

Program Evaluation

Jonathan Howland, PhD, MPH

Spring 2005 SB721 Class 12AB 1

Levels of Evaluation

Level	Question	Example of Method
Formative	Does the program make sense?	Focus groups with target population.
Process	Was the program implemented as intended?	Measure participant compliance with program regimen.
Outcome	Was the program effective?	Randomized trial or quasi-experiment

Examples of Formative Evaluation

- Focus groups
- Field observation
- Interviews with program stakeholders (e.g. health care providers, patients)
- Public forums (e.g. community meetings)

3

Examples of Process Evaluations

- Participant attendance/compliance records
- Direct observation of program implementation
- Participant satisfaction questionnaires
- Log book of critical implementation events
- Scanning media for current events that might affect program implementation of outcomes
- Interviews with program staff

4

Reasons for Conducting a Process Evaluation

- Accepting the null hypothesis (program did not result in hypothesized change)
- Replication
- Program management
- Accountability

5

Types of Causal Study

Study	Hypothesis	Causal logic	Examples
Observational	Does exposure X cause health problem Y?	Coincidence of X and Y in time and space	<ul style="list-style-type: none"> ▪ Correlation ▪ Case-control ▪ Cohort
Experiment	Does program X improve health outcome Y?	Wiggle X to observe change in Y	<ul style="list-style-type: none"> ▪ RCT ▪ Cluster RCT ▪ Quasi-experiment

6

Scientific Method for Establishing Causality

X → causes → Y,

if no reasonable alternative cause (c) can be identified

Example: The diet program was effective because no other explanation for participant weight loss could be identified. ,

Causality in Jurisprudence:

Prosecution: Pete killed Stan because no one else could have done it.

Defense: Joe might have killed Stan because Joe had a motive and an opportunity.

Trial: Court Trial

Method: Jury of Pete's peers.

Causality in Program Evaluation:

Hypothesis 1: The new diet program caused weight reduction because no other factors influenced participants.

Alternative Explanation: Participants were all infected with gastro-intestinal disease.

Trial: Randomized Control Trial (RCT)

Method: Peer review of paper reporting of study results.

Types of Validity

Type	Question
Internal	Is the causal relationship between the program and the outcome real or spurious (confounded)?
External	If the program worked in a randomized trial using recruited volunteers, will it also work with another population?

- ### Threats to Internal Validity
- Testing
 - History
 - Instrumentation
 - Maturation
 - Differential Attrition
 - Statistical Regression
 - Selection
 - Contamination
 - Hawthorn Effect
 - Chance

- ### Testing
- The pretest measurement by itself has an impact on the posttest measurement
 - Practice effects
 - Self-reflection
 - Awareness of hypothesis
 - Social desirability
 - Reactivity

History

- There may be unplanned events that occur between the first and second measurements
 - Programs
 - Policies
 - Events
- The more time that elapses between measurements, the greater the threat from historical effects

13

Instrumentation

- The measurement instrument, or the criteria for recording behavior, change between the pretest and the posttest
 - Context of survey questions
 - Interviewer skills
 - Archival records

14

Maturation

- Every person changes over time
 - Development
 - Aging
 - Role transition
 - Seasonal effects

15

Attrition

- There might be unequal dropout rates among the experimental and comparison groups
- Those dropping out from one group may be different from those dropping out of another group

16

Statistical Regression

- If a study sample was chosen due to extreme behavior, then over time that behavior will tend to regress toward the overall group average
 - For anyone with an extreme pretest score, the best prediction is that their posttest score will be less extreme

17

Selection

- There are unmeasured differences between an experimental group and a nonequivalent comparison group
 - Motivation
 - Family and social support structures
 - Or any number of other things

18

Contamination

- The control group was inadvertently exposed to the intervention program.

19

Hawthorne Effect

- The very fact of being in an evaluation can result in changed behavior
 - Similar to a placebo effect in medical research

20

Chance

- The outcome was random and had nothing to do with the intervention program.

