Obesity-Related Policy Evaluation Webinar Series

Session 2

Addressing Pitfalls to Research in Real World Settings

Dr. Kathryn E. Newcomer

The Trachtenberg School of Public Policy and Public Administration

The George Washington University

April 3, 2009

Instructional Objectives

- Identify common obstacles to collecting credible evidence in the field
- Provide suggestions for addressing typical constraints on field research

Issues Addressed in This Session

1. Brief Recap of Webinar 1
2. Evaluation Approach
3. Pitfalls **Before** Data Collection
4. Pitfalls **During** Data Collection
5. Pitfalls **After** Data Collection

Section 1:

Brief Recap

What Do We Mean by Obesity-Related Policies?

- Public policies at federal, state, and local levels that may affect diet/physical activity and/or reduce obesity prevalence

  Examples:
  - Federal mandated school wellness policies
  - Calorie labeling in restaurants
  - Fruit and vegetable carts, e.g., New York City
  - Implementation of parks or bikes trails
  - Federal funding to create incentives for bicycle commuting
  - Government subsidies/vouchers to purchase fruit and vegetables

- Contrasts with individual-level, behavioral interventions

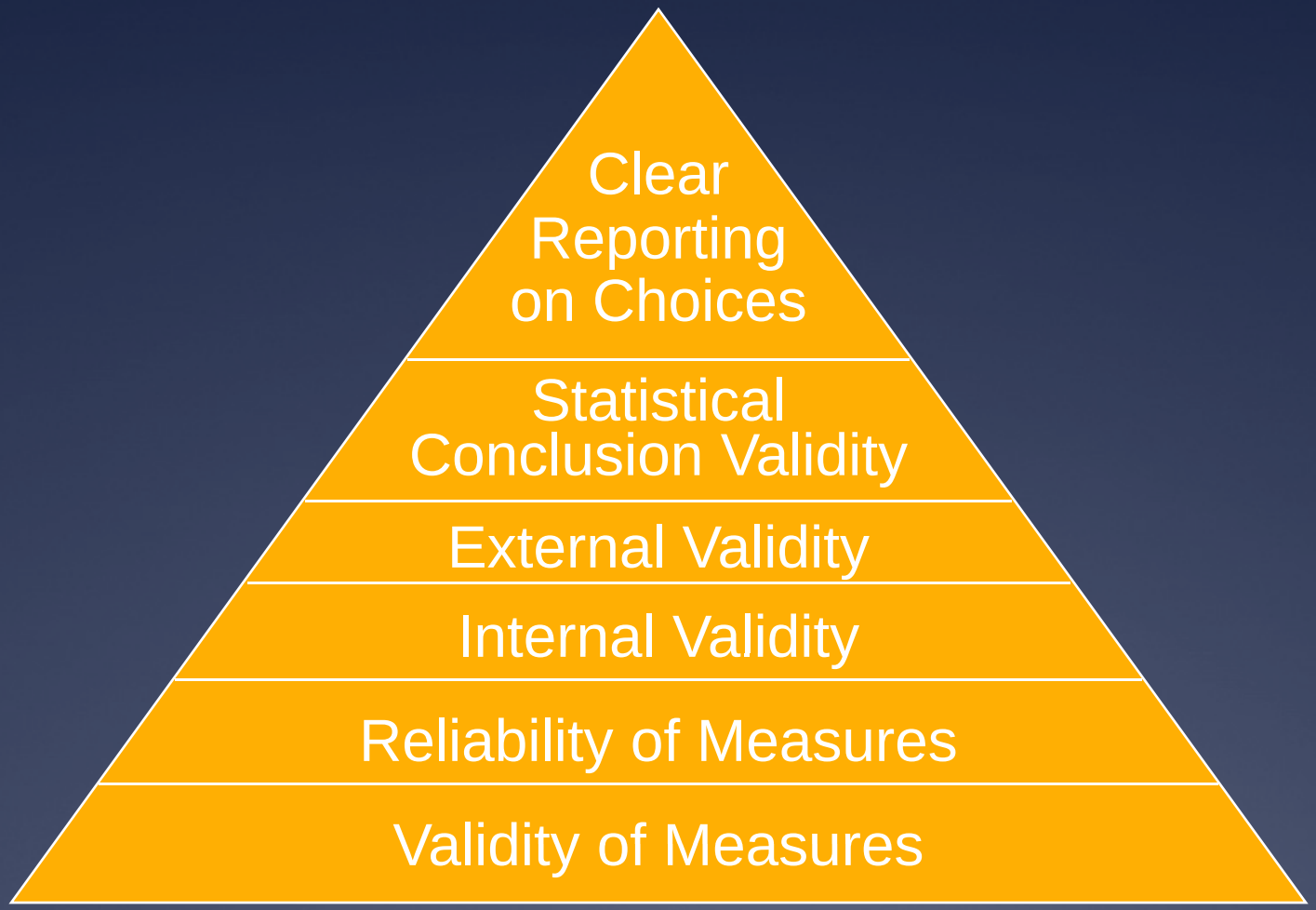
These Policies Operate in the Outer Two Rings of the Social Ecological Model

