DESIGNING AND IMPLEMENTING PROGRAM EVALUATIONS FOR ACCOUNTABILITY COURT PROGRAMS

Orion Mowbray, PhD

Michael Robinson, PhD

University of Georgia

School of Social Work

September 14, 2020

SESSION OVERVIEW

I. Purposes and foundations of evaluations

II. Creating evaluation designs - logic models and evaluation approaches

III. Assessing for appropriate evaluation costs

IV. Appropriate measurement selection associated with desired evaluation outcomes

V. What to look for in assessing prior evaluation research

I. PURPOSES AND FOUNDATIONS OF EVALUATIONS

- Program evaluation is applied research used as part of the managerial process
  - Conducted to aid those who must make administrative decisions about human services programs

- Program evaluation follows a logical, orderly sequence of investigation

- Usually involves making comparisons within or between groups

I. PURPOSES AND FOUNDATIONS OF EVALUATIONS

- Why evaluate programs scientifically?

- Because anecdotal or "case studies" don't produce data at a program level.
  - They can never determine program effectiveness

- Perspective and roles as practitioners often involves subjective opinions that may not represent the overall experiences of program participants.

- To have evidence, the conclusions we draw must be based on factual, verifiable evidence, and <u>not on opinion.</u>

I. PURPOSES AND FOUNDATIONS OF EVALUATIONS

- Why evaluate programs scientifically?

- Objectivity demands precision

- An **operational definition** is the way a variable or concept is to be defined and measured for purposes of the evaluation
  - As key concepts of a program are defined, vagueness disappears.

- Precision does not rule out the subjective experience in program evaluation.
  - Many great evaluators are also practitioners. The trick is to separate the roles

II. CREATING EVALUATION DESIGNS

- To create an **operational definition of what is being evaluated,** key concepts of a program must be defined

- The best way to identify key program concepts are through a logic model

- Logic model components:

Resources	Activities	Outputs	Short-term Outcomes	Impact (Long-term)
In order to accomplish our set of activities we will need the following:	In order to accomplish our objectives for the program we need to do these activities:	We can monitor our activities by counting or recording these events or products:	We expect our program of activities will lead to these changes:	We expect that the program will eventually lead to these changes:

- Generic mental health court logic model:

Inputs	Activities	Outputs	Outcomes		
			Short	Medium	Long
Crime Reduction Act	Program screening, assessment & intake	# clients screened for eligibility # clients enrolled # clients with current LSI-R scores	Divert at least 25% of caseload from prison	Reduce prison overcrowding	Greater public safety
Federal & state funding			Improve treatment outcomes	Reduce reliance on incarceration for non-violent offenders	Improved public safety
ARI Oversight Board	Psychiatric diagnoses Clinical assessment	# clients with DSM Axis diagnoses	Fewer arrests, jail stays, reconvictions	Lower criminal justice system cost to tax payers	Stronger social service safety net for people with SUD
ARI Staff					
Local ARI Program Site Staff	Team case planning	# clients with requirements and conditions	Improved restitution payments	Reduce recidivism, crime, and victimization	More effective use of tax dollars
Local court professionals	Linkage to therapeutic services	# clients with service referrals # types of service referrals # clients obtaining therapeutic services # types of therapeutic services	Better data for decision-making Community supervision for non-violent offenders with psychiatric diagnosis	Budgeting for results Client completes probation	Client improved quality of life and stability – psychiatric, social, and financial
Mental health & SUD treatment providers					
Community restorative boards (CRBs)	Community supervision, incentives & sanctions	#In-person visits with clients # Sanctions & Incentives # Status changes # Technical violations	Improved assessment outcomes Fewer arrests Fewer revocations and PTRs	Improve psychiatric outcomes/medication management Meet service needs	Reduce incidence of inmates with psychiatric diagnosis from ARI MHC sites
Target Population: Non-violent, high risk, persons with a psychiatric diagnosis	Program monitoring and evaluation	# program graduates # revocations # arrests during the program # sent to IDOC % reduction goal attained	Treatment retention Program graduations CRB resolutions	Demonstrate program effectiveness using program evaluations	
ICJIA staff					

II. CREATING EVALUATION DESIGNS

- Types of evaluation designs

- All evaluation research centers on causality
  - Does the program/practice/intervention cause a change among people?

