Welcome

to the

IAMSE Webcast Audio Seminar

Please Observe the Following

- Do NOT at any time put this call on "hold"
- Be conscious of your background noise
- Speaker phones use mute button or *6
- Problems? Contact julie@iamse.org

Decisions, Designs, and Data

Doing Quantitative Research in Medical Education

14 October 2004

Larry D. Gruppen, Ph.D.

Overview

1. **Hypotheses**
2. **Study design**
3. **Measurement**
4. **Sampling**
5. **Data analysis**
6. **Interpreting the results**
7. **Resources**

3

Illustrative Example

- Basic science lectures videotaped
  - Originally for viewing by absent students
- Technology afforded web-based distribution, random access, and variable replay speeds
- Student demand for broadened (universal) access
- Faculty concern about impact on learning (specifically, lecture attendance)

4

1. Hypotheses

- **What is the purpose of the study?**
  - **Why are you doing this?**
  - **What do you hope to achieve?**
  - **What are the questions you want to answer?**

5

1. Hypotheses: Videotaped Lectures

- **Possible hypotheses**
  - **Students who attend lecture will learn more than students who view the streaming video**
  - **Students who use the streaming video will be more efficient and use the spare time to study more deeply and broadly**
  - **Streaming video will be a more cost-effective method for delivering didactic teaching than traditional lectures**

6

2. Study design

- **Study design consists of:**
  - **What you are measuring**
  - **The participants and how they are assigned**
  - **The intervention**
  - **The sequence and timing of measurements and interventions**

7

2. Study Design

- **All studies have a design. The question is "How good is it?"**
- **All designs are compromises**
  - **Money**
  - **Time**
  - **Ethics**
  - **Good-will**

8

2. Study Design: Comparison Group

- Pre-post design - compare intervention group to itself
- Non-equivalent control group design - compare intervention group to an existing group
- Randomized control group design - compare to equivalent controls

9

3. Measurement

- Three basic questions
  - What do you measure?
  - How do you measure?
  - How well do you measure?

10

3. Measurement: What do you measure?

- Outcomes need to be aligned with the hypotheses and purposes of the study
- Covariates: factors that co-vary with the responses
- Multiple variables for underlying factor, dimension, or construct
- Any feature that is relevant to the environment

11

3. Measurement: How do you measure?

- Operationalization is essential to the conduct of the study

- Counting events
- Measuring time and physical quantities
- Externalizing internal (psychological) states, events, and processes

12

3. Measurement: Educational Outcomes

- Knowledge
  - Tests (MCQ, essay, oral)
- Attitudes
  - Questionnaires, surveys
- Behavior or performance
  - OSCEs, standardized patients, direct observation

13

3. Measurement: How Well ...

- Reliability — score accuracy or stability
  - Would the score be reproduced if tested again?
  - Would the score be reproduced by different raters?
- Validity — score meaning
  - Does the score measure what you intend to measure?

14

3. Measurement: Assessing Reliability

- Test-retest
- Split-half test reliability - form equivalence
- Cohen's coefficient alpha - internal consistency
- Inter-rater reliability

15

3. Measurement: Assessing Validity

- Face - looks like it measures the right thing
- Content - content is representative of the larger content domain (often based on expert judgment)
- Criterion (predictive) - scores on the measure predict performance in 'real world'
- Construct - measure reflects underlying theoretical construct (e.g., professionalism)

16

4. Sampling

- Size: larger samples enable detection of smaller effects or relationships
- Population: sample must represent the target population
- Random selection: avoid biases in sample

17

4. Sampling: Biases

- The study focus (sample) is not the same as the study target (population)
- Respondents/volunteers not the same as non-respondents/non-volunteers
- Sampling from top and bottom of distribution (regression to mean)

18

5. Data Analysis

- Two basic statistical concepts
  - Central tendency: mean, median, mode
  - Variation: variance, standard deviation, range
- Number of dependent and independent variables
- Scale of measurement
  - Nominal
  - Ordinal
  - Continuous

19

6. Interpreting the results

- Presuming the study had a clear hypothesis, was well-designed, well-executed, appropriately analyzed…
- Two additional issues in reporting
  - Effect size and confidence intervals vs. statistical significance ($p<.05$)
  - Cost-effectiveness

