Producing High-Impact Results: Where Do You Start?

Dr. Kathryn E. Newcomer

The Trachtenberg School of Public Policy and Public Administration

The George Washington University

June 17, 2009

This Session is Designed to Help You:

- Recognize critical considerations to address during the planning phase, including…

  - Thinking critically about the desired end product(s) of your work

  - Identifying ways to strengthen evidence

  - Supporting compelling findings and feasible recommendations

Common Steps

- Most evaluations and/or inspections can be broken down into four main steps:
  - Planning
  - Data Collection
  - Analysis
  - Writing

Think About Your Team's Time Allocation

- What proportion of the total amount of time spent on a study do you typically devote to:
  - Planning?
  - Data Collection?
  - Analysis?
  - Writing?

Time Spent Matters

- Planning takes time, but it saves time
- For example…
  - Insufficient time spent anticipating obstacles can lead to great effort to overcome those obstacles when in the field
  - A lack of contextual understanding can reduce the feasibility of recommendations

During planning deliberations, brainstorm about…

- Targeting your audience(s) appropriately

- Determining what type(s) of evidence will be compelling

- Anticipating challenges in addressing objectives and questions

- Identifying potential difficulties with data sources, data collection, and analysis

Think Strategically…

About the desired end product(s)

Envision the End Product from the Beginning

- What do the audience(s) expect?
- What are the key comparisons you will need to provide?
- Will you need multiple types of evidence?

Envision the End Product from the Beginning

- Will you provide data tables?
- What sorts of disaggregated analyses will you need to provide?
- Will you want descriptive quotes, stories or pictures to support the presentation?

Think Strategically about Reporting

- Plan to convey the methodological integrity of your work from start to finish
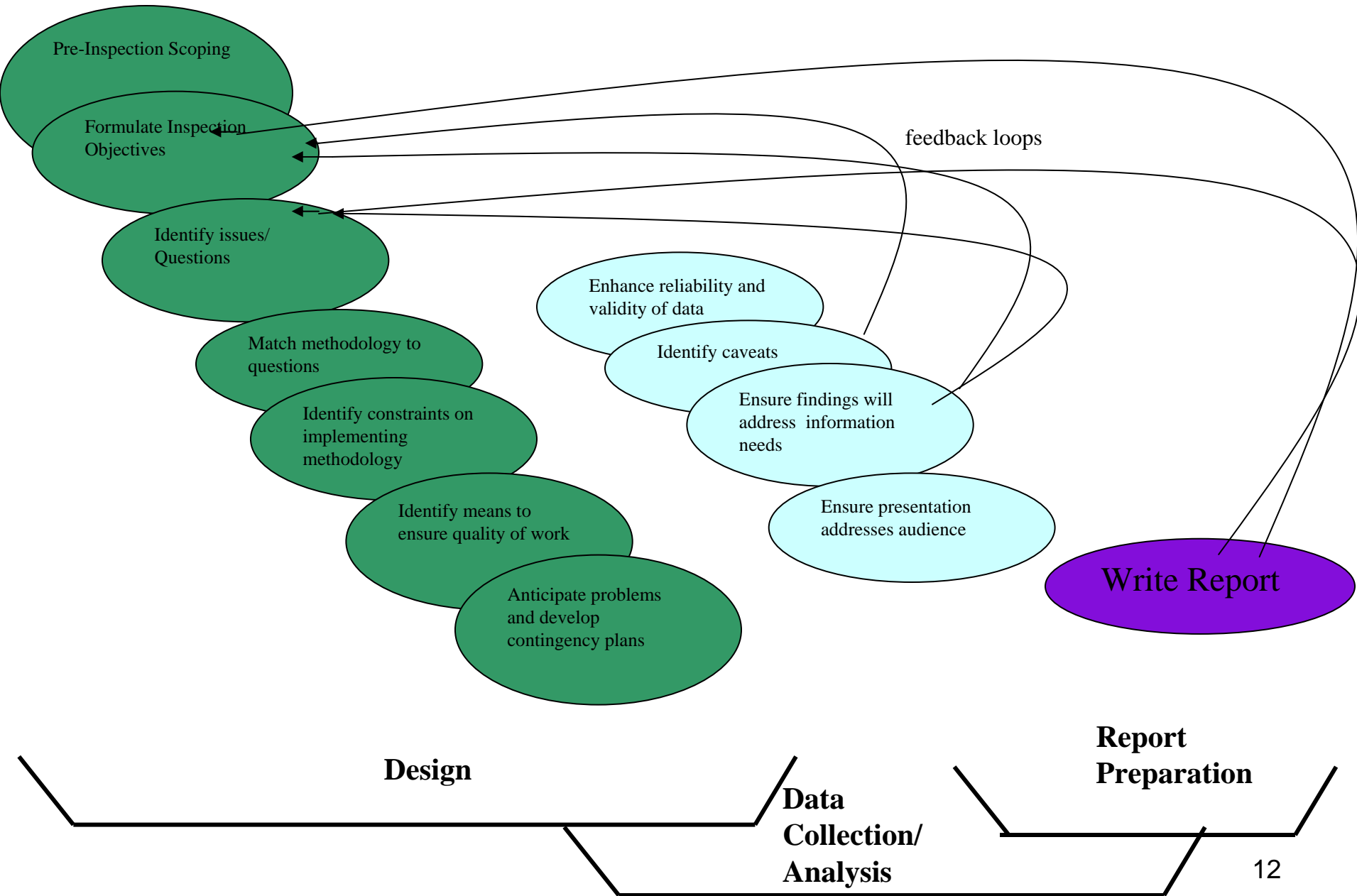- Think through what you want your key messages to be