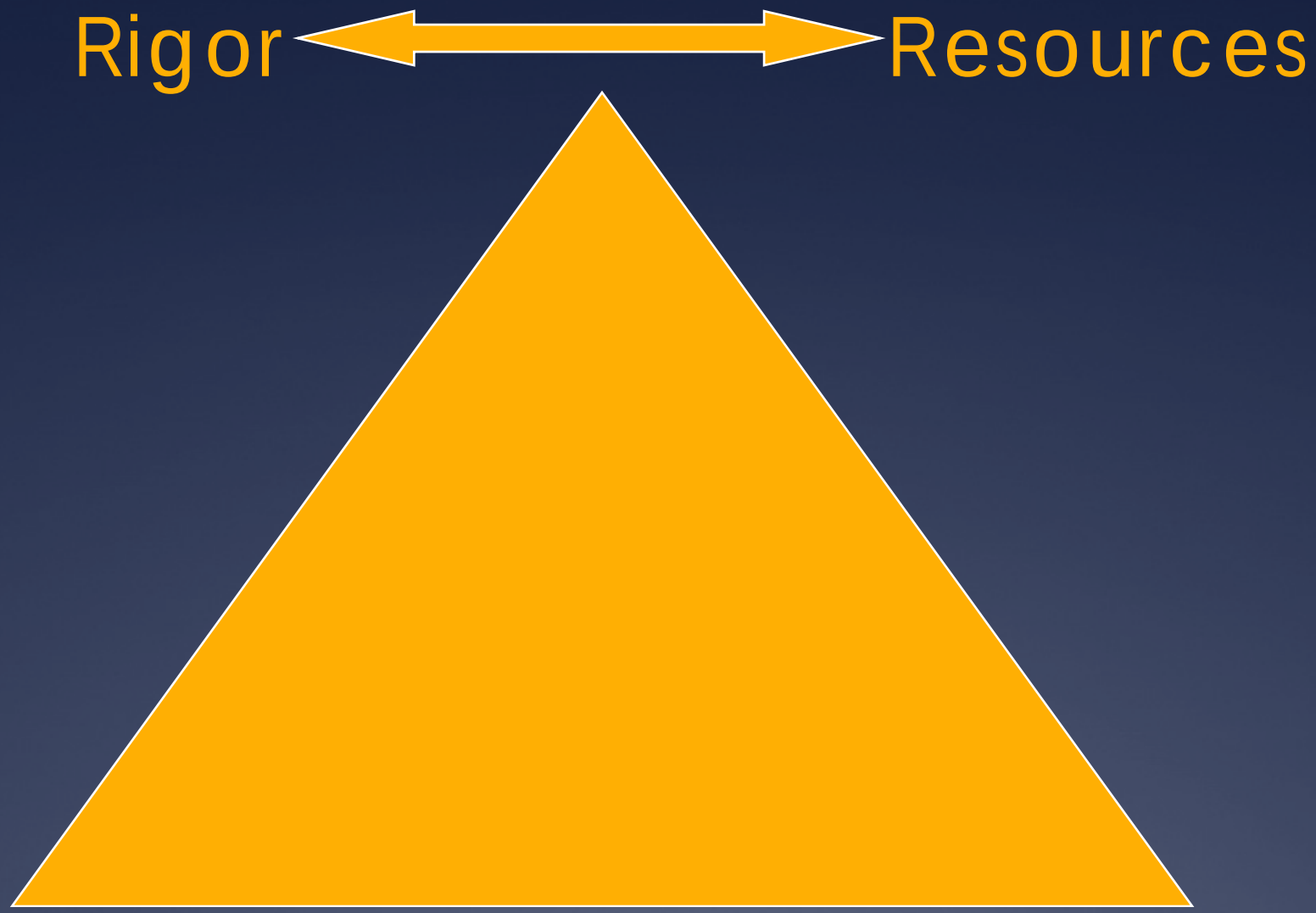
Source: Orleans, 2007

Credibility of our evaluation work is dependent upon…

the **Methodological Integrity** of our work!

Methodological Integrity is affected by a variety of decisions about...





There is a delicate balancing act between methodological rigor and the resources available!

Validity of Measures

Measurement validity is concerned with the accuracy of measurement

Are we accurately measuring what we intend to measure?

Reliability of Measures

Reliability is the extent to which measurement can be expected to produce similar results on repeated observations of the same condition or event

- Reliable measures mean that operations consistently measure the same phenomena
- Reliable measurement means consistently recording data with the same decision criteria

Internal Validity

Internal validity is concerned with our ability to determine whether X caused Y and in what magnitude

Are we able to establish definitively whether there is a causal relationship between a specified cause and potential effect?

External Validity (Generalizability)

External validity is concerned with our ability to generalize beyond the groups or context being studied

Are we able to generalize from the results?

Statistical Conclusion Validity

Statistical conclusion validity is concerned with our ability to detect an effect, a relationship, or a factor, if it is present, and the magnitude of an effect

Do the numbers we generate accurately detect the presence of a factor, relationship, or effect?

Section 2:

Evaluation Approach

What is Your Evaluation Approach?

Focus

To what extent will the evaluation be concerned with assessing implementation processes and to what extent will it measure the outcomes *or* impact of the intervention?

For example: In changing a school policy of increased physical education (PE), do you focus on measures such as:

- Numbers of PE teachers per school district, or changes in student minutes for PE per school district

or