20

6. What p<.05 *IS*

- The probability of getting these results, given the null hypothesis is true
- A dichotomous judgment (not 'highly' significant or 'marginally' significant)
  - Set your p criterion *a priori* and let that determine your decision about the data

21

6. Interpretation: What p<.05 is *NOT*

- NOT the probability that the null hypothesis is true, given these results (NOT that a smaller p means a less likely null)
- NOT the probability of getting the same result after replication
- NOT an indication of 'importance'
- NOT a basis for comparing studies or results, i.e., p<.0001 not 'better' than p<.05

22

6. Interpretation: Effect Size & C.I.

- Instead of p<.05, focus on effect size measures and confidence intervals
  - ES tells you the magnitude of the effect of the intervention or of the observed relationship
  - CI tells you the precision of the results

23

6. Interpretation: Effect Size

- How large is the effect?
  - Dozens described in the literature
    - Standardized mean difference (d) for t-tests
    - R^2 for multiple regression
    - η^2 for analysis of variance
  - Unit free measures

24

6. Interpretation: Interpreting Effect Sizes

- Rules of thumb for “small,” “medium” and “large” effects
- Many can be converted into “proportion of variance” values
- Most educational effects are “small” to “medium”
- Little experience in interpreting them (more complex than statistical significance)

25

6. Interpretation: Confidence Intervals

- Tell you everything p<.05 does & much more
- e.g., mean diff. = 1.2 (95% CI 0.3 to 2.1)
  - Statistically significant since CI excludes 0
  - Tells you the range of possible values
  - CI is symmetrical, NHST is asymmetrical
- Can compare CIs among studies
  - Mean diff = 2.0 (95% CI -1.0 to 5.0)
  - Mean diff = 2.0 (95% CI 1.5 to 2.5)

26

6. Interpretation: Cost-effectiveness

- Becoming more common in clinical research
- Little used in education
  - Difficult to account for costs
  - Historical neglect of *size* of effects
- Essential for pragmatic use of educational research - how much bang for the buck?

27

7. Resources

- Selecting statistical procedures
  - http://www.vaneck.us/microsiris/Statistical%20Decision%20Tree/
  - http://www.graphpad.com/www/Book/Choose.htm
- Other statistics resources
  - Web center for social research methods
    - http://www.socialresearchmethods.net/
  - StatPages.net
    - http://members.aol.com/johnp71/javastat.html

28

7. Design & analysis

- Pedhazur E, Schmelkin LP. Measurement, design, and analysis: an integrated approach. Hillsdale, NJ: Erlbaum, 1991.
- Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Elrlbaum, 1988.
- Kirk RE. Experimental design: procedures for teh bahavioral sciences, 2nd ed. Pacific Grove, CA: Brroks/Cole, 1982.

29

7. Design & analysis

- Myers JL, Well AD. Research design and statistical analysis. Hillsdale, NJ: Erlbaum, 1995.
- Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston, MA: Houghton Mifflin, 1979. (the classic)

30

7. Reliability and Validity

- http://seamonkey.ed.asu.edu/~alex/teaching/assessment/reliability.html
- http://www.socialresearchmethods.net/kb/introval.htm
- Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills: Sage; 1979.
- Feldt LS, Brennan RL. Reliability. In: Linn RL, ed. Educational measurement. 3rd ed. ed. New York: Macmillan; 1989.

31

7. Effect Sizes & Confidence Intervals

- Smithson M. Confidence Intervals. Vol 07-140. Thousand Oaks, CA: Sage Publications, Inc.; 2003.
- Colliver JA. Call for greater emphasis on effect-size measures in published articles in Teaching and Learning in Medicine. Teaching and Learning in Medicine. 2002;14(4):206-210.

32

7. Null Hypothesis Statistical Testing

- Kirk RE. Practical significance: A concept whose time has come. Educational and Psychological Measurement. 1996;56:746-759.
- Shaver JP. What statistical significance testing is, and what it is not. Journal of Experimental Education. 1993;61:293-316.
- Carver RP. The case against statistical significance testing, revisited. Journal of Experimental Education. 1993;61:287-292.

33

7. Null Hypothesis Statistical Testing

- Lingren BR, Wielinski CL, Finkelstein SM, Warwick WJ. Contrasting clinical and statistical significance within the research setting. Pediatric Pulmonology. 1993;16:336-340.
- Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. Psychological Methods. 2000;5(2):241-301.

34