



Obesity-Related Policy Evaluation Webinar Series

Session 3

Enhancing the Usefulness of Evidence to Inform Practice

Dr. Kathryn E. Newcomer

The Trachtenberg School of Public Policy and
Public Administration

The George Washington University

May 1, 2009

Instructional Objectives

- Provide guidance for evaluators of obesity-related policies to help them bolster the credibility of evidence of policy intervention effectiveness to policymakers
- Identify ways to present findings to bolster confidence in data
- Offer tips on documenting adaptation of interventions to inform practice

Why Is the Quality of Evidence So Important?

- Calls for evidence-based policy making, evidence-based practice, and evidence-based management are here to stay
- Foundations and professional organizations, such as the Cochrane Collaboration and the Campbell Collaboration, are drawing attention to the need for higher levels of evidence to inform public policy debate
- The federal government has embraced the notion of evidence-based management as seen in the Office of Management and Budget (OMB) guidance that random control trials (RCTs) are the “gold standard” for producing credible evidence, and the What Works Clearinghouse at the Department of Education

What Are Implications of the Focus on Evidence in Evaluation Practice?

- Higher demands placed upon evaluators to demonstrate the quality of the evidence they produce
- Lack of a clear, shared understanding about when evidence is good enough
- Anxiety about how to produce high level evidence in fieldwork where random assignment is simply not an option

Issues Addressed in This Session

1. Criteria for judging the credibility of evidence
2. Bolstering confidence in results
3. What is needed to facilitate the transferability of policy interventions

Section 1:

Criteria for Judging the Credibility of Evidence

How Might We Source Criteria for Judging Evidence?

- In the public policy arena there are multiple audiences who bring different perspectives, values, and priorities to the assessment of evidence
- Obesity policy research brings together a number of “unlikely partners” with different training and perspectives, e.g., urban planners, economists, nutritionists, geographers, psychologists, exercise scientists, public policy researchers
- Adherence to social science methodological norms and accepted practice is a place to start, but may not be enough
- Producing actionable evidence in the public domain entails consideration of policy-relevant criteria
- The Government Accountability Office conducts policy-relevant research in the public eye and offers a useful rubric, called the “rule of evidence”

The Rule of Evidence*

- ❑ **Competence:** Was the methodology used to collect the evidence competently executed by competent professionals?
- ❑ **Relevance:** Does the evidence address the evaluation question?
- ❑ **Sufficiency:** Is the evidence convincing to the consumers of the evaluation?

* The Rule of Evidence from the GAO “Yellowbook” of Government Auditing Standards

Competence

What Constitutes Competence?

- Valid and widely accepted measures (**measurement validity**)
 - Selecting credible measures is more challenging in this area, as the measures of the food and physical activity environments are “first generation measures” in many ways
- Reliability in measurement procedures (**reliability**)
- Data and logic supporting inferences about the causal linkage between the intervention and the observed outcomes (**internal validity**)

Competence Cont'd

. . . and, if samples are used:

- The ability to generalize beyond the groups or context being studied (**external validity**)
- The ability to generalize statistical findings beyond the sample (**statistical conclusion validity**)

Competence will be judged by well-founded and clearly explained decisions about...



Conveying Competence

- Communicate **clearly** and **effectively** about the choices you made and procedures you employed during evaluation design, data collection, and analysis. More to follow!

Relevance and Sufficiency

How Do We Judge?