Think Strategically About The Evaluation Process

- View the inspection process as a systemic, yet not always linear, process
- And keep your focus on the desired outcomes, i.e., the answers to well framed questions

The Non-Linear Evaluation Process

Pre-Inspection Scoping

Formulate Inspection Objectives

Identify issues/ Questions

Match methodology to questions

Identify constraints on implementing methodology

Identify means to ensure quality of work

Anticipate problems and develop contingency plans

Enhance reliability and validity of data

Identify caveats

Ensure findings will address information needs

Ensure presentation addresses audience

feedback loops

Write Report

Design

Data Collection/ Analysis

Report Preparation

12

Objectives Drive Planning:
But Questions Drive Your Work

Objectives versus questions:

- Objectives capture what the evaluators want to achieve, e.g., measure conditions, identify successful practices, assess compliance, estimate error rates

- Specific, focused questions state in measurable terms what will be addressed with empirical data

Framing Questions Well is Key

Basic types of researchable questions:

- Descriptive
- Compliance
- Effectiveness
- Prospective
- Explanatory

Framing Questions (Continued)

- **Descriptive:**
  - What are the policies within Commerce and NOAA that govern the creation, review, and dissemination of scientific research, particularly press releases and fact sheets?
  - Do Medicare contractors have uniform post-payment procedures to identify and recover payments for deceased beneficiaries?

Framing Questions (Continued)

- **Compliance:**
  - Was the hurricane fact sheet released on NOAA's web page in September 2006 created, reviewed, and released consistent with NOAA and/or Department policies?
  - Are suppliers in South Florida complying with Medicare Supplier Standards?

Framing Questions (Continued)

- **Effectiveness:**
  - To what extent have the Justice Department's efforts to implement the Sex Offender Registration and Notification Act increased the number of fugitive sex offenders investigated, arrested, and prosecuted by the Department?
  - To what extent was the FAA's network intrusion-detection capability effective in monitoring ATC cyber-security incidents?

Framing Questions (Continued)

- **Prospective:**
  - What are promising approaches to reduce child support arrearages that warrant broad consideration?

Framing Questions (Continued)

- **Explanatory:**
  - How are nursing homes coordinating with State and local emergency preparedness agencies to prepare their response to hurricanes?

Questions Indicate Study Needs

- Compliance, effectiveness, and prospective questions require criteria

- Effectiveness and prospective questions require consideration of cause-effect logic and comparisons, e.g., counterfactuals – what would have happened (or will happen) without the program or policy?