- Two ways causality is examined
  - Pre-experimental (quasi-) design studies
  - Experimental design studies

- Pre-experimental design studies
  - Rank low on the evidence base ladder

1. One shot case study
   - X $\rightarrow$ O
   - X is the stimulus, or intervention and O is the observation
2. One group pre-test/post-test design
   - $O_1 \rightarrow X \rightarrow O_2$
3. Post-test only design with nonequivalent groups

   X   O

   O

- Experimental designs
  - Most valid, best evidence, hardest to make work

- Traditional: Pre-test/post-test control group design

- Random assignment to treatment and control, pre-test and post-test

- Depicted as:

$$R \quad O_1 \quad X \quad O_2$$
$$R \quad O_1 \quad \quad O_2$$

III. APPROPRIATE EVALUATION COSTS

- Program evaluation is a critical component of program delivery and requires investment to produce usable results

- Experienced evaluators should always offer a range of options for evaluation costs

III. APPROPRIATE EVALUATION COSTS

- Many SAMHSA grants state that no more than 10% of a project budget should be allotted to data collection (evaluation) activities

- Ideas for saving costs:
  - Collect data in house and make available for evaluators
  - Build evaluation activities and costs into existing budgets and grants
  - Colleges and universities often have multiple staff available to contribute to project
  - Consult with other organizations and partners about the right evaluation for the project prior to hiring an evaluator
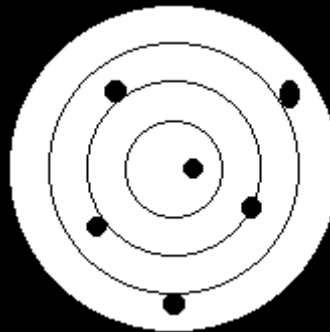
IV. APPROPRIATE MEASUREMENT SELECTION

- Two main topics in measurement: Reliability and validity

  - Reliability: The accuracy of a measure whether items within a measure are related in some meaningful way

  - Validity: The extent to which any instrument measures precisely what it intends to measure
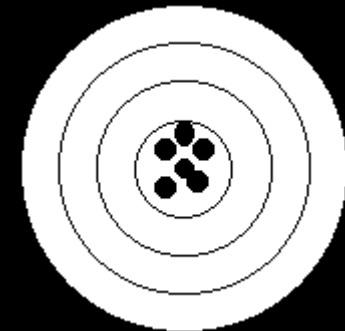
- Reliability:
  - A measurement technique to assess overall agreement across respondents
  - The more reliable the measure, the less <u>random</u> error

Reliable but
not valid

Neither reliable
nor valid

Valid and
reliable

- Quick overview of validity:

- Face validity
  - A measure appears to measure what it is supposed to measure

- Content validity
  - The degree to which a measure covers the range of meanings included within the concept

IV. APPROPRIATE MEASUREMENT SELECTION

- ## Criterion-related validity
  - Usually Predictive Validity – weather a measure can predict a something that will occur in the future

- ## Construct validity
 - Convergent/Discriminant validity – whether a measure agrees/disagrees with other measures of same topic

V. ASSESSING PRIOR EVALUATION RESEARCH: AN EXAMPLE

- A Springfield, IL. Trial Court receives a $2 million grant to provide mental and behavioral health services within the Municipal Court and Drug Court.

- Award will provide care for 100 persons

- Funds will provide case management, co-occurring disorders treatment, peer support, vocational supports, and trauma-informed care into a single, coordinated service delivery approach