21

		Confounder Drives	
		> Null Hypo	< Null Hypo
Results	Sig. Diff	OK	Confounded
	No Sig. Diff	Confounded	OK

22

One Group Pre-test Post-test Design

O X O

23

One Group Pre-test Post-test Design

Example:

Pre-test Mean *Post-test Mean*
 150 lbs X 120 lbs

Are there any other factors that may have caused this change from pre-test to post-test?

24

Nonequivalent Comparison Group Design (Quasi-Experiment)

E O X O
C O O

25

Nonequivalent Comparison Group Design (Quasi-Experiment)

Example:

	<i>Pre-test Mean</i>		<i>Post-test Mean</i>
E	150 lbs	X	120 lbs
<u>C</u>	143 lbs		147lbs

Are there any other factors that may have caused this change from pre-test to post-test?

26

Drawbacks to Nonequivalent Comparison Group Design

- Group assignment is not random
- The two groups can differ from one another in many ways, some of which might affect the changes in outcome measures
- Various pretest measures might show that the groups are very similar – but we can't measure everything

27

The True Experiment

- Controlled trial
- Random assignment among experimental and control groups.
- Protects against many threats to internal validity

E O X O
R C O O

28

The True Experiment

Example:

	<i>Pre-test Mean</i>		<i>Post-test Mean</i>
E	150 lbs	X	120 lbs
C	150 lbs		147lbs

Are there any other factors that may have caused this change from pre-test to post-test?

29

Almost Perfect...

- This design controls for almost all threats to validity
- Absence of selection effects should be verified
 - Did the randomization work?

30

Remaining Threats to Validity

- Differential attrition
- Treatment group exposure to other program
- Contamination
- Hawthorn effect
- Chance

31

True Experiments: Cluster Trials

- Cluster Randomized Trials or Group Randomized Trials (GRTs)
 - Trials designed to evaluate group interventions that assign intact groups (e.g., schools, workplaces) to experimental conditions
- Individually Randomized Group Treatment (IRGTSS) Trials
 - Trials that assign individuals to study conditions when interventions are delivered to groups (e.g., exercise groups)

Both designs involve potential correlation among observations within treatment groups that must be adjusted for in power calculations

32

Analysis of Randomized Trial

33

Analysis of Nonequivalent Groups (Quasi-experiment)

34

Types of Errors in Hypothesis Testing

Types of error in hypothesis testing		Reality	
		Difference	No difference
Study Finding	Difference ¹	OK	Type I
	No difference ²	Type II	OK

¹Reject null hypothesis ²Accept null hypothesis

35

EVALUATING AN INTERVENTION TO REDUCE NEW STIs

36

WHAT IS THE PUBLIC HEALTH PROBLEM?

- About 25% of STD clinic patients return with a new infection
- This group may constitute a core that serves as a reservoir of disease in the community
- STIs that involve lesions can enhance transmission of HIV

37

CAN ANYTHING BE DONE TO REDUCE THE RATE OF NEW INFECTIONS?

Several studies have shown that intensive counseling of clinic patients can reduce rates of new infections

Spring 2005 SB721 Class 12AB

38

IS THE STD CLINIC A GOOD PLACE TO INTERVENE FOR THIS PROBLEM?

Spring 2005 SB721 Class 12AB

39

WHAT SHOULD THE BEHAVIORAL OBJECTIVE BE FOR THE INTERVENTION?

Spring 2005 SB721 Class 12AB

40

WHAT ARE THE BARRIERS TO CONDOM USE IN THE TARGET POPULATION?

Spring 2005 SB721 Class 12AB

41

WHAT MEDIUM SHOULD BE USED FOR THE INTERVENTION?