- Explanatory questions typically require qualitative data collection processes

Questions to Findings

- For each question, we need to envision the appropriate:
  - Evidence/Data Needed
  - Likely Data Source
  - Data Collection Strategy
  - Contingency Plans for Collecting Data
  - Data Analysis Plan
  - Potential Limitations to the Data
- However, the questions and data plans may need to be modified along the way.

Message #1:

Pay Now or Pay Later

Strengthen Evidence

Focus on the Quality of Evidence Throughout The Evaluation Process

- Calls for evidence-based policy making, evidence-based practice, and evidence-based management have raised expectations

- The political context surrounding OIG work calls for higher standards of evidence and transparency than in many arenas of evaluation

Criteria for Judging Evidence

Quality Standards: Rule of Evidence

"Evidence supporting inspection findings, conclusions, and recommendations should be **sufficient, competent and relevant** and should lead a reasonable person to sustain the findings, conclusions, and recommendations."

Judging Evidence

How do we judge sufficient, competent, and relevant?

- Experience and professional standards can help but will not dictate the "correct" choices.
- Professional judgment plays a big part in deciding what and how to evaluate.
- Think about what will constitute "compelling" evidence.
- Recognize the diversity in the audience's perspectives on evidence, e.g., lawyers, economists.

What Constitutes Competence?

- Valid and widely accepted measures
- Reliability in measurement procedures
- Data and logic supporting inferences about conditions and recommendations

What Constitutes Competence? - Continued

. . . and, if samples are used:

- the ability to generalize beyond the groups or context being studied, and

- the ability to generalize statistical findings beyond our sample

How Is Relevant and Sufficient Judged?

- The goal is to produce convincing data that constitute "compelling" and understandable evidence to support findings and recommendations
- Resources must be balanced with rigor

Strengthening Evidence

- Questions and circumstances will limit our choices of evidence

- We may have different levels of comfort with different types of evidence, but beware of the "law of the hammer"

Anticipate and Address Limitations

Anticipate pitfalls to:

- strengthening the quality of our evidence

- ensuring findings and recommendations are credible and supportable

Common Evidence Pitfalls

1. **Failure to assess whether a policy or program is evaluable yet – will evidence even be available?**

2. **Not devoting sufficient time and deliberations to identify criteria for measuring quality, implementation and outcomes.**

3. **Failure to assess the quality, completeness, and accuracy of the data.**

Common Pitfalls - continued

4. **Failure to pretest data collection instruments and train data collectors appropriately.**

5. **Failure to collect sufficient data to support findings and recommendations.**