Springfield Drug Court awarded $2-million grant for substance abuse and mental health programs
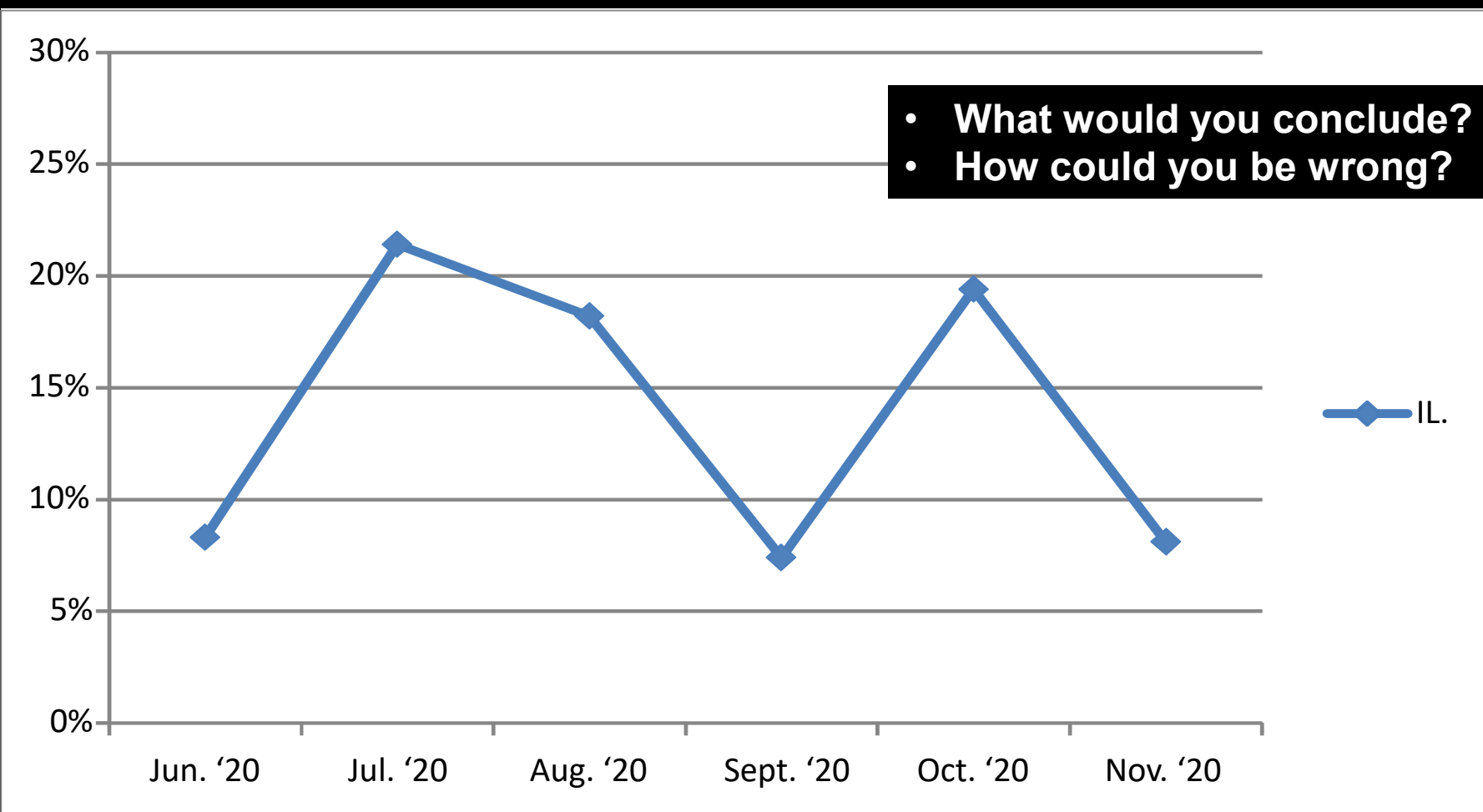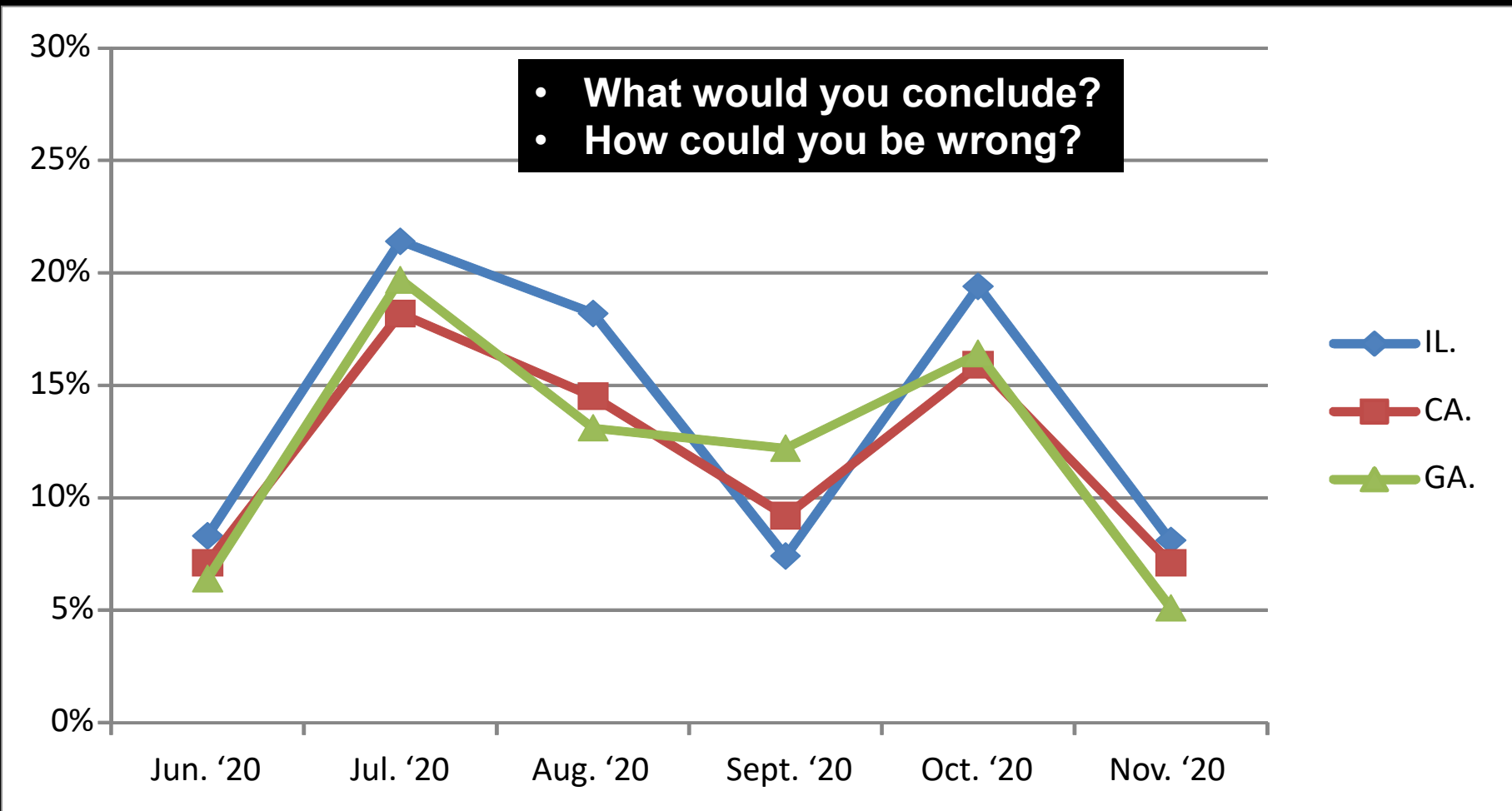
V. ASSESSING PRIOR EVALUATION RESEARCH: AN EXAMPLE