- Experience and professional standards can help but will not dictate the “correct” methodological choices
- Resources must be balanced with rigor
- Producing “compelling” data may well be harder when you are evaluating policy interventions in the field without much control over the context
- The goal is to produce convincing data that constitute “compelling” and understandable evidence to inform practice

Section 2:

Bolstering Confidence in Results

Bolstering Confidence in:

- ❑ Measurement
 - Measures and Reliability
- ❑ Inferences
- ❑ Numbers

Measurement

- Explain why specific *measures of outcomes* were selected

For example: Based on previous work, expert judgment

- Use *tables and graphic displays* to provide supporting documentation backing up your work

For example: Use a cross walk showing how your measures compare to measures used in previous work

Health Care Access for Child Care Workers Behaviors to Be Measured (Supported by Previous Research)

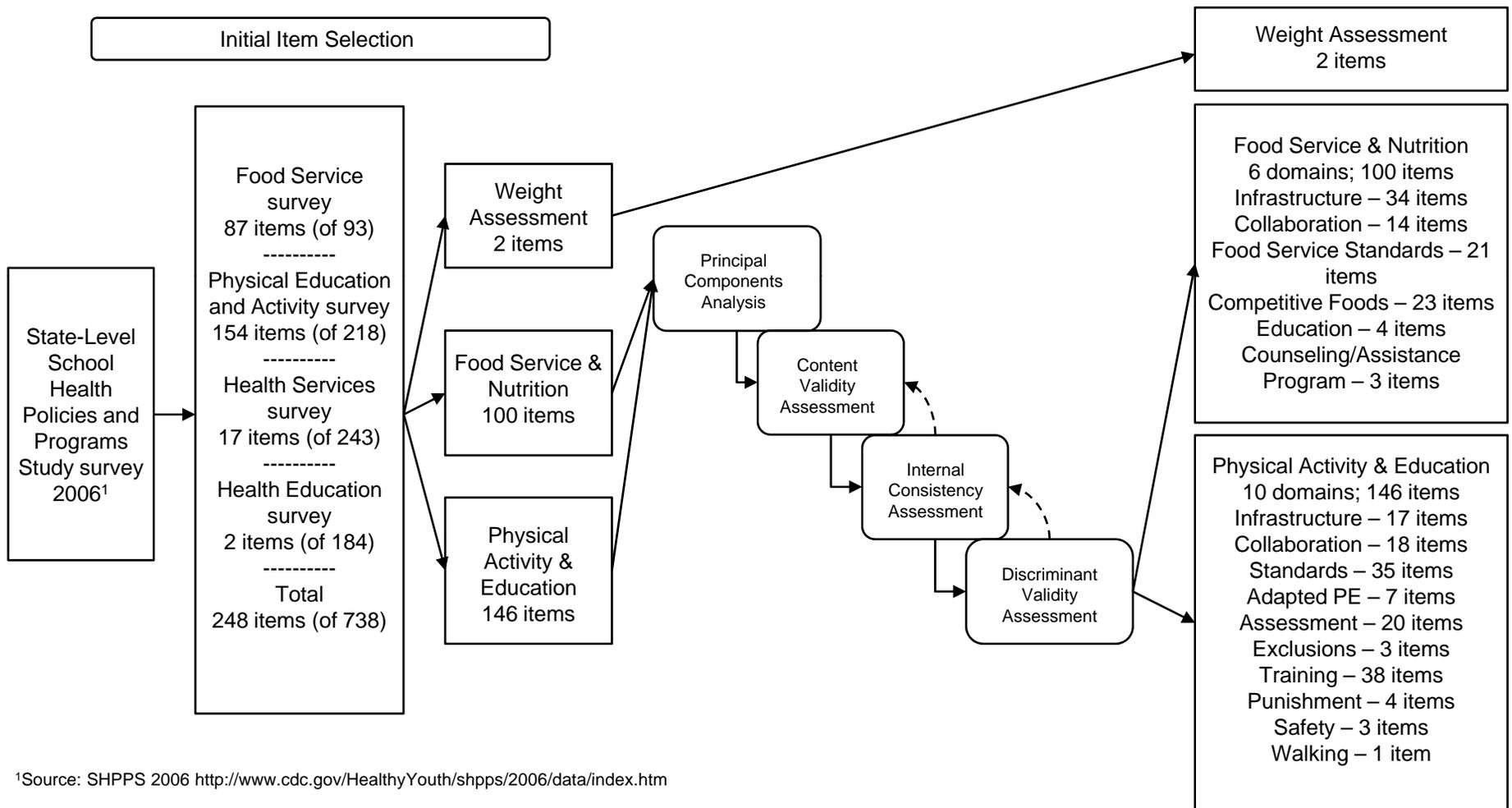
<u>Health Care Access Outcome Measures</u>	<u>Source</u>
Delay in seeking needed care (any reason)	Burstin, et al (1998); Baker, Shapiro, and Schur (2000); Ayanian, et al (2000).
Primary physician	Burstin, et al (1998); Weinick, Zuvekas, Shi (2000).
Preventative procedures (including, but not limited to, check ups, vaccines, mammograms, colonoscopies)	Burstin, et al (1998); Ayanian, et al (2000).
Annual physical exam	Baldwin, et al (2007); Burstin, et al (1998); Ayanian, et al (2000).
Use of prescription medication for diagnosed problems	Stuart and Grana (1995).
Access to contraceptive/infertility treatments	King and Meyer (1997).
Follow up visits (regardless of provider type)	Burstin, et al (1998); Hadley (2007); Fortney, Rost, and Zhang (2000).