6. **Failure to address non-response issues and other sample size issues.**
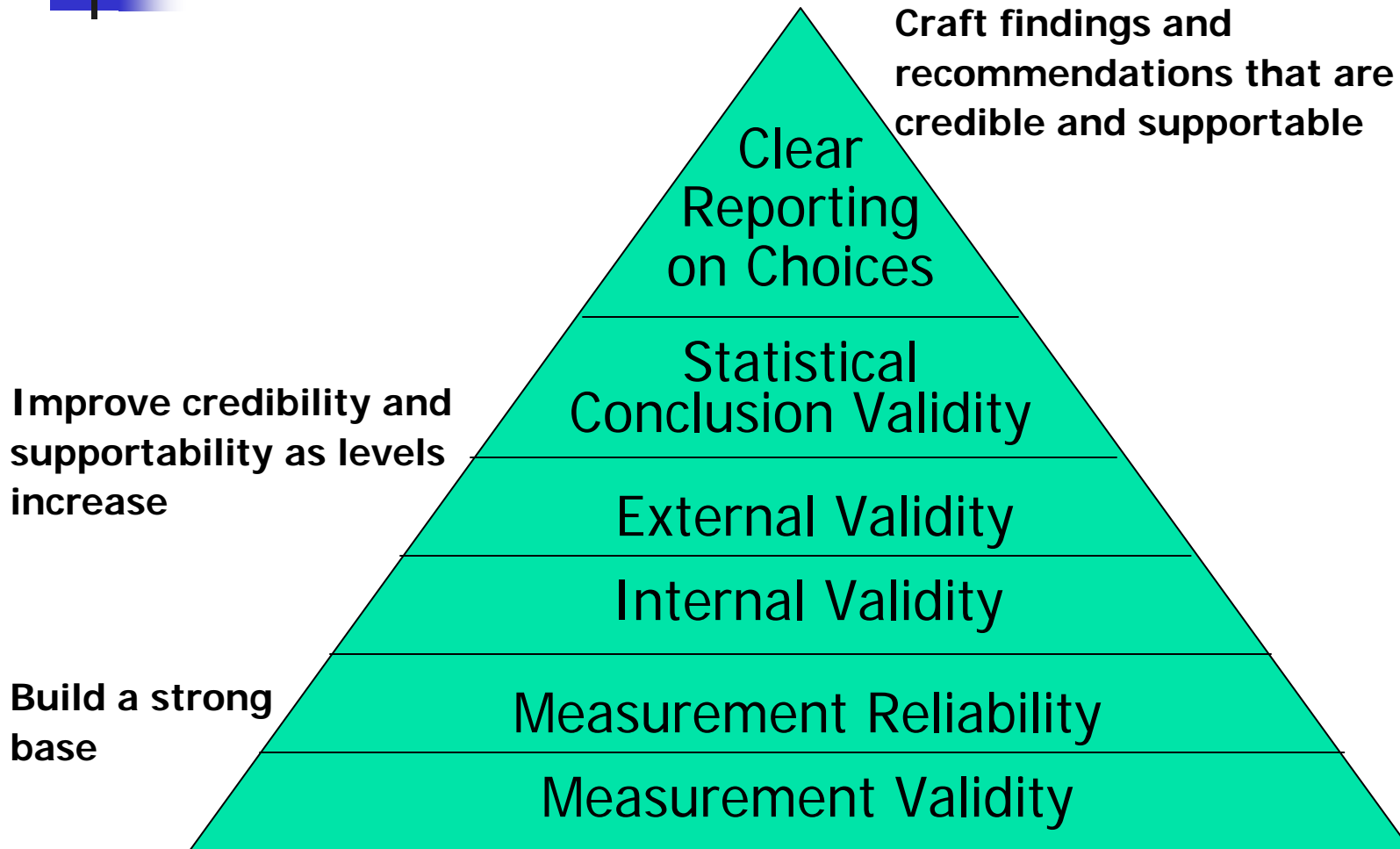
Common Pitfalls - continued

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

8. **Generalizing beyond the confines of a sample, or the limits of study sites.**

9. **Failure to adequately support findings and recommendations with specific data**.

10. **Incomplete or unclear presentations of analyses.**

Provide Credible and Supportable Findings and Recommendations

The Pyramid of Strength

Craft findings and recommendations that are credible and supportable

Clear Reporting on Choices

Statistical Conclusion Validity

Improve credibility and supportability as levels increase

External Validity

Internal Validity

Build a strong base

Measurement Reliability

Measurement Validity

Pyramid Base: Measures Need to be Valid and Reliable

<u>Measurement Validity</u>: are we measuring what we claim to measure accurately?

<u>Measurement Reliability</u>: will measures we use produce similar results on repeated observations of the same condition or events?

Strong Base: Face Validation

Questions to ask ourselves:

- Are the measures (or records) generally accepted by the administrators as accurate?
- Have you consulted experts to get their views on the adequacy of any newly constructed measures?

Strong Base: Content Validation

Question to ask ourselves:

- Rather than relying on only one measure, do you have a set of measures that experts view as adequately representing the potential pool of similar measures?

Strong Base: Criterion Validation

Question to ask ourselves:

- Rather than use testimonial evidence from one source, have you employed alternative ways to get estimates of the condition, also known as **triangulation,** that are respected as extremely reliable?

Strong Base: Predictive Validation

Question to ask ourselves:

- Do the measures predict subsequent behaviors in ways consistent with existing theory and knowledge and our past experience?