- **What would you conclude?**
- **How could you be wrong?**
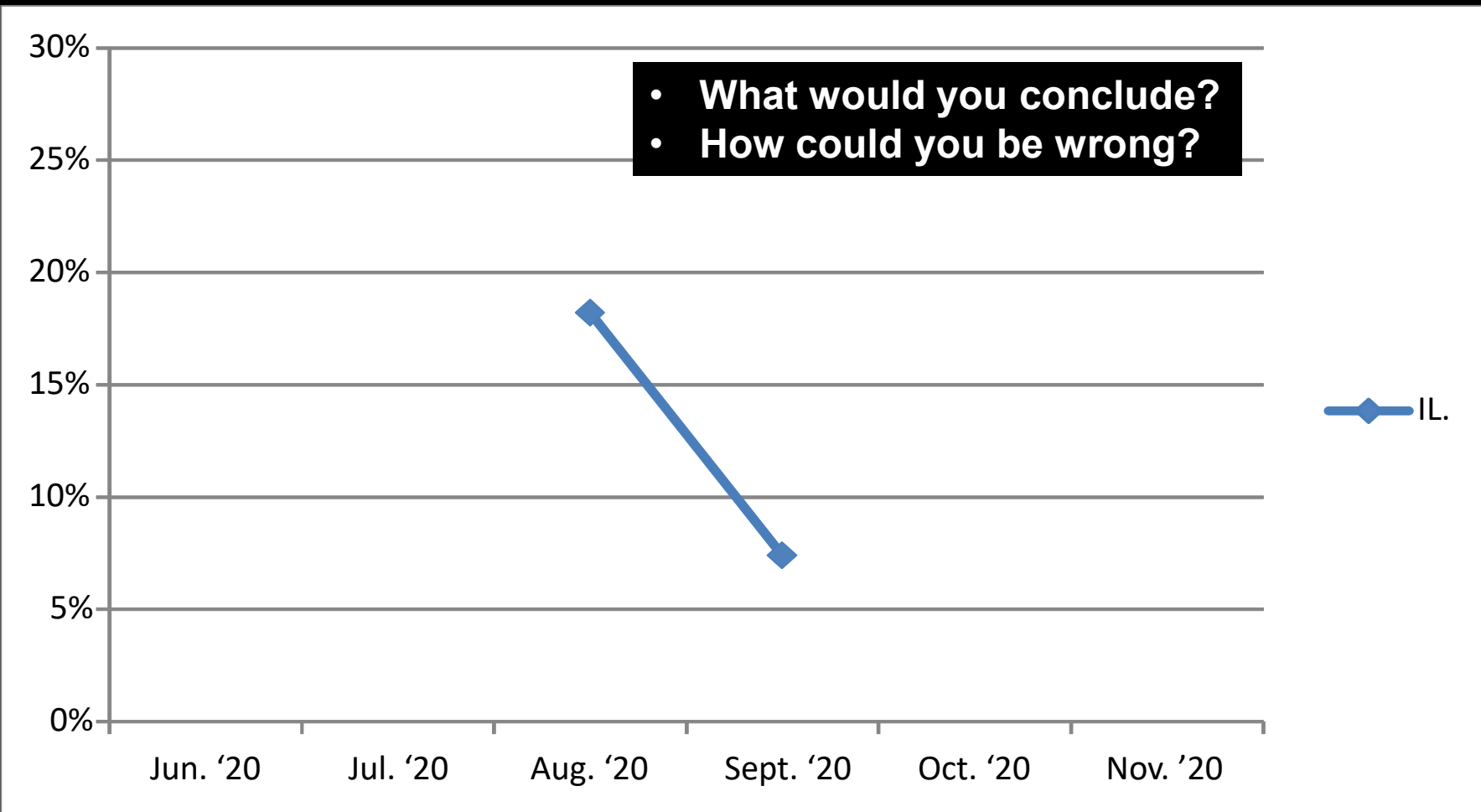
Note: Results presented are not real program results. This is an illustrative example