Measurement Cont'd

- Explain clearly how the *sets of indicators* were identified and how indices were constructed from the pool of indicators

For example: Through analytical processes such as factor analyses or expert reviews (see next slide for an example)

Figure 1. Schematic diagram of the hybrid expert and empirical process used to create state obesity prevention policy domains.



¹Source: SHPPS 2006 <http://www.cdc.gov/HealthyYouth/shpps/2006/data/index.htm>

Measurement Cont'd

Reliability in Measurement Procedures

- Explain how you trained observers and interviewers so that they *consistently applied comparable criteria*
- Explain what *quality control procedures* were used to ensure consistent measurement in the field
- Describe any *inter-coder reliability checks* that were undertaken and provide statistical measures conveying reliability if used

For example: The most commonly used measure is Cronbach's alpha

Inferences

- Offer *convincing statements* that “plausibly attribute” measured outcomes to the intervention
- Identify *other events and processes* that occurred in the neighborhoods or schools that may have also affected the achievement of desired outcomes

For example: Parents in a school district may also adopt food policies from schools in their religious settings. Thus, any change in overall dietary behavior may be due in part to this adoption

Inferences Cont'd

- Discuss the *time dimension* — the possibility that the time needed for the intervention to change attitudes or behavior may be longer than the time allotted in the evaluation to measure outcomes.

For example: If the policy intervention is in schools, you may need to work with (1) the principal and (2) the wellness committee, (3) change language in the written policy, and (4) examine the impact of the policy change in practice. All of these things may take an entire school year!

Inferences Cont'd

- Describe *implementation* of the intervention with sufficient detail so that attributing outcomes to the intervention is reasonable

For example: Describe other activities that were happening at the time as the policy intervention, (e.g., change to the competitive food policy) while describing the policy change to using food for fundraising and the practice of birthday treats in classrooms, etc.

Inferences Cont'd

- Discuss explicitly to whom and to where the results can be reasonably generalized. Using the example of a new park:
 - Discuss what efforts were undertaken to corroborate results offered through self-reporting (describe results of observations at the park)
 - Explain why the sample was of sufficient size, with adequate sampling of important subgroups to permit generalizing to those sub-groups
 - Discuss the response rates for any samples
 - Clarify what efforts were undertaken to probe for systematic differences between respondents and non-respondents (i.e., non-response bias)

Numbers

- Be clear about your *denominators*

For example: are you providing rates or proportions, or both?
Always clarify

- Calculate percentages on the *appropriate base*

For example: 85% of the treatment group scored high on the criterion, not 32% of those scoring high were in the treatment group

- Distinguish between statistics showing the statistical significance of relationships, and statistics measuring the strength of relationships or the magnitude of effects