Measurement Reliability Means:

- Operations will consistently measure the same phenomena, and
- Data will be recorded accurately and consistently using the same decision criteria.

Typical Limitations With Measurement Reliability

- Variability in record-keeping across sites, care providers, or jurisdictions?

- Inter-coder or interviewer variation?

- Differential responses from members of different cultural or age groups?

Mid-Levels of Pyramid

1. **Internal validity**: the relationship between a specific "cause" and the observed (or intended) effects

2. **External Validity**: the ability to generalize beyond the groups or content studied

3. **Statistical Conclusion Validity**: the ability to generalize beyond a sample (make quality projections to study population)

Internal Validity

Internal validity is concerned with our ability to determine whether a policy, rule or regulation (X) caused (or will cause) the intended outcome or result (Y) and in what magnitude:

- Are we able to make a causal inference between a policy/rule and observed effects?
- Are we confident that a recommended action will have the intended effect?

Internal Validity in Evaluations

- **Condition**: is typically what an evaluation is examining, and it constitutes the "dependent variable" (Y) in research terminology, as it is the targeted outcome, such as satisfaction of Medicare beneficiaries, or mismanagement, or quality care in nursing homes.
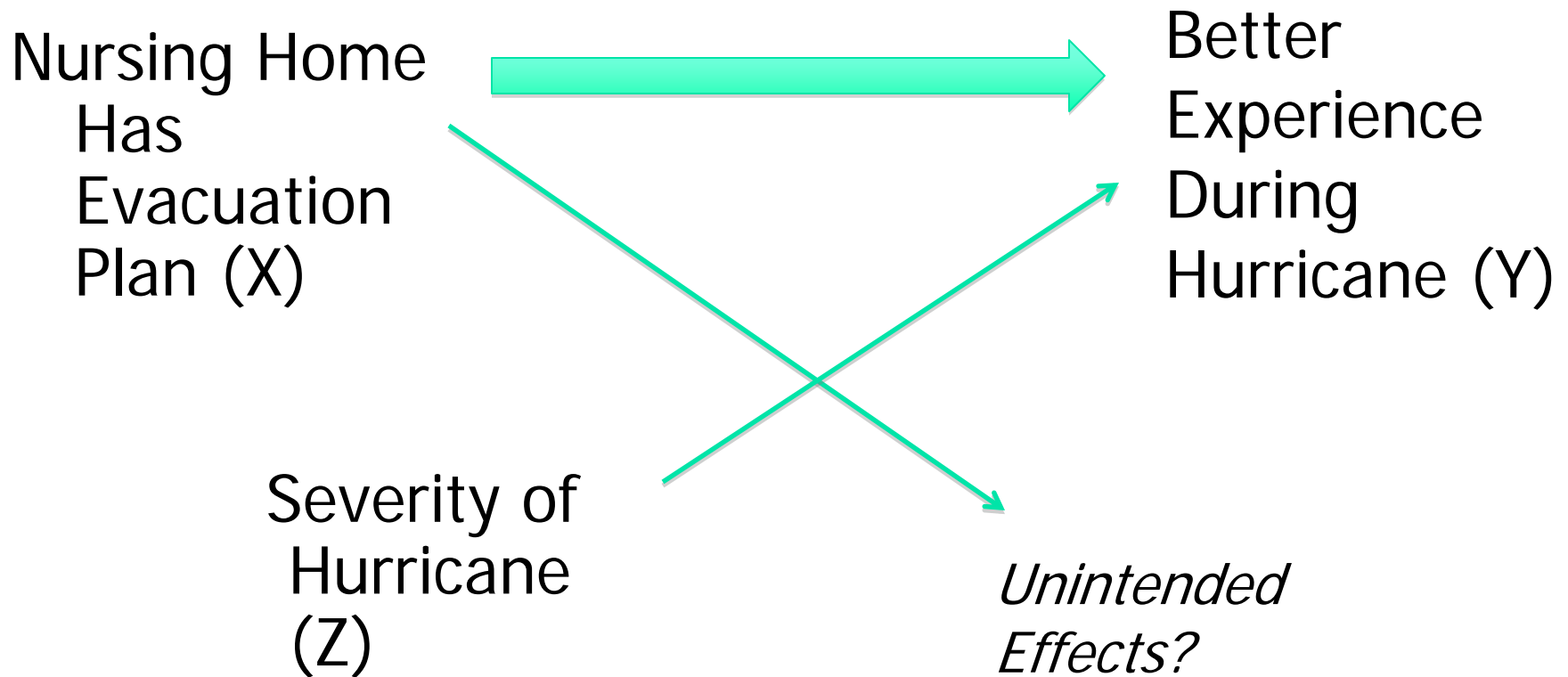
Internal Validity in Evaluations