- Students' overall activity levels, or students' BMI

Note: Even if impact is the primary focus, examine implementation!

Evaluation Approach Cont'd

Inquiry

Is the evaluation focused solely on measuring achievement of stated objectives, or will it include some open-ended exploration for unintended effects of the intervention (positive or negative)?

For example: Will you survey principals, teachers, and/or students for their feedback on the required increase in PE?

- Unanticipated consequences may possibly include reduced time for recess and lunch breaks or increases in test scores.

Evaluation Approach Cont'd

Data

What sorts of data will be required to address the evaluation questions?

- Quantitative or qualitative, or both?
  - E.g., you may develop or use an instrument to code language in a school policy for strength of implementation, or use semi-structured interviews with teachers, principals, and others
- Data collected solely by neutral observers, or by study participants as well?
- Cross-sectional or longitudinal?
  - I.e., collect data at one point in time—or over several points, before, during, and after implementation

Evaluation Approach Cont'd

Comparisons Desired

What sorts of comparisons will be helpful to demonstrate the effectiveness of the policy intervention?

- Pre- and post-intervention observations of one of more treatment groups? E.g., between school districts that implement the same policy change.
- Differential effects for different sub-groups of policy beneficiaries? E.g., between age groups.
- Post-intervention observations of a treatment group and comparison group? E.g., between a school district that changes its policy regarding PE versus one that does not.

Methods Choices Matter

The choices we make, as well as unforeseen obstacles we encounter, when conducting evaluation work typically affect more than one of the types of validity and reliability.

Therefore, think systematically about how your findings are affected when choices are made prior to, during, and after data collection!