V. ASSESSING PRIOR EVALUATION RESEARCH: COMMON PROBLEMS

1) Unclear why study was done

2) Sample size issues

3) Method doesn't reveal what was actually done

4) Rationale for statistical test is not described

5) Limitations aren't fully discussed

6) Inferences on study results too big

V. ASSESSING PRIOR EVALUATION RESEARCH

- **Issue 1: Why was study done in the first place?**

- This SHOULD be found in the introduction
  - Introductions outline study rationale.

- Comes in an explicit statement in the form of:
  - Statement of purpose, Objective, Research question, or Hypotheses

- What makes a good introduction?
  - Offers point of the study
  - Justifies why this study is done
  - Adequate summary of literature

V. ASSESSING PRIOR EVALUATION RESEARCH

- **Issue 2: Sample size**

- What are the consequences of too large/small sample size?

- Type 1 error: False positive - a difference in the sample, not in the pop.
  - Can occur in large sample

- Type 2 error: False negative - there's no difference in the sample, there is in the pop.
  - Can occur in small samples

V. ASSESSING PRIOR EVALUATION RESEARCH

- **Issue 3: Method doesn't reveal what was actually done**

- What to look out for/what's not reported
  - Collapsed variables
  - A presentation of how the **dependent variable** is distributed

- Level of measurement
  - Nominal, Ordinal, Interval or Ratio

- Level of measurement will determine appropriate statistical test

V. ASSESSING PRIOR EVALUATION RESEARCH

- **Issue 4: Rationale for statistical test is not described**

- What's the statistical analysis?
  - Inferential vs. descriptive

- Which type of test: **Bivariate vs. Multivariate?**
  - Were other factors accounted for or not?

- Statistical tests go beyond p values and significance
  - Issues with effect sizes, the power of the test can influence test results

V. ASSESSING PRIOR EVALUATION RESEARCH

- **Issue 5: Limitations aren't fully discussed**

- Limitations section follows a summary of findings and is incorporated in the discussion section.

  - Simple space does not allow full discussion of limitations
  - They should address what's in our list here, but sometimes they do not
  - When criticizing a study, look to see if the criticism is mentioned in the limitations.
  - **No study is perfect; limitations will always emerge**

- **Issue 6: Inferences on study results are too bog**

- In discussion/conclusion section of a study, the author is expected to
  - Summarize findings
  - Offer implications for practice

- Sometimes the implications go way too far or attempt to make findings generalizable when they are not

CLOSING

- Rigorous evaluations are hard to do
  - They require up front planning, using a logic model and team agreement on what should be evaluated and how

- There are many out there willing to help do evaluation work
  - But a reasonable evaluation budget and understanding of what's needed is necessary to make partnerships work

- Appropriately reliable and valid measures, especially when doing in-house data collection are critical in the evaluation process

- Assessing the utility of prior evaluation research requires a discerning eye and an understanding of how research is presented/published

THANK YOU

Orion Mowbray, PhD

University of Georgia

School of Social Work

Athens, GA 30677

omowbray@uga..edu

734-260-4730

Michael Robinson, PhD

University of Georgia

School of Social Work

Athens, GA 30677

marobi01@uga.edu

205-219-2043