- **Cause**: is typically what an evaluation focuses on as a factor that is contributing to deficiencies in the condition, and it will constitute the independent variable (X) in research terminology, **and** it may also be the recommended intervention to improve the condition.

Make Plausible Inferences with a Recommended Intervention

Nursing Home Has Evacuation Plan (X) ⟶ Better Experience During Hurricane (Y)

Severity of Hurricane (Z)

Unintended Effects?

Note: The Goal in Your Work is Typically Not to Establish Causation

- However, you need to critically examine the inferences you make to support your findings and recommendations!
- Are they plausible based on the data you have and do they seem logical?

External Validity

To generalize beyond the groups or context being studied, we need to:

- Identify the subgroups in the population of interest to us, so that we have large enough sub-samples of the groups of interest to analyze.
- However, even though probability sampling is used, we need to examine a sample to ensure that it is **truly representative** of the population to which we hope to generalize on demographic variables of interest to us, e.g., age, race?

Statistical Conclusion Validity

- Statistical conclusion validity is concerned with our ability to precisely estimate the magnitude of an attribute or the strength of a relationship based on sample values.

And in Inspections/Evaluation Work...

- We may undertake:
    - Universe analysis (say of all records),
    - Sample analysis without trying to project beyond the sample, or
    - Sample analysis with intent to project results beyond the sample.
- And it is only with the third instance that the use of inferential statistics is appropriate!

When We wish to Draw Inferences from Samples We Test for the "Statistical Significance " of our Sample Results

Statistical significance refers to the generalizability of numbers generated in a sample.

- It reflects the size of the sample, the dispersion of the sample data, and the magnitude of the effect of one variable on another.

- The statistic denoting statistical significance does not, by itself, convey how large the effect of the intervention is.

Statistical Significance versus Importance of the Measures?

When reporting results we must report both:

1. Whether or not the sample size allows us to conclude that a statistic can be generalized from the sample to the population, AND