Section 3:

Pitfalls **Before** Data Collection

Policy, Program, or Theory Feasibility

Preparation for Data Collection

Policy, Program, or Theory Feasibility

1. Failure to assess whether the policy or program is evaluable

   - Investigate where the school district policy been implemented and sufficiently explained at the school level and where it has not
   - Ask administrators and teachers whether they have received sufficient training about new curricula or procedures
   - Find baseline data—or construct a strategy to reconstruct some baseline data

   For example: If evaluating calorie labeling, and one restaurant implemented changes before others, survey regular customers and/or store owners for their recollection baseline conditions.

Threats to methodological integrity

Policy, Program, or Theory Feasibility Cont'd

2. Starting data collection too early

   - Examine whether new calorie labeling requirements have been introduced long enough so that store or restaurant customers are aware and have become comfortable reading them
   - Examine the extent to which a new curriculum or training program been completely implemented
   - Review previous research (and theory!) to identify an adequate timeframe for physical activity or eating behavior modification to start showing effects

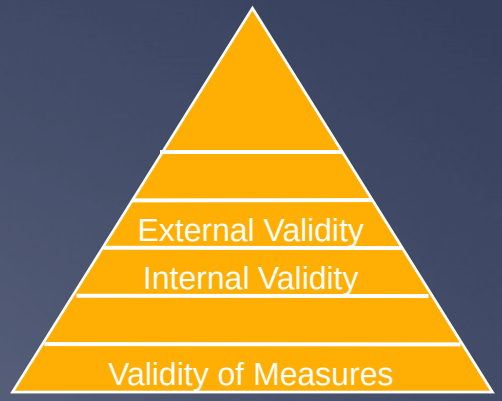
Threats to methodological integrity

Policy, Program, or Theory Feasibility Cont'd

3. Not devoting sufficient time and deliberations to identifying criteria for measuring implementation and outcomes

   - Review previous research to identify widely accepted evaluation criteria and any flaws in previously used criteria

   - Examine previous research (or contact experienced researchers) on measures used in calorie labeling evaluations, such as sales receipts, menu checklists, surveys of customers for changes in behavior or knowledge



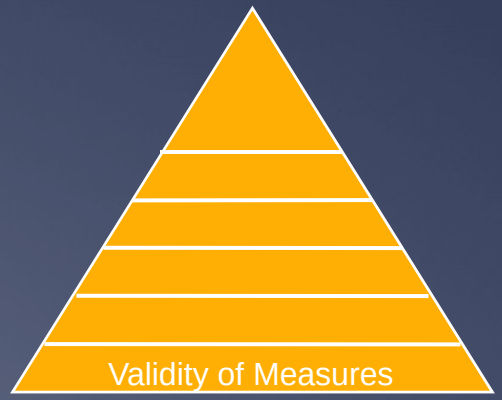
Threats to methodological integrity

Policy, Program, or Theory Feasibility Cont'd

4. Failure to identify a comprehensive set of measures to detect both intended and unintended outcomes of the intervention

   - Make plans to use multiple data collection methods to detect implementation, behavioral, and other outcomes from more than one source

   - Examine cafeteria sales receipts and menu options, in addition to conducting interviews with students or teaching staff



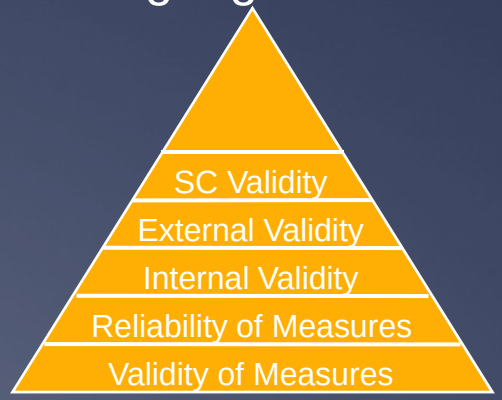
Threats to methodological integrity

Preparation for Data Collection

5. Failure to pretest data collection instruments appropriately

   - Pretest surveys, interview protocols, diaries, and observation checklists with samples of the appropriate respondents—such as teens, children, parents of diverse educational backgrounds, and the elderly