Spring 2005 SB721 Class 12AB

42

STI INTERVENTION STUDY: VIDEO VIGNETTE TOPICS	
VIGNETTE #	VIGNETTE TOPIC
1	Shows how a woman can negotiate condom use with her primary partner
2	On the street interviews about how to make condom use more fun
3	A rap song with visuals showing the proper way to use a condom
4	Shows how to tell a partner, whom you may have infected, that they have to go to the STD clinic for an exam
5	Shows how to talk to a casual partner about using condoms
6	Shows how to respond to excuses for not wearing condoms
7	A monologue about not allowing alcohol or drugs to inhibit condom use
8	Shows how a man can talk to his primary partner about using condoms
9	An infomercial showing the different types of condoms available
10	A monologue about how the HIV virus is spread from person to person

WHAT IS THE BEST RESEARCH DESIGN TO USE FOR EVALUATING THIS INTERVENTION?

Spring 2005 SB721 Class 12AB 44

WHAT SHOULD THE OUTCOMES BE?

Spring 2005 SB721 Class 12AB 45

HYPOTHESIS THAT INTERACTIVE VIDEO IS MORE EFFECTIVE THAN LINEAR VIDEO?

Spring 2005 SB721 Class 12AB 46

WHAT SHOULD THE CONTROLS RECEIVE?

Spring 2005 SB721 Class 12AB 47

WHAT IS THE REAL RESEARCH QUESTION?

Spring 2005 SB721 Class 12AB 48

ROUTINE CARE CHECKLIST

ROUTINE CARE PATIENT EDUCATION PROTOCOL CHECK LIST AS MANDATED BY STATE CONTRACT

- Counseled on medical instructions
- Administration of medications
- Counseled on medication side effects
- Counseled on complications of health problem
- Patient verbalizes understanding of treatment instructions
- STI literature provided
- STI education provided
- Counseled on disease symptoms
- Counseled on disease transmission
- Counseled on disease prevention
- Counseled on disease symptoms
- Counseled on response to future STIs
- Need for follow-up visit discussed
- Asked to return to clinic for follow-up visit
- Partner referral
- HIV counseling
- HIV test offered
- HIV literature provided
- Condoms provided

49

STI INTERVENTION STUDY DESIGN

Assignment	Groups	Baseline	Intervention	Follow-Up		
				6 wk	6 mo	1 yr
R	E1 N=500	O ₁	X Interactive Video	O ₂	O ₃	O ₄
	E2 N=500	O ₁	X Linear Video	O ₂	O ₃	O ₄
	C N=500	O ₁	Routine Care	O ₂	O ₃	O ₄

50

THREATS TO VALIDITY?

TESTING	
HISTORY	
INSTRUMENTATION	
SELECTION BIAS	
MATURATION	
ATTRITION	
REGRESSION	
CONTAMINATION	
HAWTHORN EFFECT	

51

WHAT WOULD YOU MEASURE FOR PROCESS EVALUATION?

52

PROCESS EVALUATION

➤ Direct observation of study implementation
➤ Review routine care checklists
➤ Participant satisfaction interviews
➤ Event log book
➤ News media scans
➤ Enrollment/refusal rates
➤ Follow-up rates
➤ Monthly reports

53

WHAT WERE THE STUDY RESULTS?

54

POSITIVE AT ENROLLMENT – MALES AND FEMALES 6 WEEKS				
GROUP	POSITIVE	NEGATIVE	TOTAL	% POSITIVE
TREATMENT	54	134	188	29%
CONTROL	46	87	133	35%
TOTAL	100	221	321	31%
OR = 1.3, p= 0.26				

55

POSITIVE AT ENROLLMENT – MALES 6 WEEKS				
GROUP	POSITIVE	NEGATIVE	TOTAL	% POSITIVE
TREATMENT	48	89	137	35%
CONTROL	38	55	93	41%
TOTAL	86	144	230	37%
OR = 1.3, p= 0.37				

56

POSITIVE AT ENROLLMENT – FEMALES 6 WEEKS				
GROUP	POSITIVE	NEGATIVE	TOTAL	% POSITIVE
TREATMENT	6	45	51	12%
CONTROL	8	32	40	20%
TOTAL	14	77	91	15%
OR = 1.9, p= 0.28				

57

CONCLUSIONS

- Could have been a type 2 error
- Should have done better formative research
- An intervention with a small effect (6 weeks) to which many are exposed because it is cheap could yield a big effect for the community

58