2. The importance and relevance of the magnitude of the point estimate or relationship conveyed by the statistic.
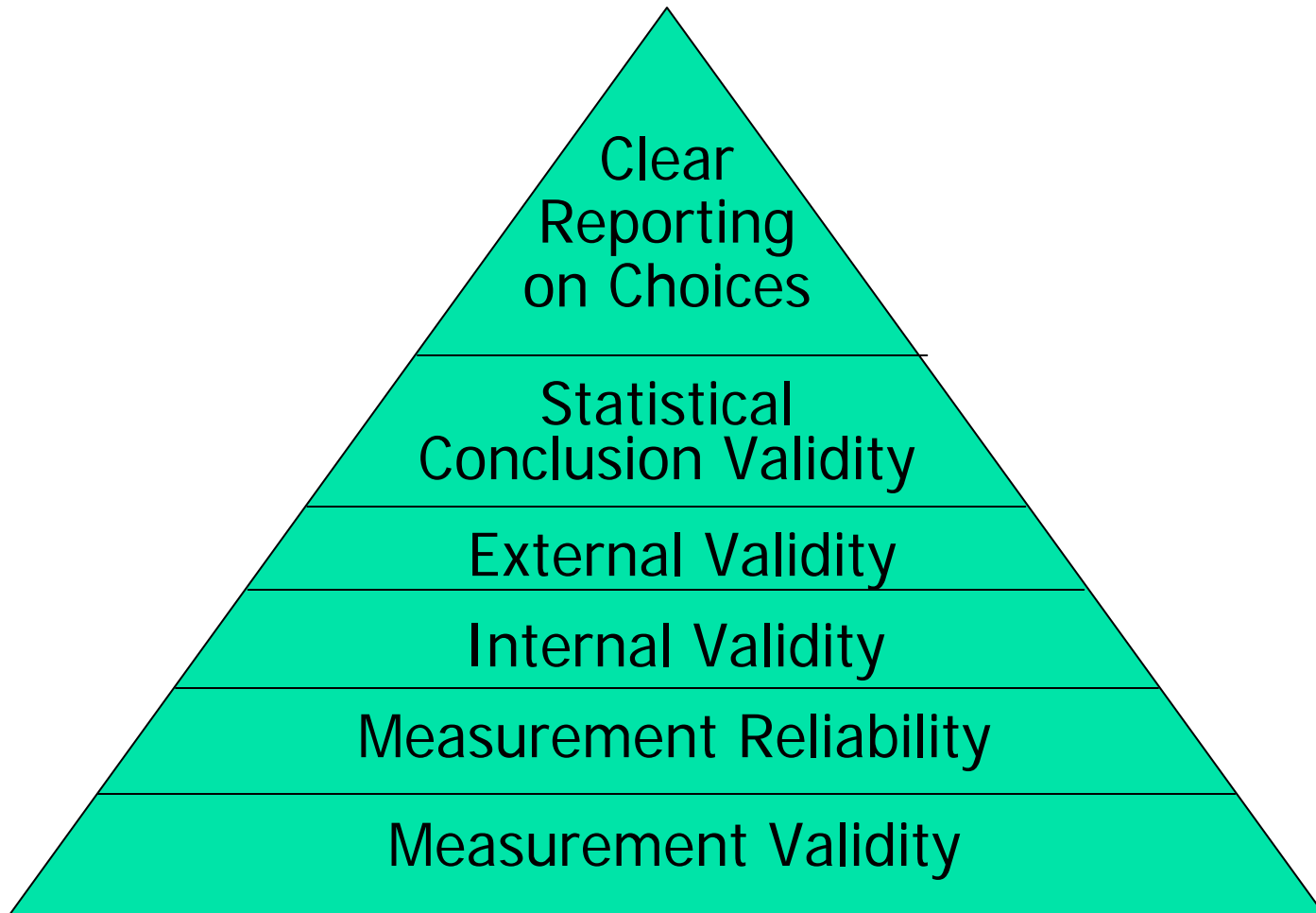
And When Presenting Both Quantitative Data and Qualitative Data – Think Comparison!

- Use graphs and tables effectively to reveal:
    - Relevant relationships
    - Pertinent trends
    - Notable differences
    - Surprising inconsistencies
- And label them clearly – as they need to stand alone!

Clear Reporting on Choices is the Top of the **Pyramid of Strength**!