   - Ensure that surveys, interview protocols, diaries, and observation instruments, e.g., vending machine inventories, have been translated appropriately into an adequate number of languages



Threats to methodological integrity

Preparation for Data Collection
Cont'd

6. Use of too few indicators of intended program or policy outcomes

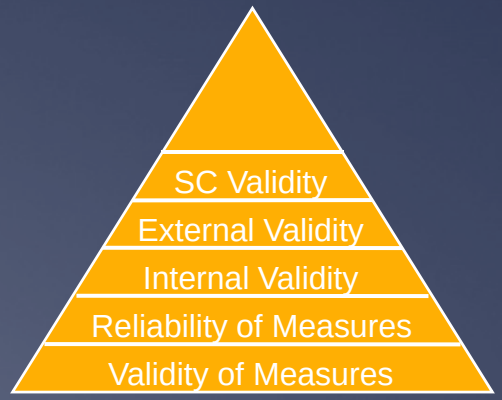
- Develop backup plans to ensure that you are able to collect sufficient data on the targeted outcomes
- Identify several ways to record:
  - ✓ Changes in PE requirements in schools, such as evaluating language in policies, as well as actual minutes of PE in schools via survey (observed, and/or self-report from principals/teachers)
  - ✓ Physical activity, such as using accelerometers in addition to activity diaries, counting bicycles, active play on playground count/observation

Threats to methodological integrity

Preparation for Data Collection
Cont'd

7. Inadequately training data collectors

- Make sure all data collectors have received sufficient training with the use of surveys, interview protocols, and observation checklists. For example, how to count broken street lamps, record condition of sidewalks and playground equipment safety appropriately
- Whenever pretests identify possible areas in which clarification may be required, provide the collectors with clear guidance on how to provide such clarity
- When there are likely to be respondents of multiple cultural backgrounds, make sure that data collectors have received adequate guidance about potentially relevant cultural sensitivities



Threats to methodological integrity

Section 4:

Pitfalls **During** Data Collection

Research Procedures

Measurement Constraints

Reactivity

Composition of Sample

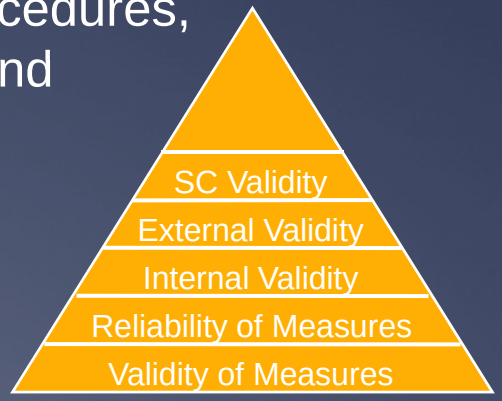
Flawed Comparisons

Research Procedures

8. Failure to identify and adjust for changes in data collection procedures that occur during the measurement period

   - Whenever observers who are using checklists or surveys in the schools, stores, or restaurants report important additional aspects that were not originally included, ensure that they are added and revisits made

   - Whenever access to respondents, or willingness to participate in interviews or in other collection procedures, is lower than expected, make efforts to identify and include other respondents



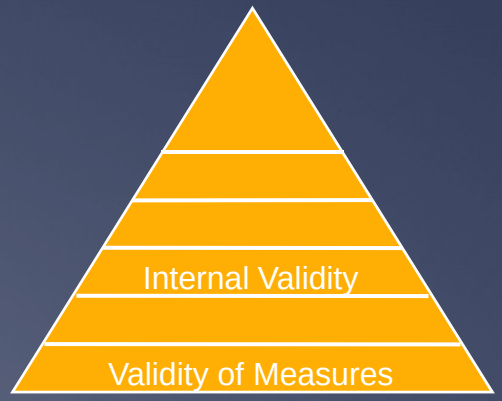
Threats to methodological integrity

Research Procedures Cont'd

9. Collecting too much data and not allowing adequate time for analysis of the data collected

   - When employing semi-structured interviews and focus groups, provide adequate resources to conduct time-consuming qualitative data analysis