Numbers Cont'd

Credible Statistics

➤ When drawing statistical generalizations from your sample:

- Clarify the *level of confidence* you have in your results – such as 95% or 99%
- Clarify the *margin of error* around your statistical estimates; use confidence intervals around point estimates

For example: “We can conclude with 95% confidence that purchases of fruits in the cafeteria increased 8–12%,” rather than simply claiming an increase of 10%

- Use *confidence intervals* around virtually all statistics, such as differences in means

For example: “The 95% confidence interval around the difference in minutes spent watching television between girls and boys was between 14.5 and 34.5 minutes.”

Numbers Cont'd

Credible Assumptions Underlying the Numbers

- Clarify how the statistical technique used was appropriate, given the level of measurement or the distributions of the variables

For example: Do not compare means on variables with a short scale or limited variance — a 1 to 5 Likert scale, or a longer scale where most respondents chose only a 0 or 1 — but compare the distributions with a bivariate table and use the chi square statistic to test statistical significance

Provide *sensitivity analyses* when warranted — show how varying any important assumptions about rates or measures might change your estimates of results

Upgrade of Urban Cycling Routes

	<u>Before</u>	<u>After</u>
Sample	n=210	n=185
Frequency of Walking or Biking to Work (Mean) (reported on a 1-5 scale)	2.5	2.6

	<u>Before</u>	<u>After</u>
How Frequently Do You Walk or Bike to Work?	n=210	n=185
Never (1)	60	55
Several times a year	10	8
Several times a month	10	8
Several times a week	10	9
Daily (5)	10	20
Total	100%	100%

Section 3:

What Is Needed to Facilitate the Transferability of Interventions

Transferability of Interventions

- ❑ What will others need to know to replicate the intervention?
- ❑ What are reasonable expectations of results for future replications?

What Will Others Need to Know to Replicate the Intervention?

- The **context** of the policy intervention must be described in enough detail that others may know whether it is feasible to replicate the intervention. Include:
 - Relevant demographic information about the participants in the evaluation
 - Resource requirements for implementing the intervention
- The manner in which all key components of the intervention were implemented must be described in enough **detail** that others may actually implement it

What Are Reasonable Expectations of Results for Future Replications?

- A realistic **assessment of the overestimation of effectiveness** due to the methodological aspect of the evaluation, such as participants and/or staff being highly motivated to behave in the desired direction due to the novelty of the situation
- Pertinent information about **implementation processes** so that the essential elements needed to ensure appropriate implementation, and a reasonable time dimension for observing results, is clear
- A realistic **description of all training** required to implement intervention

In Sum: Expectations Matter!

1. The **current environment**, in which calls for evidence-based interventions and evidence-based policy are prevalent, has raised the stakes for production and presentation of evidence
2. Competent execution of rigorous methodology is **necessary, but not sufficient**, for production of credible evidence
3. In addition to being **competent**, evidence must be **relevant** and **sufficient** to be convincing to the audiences for your evaluations

Expectations Matter! Cont'd

4. It is critical for evaluators to clearly convey the **credibility** of the **measures, inferences, and numbers** they generate
5. In order to facilitate replication of successful interventions, detailed information about **implementation** of the intervention is necessary to allow others to implement policies and interventions in new contexts
6. Evaluators have a professional responsibility to provide **detailed information** about their evaluations, the **context for the intervention** they evaluated, and the **implementation itself** to permit successful transference to new locations

Resources

Evaluation Texts

- * Pawson & Tilley (1997): Realistic Evaluation, Sage Publications
- * Rossi, Peter & Freeman (2003): Evaluation: A Systematic Approach, Sage Publications
- * Weiss (1997): Evaluation, 2nd Ed. Prentice Hall
- * Wholey, Hatry & Newcomer Eds. (2004): The Handbook of Practical Program Evaluation, Jossey-Bass