Pyramid levels from top to bottom:
- Clear Reporting on Choices
- Statistical Conclusion Validity
- External Validity
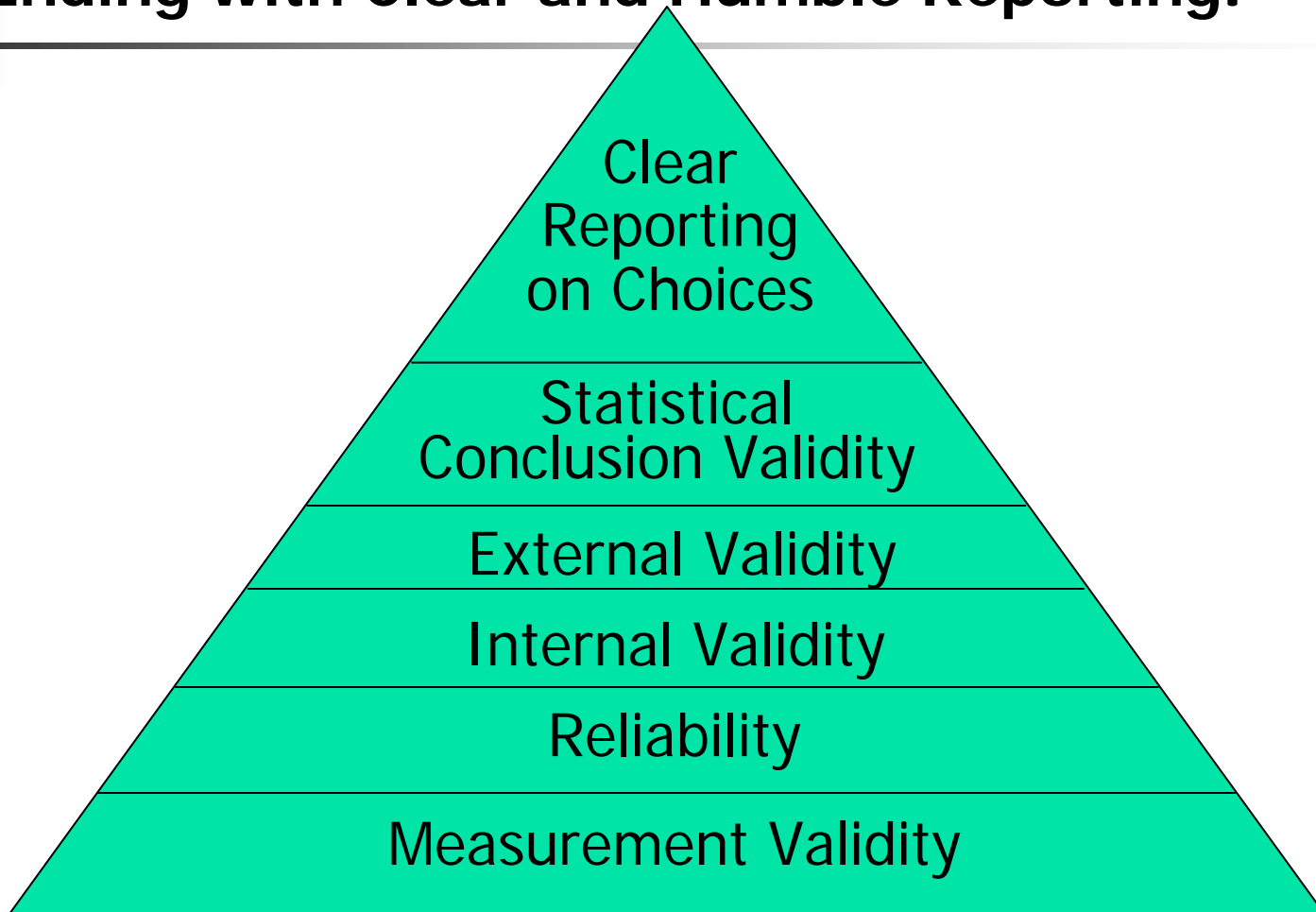- Internal Validity
- Measurement Reliability
- Measurement Validity

What Issues Bear Careful Reporting to Ensure Correct Interpretation?

- Limits to causal inferences (rival explanations and the time dimension)
- Sampling issues such as response rates, incomplete data concerns
- Limits to generalizability
- Null Results (versus negative results)

Thus, The Credibility of Inspection Findings will be built from the Ground up, starting with Wise Decisions Made During Design Deliberations and Ending with Clear and Humble Reporting!

Clear
Reporting
on Choices

Statistical
Conclusion Validity

External Validity

Internal Validity

Reliability

Measurement Validity

Message #2

Take Great Care With Measurement,
AND Be Humble In Reporting About It!

So, to Wrap Up…

- **Professional judgment** is an evaluator's most precious resource
- **Plan ahead** while designing evaluations and **anticipate challenges**
- No matter how the evaluation goes, **never sacrifice your methodological integrity**
- **Build a strong pyramid** and the impactful findings and recommendations will follow
- Even if your methodology is sound, nobody will know unless you **convey competence of the methodology in reports**

Table Discussions

- What are some successful practices you have employed during the design phase to identify appropriate questions to address?

- What is one particularly satisfying experience you have had in getting "really good" evidence?