   - Plan for adequate time for environmental audits, driving time, weather issues, store permission access



Threats to methodological integrity

Measurement Constraints

10. Inadequate conceptualization and/or measurement of the "intervention"

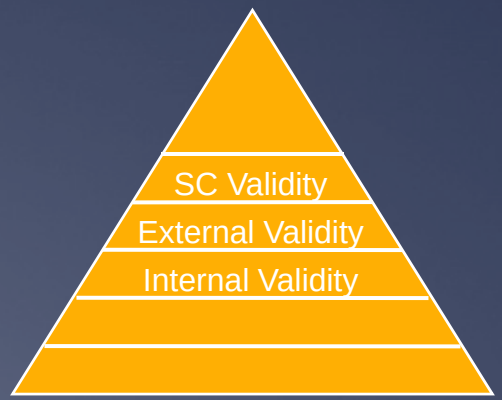
- Devote sufficient attention and time to asking administrators, teachers, or other care providers about the completeness of implementation of training or provision of healthy food options in schools
- Ask exactly what changes were made to menus in schools, not simply if any changes were made

Threats to methodological integrity

Measurement Constraints
Cont'd

11. Beginning observation when conditions or target outcomes are at an extreme level

    - Check if levels of obesity or disease rates were at higher than average levels in the schools or communities where the intervention was located, or relatively high number of vending machines in one school, or school breakfast participation is the lowest in the state.



Threats to methodological integrity

Reactivity

12. Inappropriate involvement of administrators in data collection

- If administrators in schools, community centers, or health facilities who have a stake in showing positive outcomes are involved in collecting data or reporting on behaviors (sometimes called social apprehension in responding), check on overestimation of effects by comparing the magnitude of change to previous research

Threats to methodological integrity

Reactivity Cont'd

13. Overly intrusive data collection procedures that change behaviors of staff or participants

- Check if administrators in schools, community centers, or health facilities who have a stake in showing positive outcomes in positions where they could alert participants of desired outcomes changed their behaviors to encourage the desired behaviors—leading to some contamination

- Ask if target participants were given additional incentives to show positive outcomes since they knew they were being observed, i.e., the "Hawthorne effect"

Threats to methodological integrity

Composition of Sample

14. Failure to account for drop off in sample size due to attrition

- When more than one observation is needed for each participant or unit (e.g., a school, classroom, restaurant, or grocery store), and there are difficulties gaining access for the follow-up measurement, make additional efforts to ensure the number of units with multiple measures achieves generalizable sample sizes
- *See Webinar 1 for more information on this!*

Threats to methodological integrity

Composition of Sample Cont'd

15. Failure to draw representative sample of program participants

- Examine previous relevant evaluation work to anticipate response rates among specific targeted populations
- Keep checking during data collection to ensure that adequate numbers of all subgroups of interest will be included in the final sample

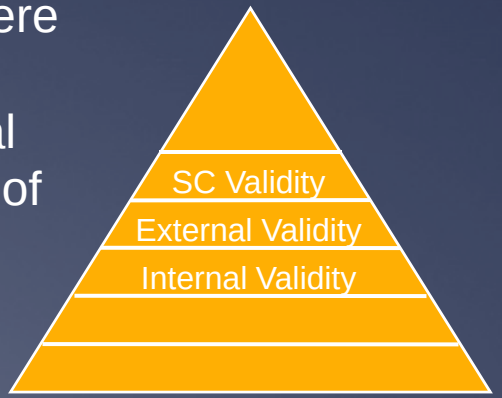
Threats to methodological integrity

Composition of Sample Cont'd

16. Insufficient number of callbacks or efforts made to protect against "non-response bias"

- When there are low response rates due to inadequate access or obstacles such as answering machines or non-English speaking respondents, make additional efforts to ensure that response rates for entire samples and important subgroups permit generalizable results

- Test while data collection is still ongoing to see if there are systematic differences in the demographics of respondents and non-respondents so that additional efforts may be undertaken to reach a good number of non-respondents through different means



