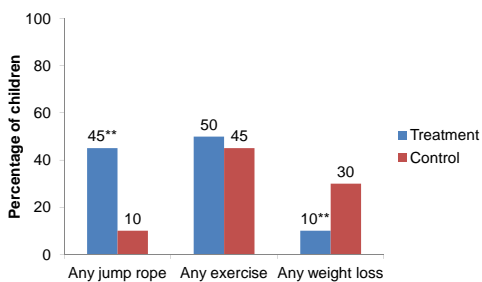
47

A Hypothetical Example

- The *Jump Start to Health* program is designed to decrease obesity in children living in communities with limited access to fresh food or safe public space for physical activities. The program organizes group jump-rope events.
- The systematic review included the results from three random-assignment evaluations.

48

Is *Jump Start to Health* Effective?



49

All Measures Are Not Created Equal

- It may be easier to effect change for some outcomes than for others
 - Attitudes versus behaviors
 - Normed versus author developed
 - Self-report versus direct observation
- Some outcomes are more distal from the program
 - Outcomes closely aligned with program goals may be easier to change than more global outcomes

50

Breadth of Impacts May Vary

- Some programs may affect a range of outcomes, while others are more limited
 - In the home visiting review, we considered diverse outcomes, including maternal health, child development, and family self-sufficiency
- A more limited variety of impacts is not necessarily problematic, if the practitioner is interested in a targeted program

51

Were the Findings Consistent?

- Multiple outcome measures often result in mixed findings: favorable, unfavorable, and no effect
- Findings may be favorable by chance
- More consistency provides more confidence in the results
 - Is there any pattern to the findings (based on domains, length of follow-up, etc.)?

52

Considering Unfavorable Results

- Some results may be unfavorable or ambiguous
 - For example, an increase in the number of days mothers are hospitalized could indicate an increase in health problems or increased access to needed health care due to participation in a home visiting program
- Are there any outcomes for which an unfavorable or ambiguous result would mean the program should not be used?

53

Shorter Follow Ups May Have Stronger Impacts

- Outcomes measured during or immediately after a program show the short-term effects of the program
- Long-term change often is a higher bar than short-term change.

54

No Easy Interpretation for the Size of Effects

- Effect size is a standardized measure of the difference between groups
- However, the meaning of the effect size is not as straightforward
 - A small effect for an outcome such as child abuse may have more clinical significance than a larger effect for maternal health
- Interpret effect sizes in context (Hill et al. 2008)
 - What is the normative expectation?
 - What has other research shown?

55

What Was the Counterfactual?

- Comparison condition can vary
 - No treatment
 - Less intensive treatment
 - Alternative treatment
- The size of the effect will likely be larger the less the comparison group received

56

What Were the Characteristics of the Samples?

- Differences between the research samples and the population of interest may translate to differences in effects
 - Age, race and ethnicity, level of disadvantage, number of children, marital status
- Who administered the program?
 - Were the providers trained by the developer?
 - What were the providers' levels of experience and/or education?

57

Note Conflicts of Interest

- A conflict of interest does not necessarily undermine the research, but it should be considered
- Example: The program developer
 - If the developer was involved in implementation, it may be of higher quality than if administered by others
 - If a developer was involved in the research, less desirable findings may not be presented

58

Quality of Implementation May Affect Results

- High-quality implementation may provide the best-case scenario for program effectiveness
- But avoid "cherry picking" studies based on quality of implementation
- Questions to consider:
 - Was the program implemented with fidelity to the model?
 - Was the program implemented in "real world" conditions?
 - What were the prerequisites of implementation?

59

In Summary

- A good evidence review includes and summarizes high-quality research
- Favorable results are a good place to start, but other details should be considered
- Even an effective program is not likely to work for every population or in every situation
- Ask what works, under what circumstances, and for whom?

60

ACTIVITY: Cookie!

- Now say you are having a party at your house. How would you determine which brand of cookies to buy for this party?
- How would you weigh the criteria and ratings?

61

Criteria for Evidence of Effectiveness

- An interagency collaboration was formed to make decisions about what threshold of evidence would be eligible for the evidence-based funding.
- This group weighed a number of factors:
 - Quality of the research
 - State of research on the field
 - Measurement quality
 - Effect size versus statistical significance
 - Favorable vs. unfavorable or ambiguous findings
 - Needs of communities for flexibility to match program to community context

62

Guiding Principles

- Systematic – we needed to be able to apply it the same way to every study
- Transparent – we needed to be able to show stakeholders where we got the information and how decisions were made
- Defensible – we needed a criteria that we could defend against critiques

63

DHHS Criteria for Evidence of Effectiveness

- What is in the criteria:
 - Statistical significance
 - Favorable findings
 - Quality of Research
 - Confidence in the effect
- What is reported on the website but not part of the criteria:
 - Unfavorable or ambiguous findings
 - Effect size
 - Measurement quality
 - Duration and replication of effect
 - Independence of the evaluator
 - Implementation factors

64

Resources: Federal Systematic Reviews

- Home Visiting Evidence of Effectiveness (HomVEE): <http://homvee.acf.hhs.gov/>
- Pregnancy Prevention Evidence Review (PPRER): <http://www.hhs.gov/ash/oah/oah-initiatives/tpp/tpp-database.html>
- CrimeSolutions.gov: <http://www.crimesolutions.gov/>
- FindYouthInfo.gov: <http://www.findyouthinfo.gov/>
- What Works Clearinghouse (WWC): <http://ies.ed.gov/ncee/wwc/>
- National Registry for Evidence-based Programs and Practices (NREPP): <http://www.nrepp.samhsa.gov/>
- CDC's Community Guide: <http://www.thecommunityguide.org/index.html>

65

Resources: Evidence-based Decision Making

- CDC's Understanding Evidence Guide:
http://www.cdc.gov/violenceprevention/pub/UnderstandingEvidence_pt1.html
- Evidence-Informed Public Health Toolkit: National Collaborating Centre for Methods and Tools:
<http://www.nccmt.ca/eiph/index-eng.html>
- Brownson, R.C., Fielding, J. E., & Maylahn, C. M. (2009). Evidence-based Public Health: A fundamental concept for public health practice. *Annual Review of Public Health*, 30, 175-201.
- Nutley, S. M., Walter, I., & Davies, H. T. O. (2008). *Using evidence: How research can inform public services*. Policy Press: Bristol, UK.

66