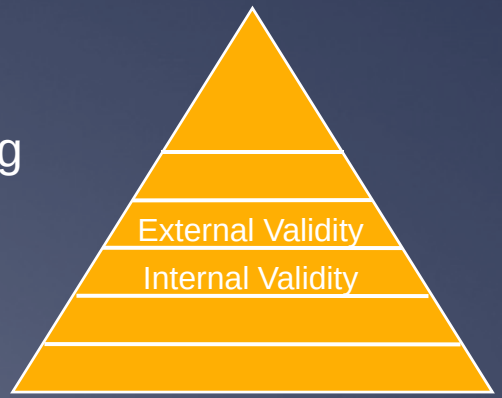
Threats to methodological integrity

Flawed Comparisons

17. Over-estimation of effects where comparison groups were not available

- When inclusion of truly comparable comparison units, such as schools or neighborhoods, are simply not available, estimate the effectiveness of an intervention by comparing the effects measured to those found in evaluations where comparison groups were included

For example: If studying the effects of calorie labeling, and a relevant comparison community is not available for your study (perhaps your comparison community also implemented calorie labeling, or your study funding does not allow comparison), review the results of other studies for results of any effects of the policy change and compare to your findings

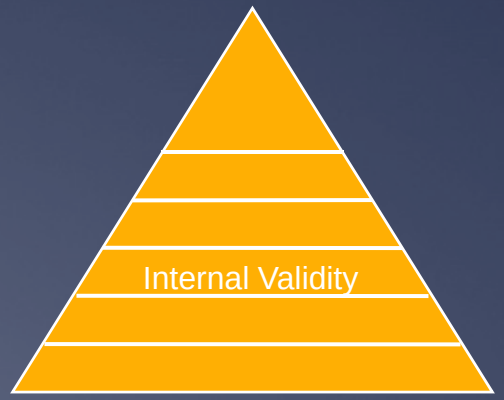


Threats to methodological integrity

Flawed Comparisons Cont'd

18. Failure to take into account key contextual factors out of the control of administrators that affect policy/program outcomes

    - Ensure that all data collectors and observers ask questions to ensure that they identify other experiences, such as education or guidance, that respondents may have received through other programs or influences outside of the school or neighborhood setting, for example, that may also positively or negatively affect outcomes



Threats to methodological integrity

Section 5:

Pitfalls **After** Data Collection

Data Analysis

Presentation of Results

Data Analysis Cont'd

19. Focusing only on overall results with inadequate attention to disaggregated results
    - Test effects for important subgroups of the whole sample to ensure that:
        - ✓ Results that are observable for only specific large subgroups are not making it appear that the intervention is effective for the entire sample
        - ✓ Inappropriate generalizations are not made about the relative effectiveness of the intervention for key subgroups

For example: If effects were seen only in a certain grade across schools, or within a subgroup of schools, or for overweight/obese children but not those of normal weight, do not generalize findings to all grades, all schools, or all children

Threats to methodological integrity

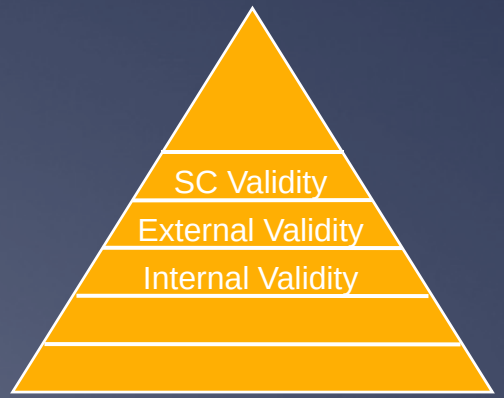
Data Analysis Cont'd

20. Applying an analytical technique without meeting important assumptions about the data

- When using a technique that analyzes the effects of more than one variable on an outcome of interest, ensure key assumptions are met

  For example: when using regression techniques, policies should be incorporated appropriately in the models (e.g., as a dummy variable)

- There are sophisticated techniques that can be employed to correct for the impact of such a key unobservable (omitted variable) when multivariate techniques are employed



Threats to methodological integrity

Presentation of Results

21. Over emphasis on statistical significance and under emphasis on practical significance of effect sizes

- Describe findings in terms of both statistical significance—which largely reflects sample size—and the policy-relevant importance of the changes in behavior, weights, or health outcomes that resulted from the intervention, and perhaps, for example, in costs savings connected with observed changes
- Effect sizes should be provided in terms that are meaningful to the audiences, perhaps in standard deviations or percentage changes

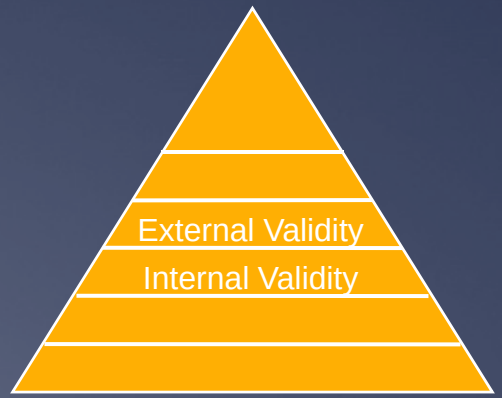
Threats to methodological integrity

Presentation of Results Cont'd

22. Generalizing beyond the confines of the sample or the limits of the study sites collected

- Provide sufficient detail on key contextual variables (e.g., neighborhood SES, race/ethnicity) both on the participants and/or community and on the nature of implementation so that others can determine when and how to adapt the intervention in another setting

- Be humble and precise in suggesting how far results can be generalized



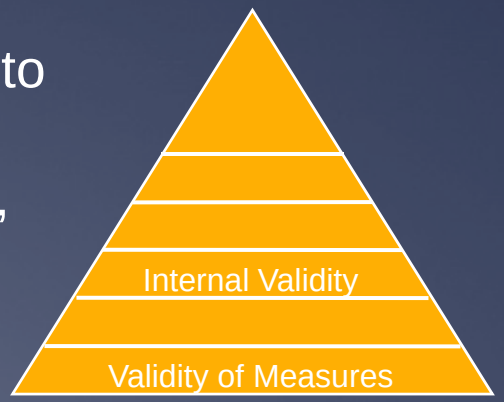
Threats to methodological integrity

Presentation of Results Cont'd

23. Failure to acknowledge the effects of multiple policy components

- Provide clear descriptions of each of the different components involved in the intervention and implementation—as well as analysis of the relative effectiveness of different components and the needed synergies among them—so others can understand how the intervention might be replicated in other settings

- *For example:* Implementing a breakfast policy to eat in classrooms. May wish to capture information on those who ate breakfast before, any improvements in school attendance, time taken from classroom instruction, janitorial response to any extra trash



Threats to methodological integrity

Presentation of Results Cont'd

24. Failure to adequately support conclusions with specific data

- Ensure that all inferences about the effectiveness of the policy intervention are supported with both logic and empirical support

  For example: do not make statements that imply that a school-level intervention could be effective in any school despite the fact that the evaluation assessed the intervention only in inner-city schools

Threats to methodological integrity

Presentation of Results Cont'd

25. Poor presentation of evaluation findings

- Make sure that methods choices and all potential methodological limitations to findings are clearly and sufficiently explained, so that future implementations of the policy intervention and future evaluations can benefit
- Offer recommendations for improving the quality of measurements that emerged, so that your evaluation can be used to guide future evaluations

Threats to methodological integrity

In Sum

1. The methodological integrity of your work may present a vulnerable target for detractors of your findings.
2. Upfront planning before data collection commences, good pre-testing of instruments, and training of all staff helps but does not guarantee that pitfalls won't hinder the ability to provide credible findings.
3. Unanticipated constraints and insufficient redundancy in measurement also can hinder your ability to produce credible results.
4. After the data are in, insufficient consideration of the applicability of analytical techniques and/or weak presentation format can reduce the credibility of findings.
5. Candid and humble reporting of limitations to the validity and reliability of findings is absolutely necessary to forestall critics.

Resources

Evaluation Texts

- 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