

Chapter 12. Evaluation Research

- Evaluation research (program evaluation)'s uniqueness lies in its purpose: to evaluate the impact of social intervention.
- Topics covered in this chapter
 - Topics appropriate for evaluation research
 - Issues related to conducting evaluation research
 - Formulating the Problem: Issues of Measurement
 - Types of Evaluation Research Designs
 - The Social Context
 - Common Problems associated with evaluation research
 - Social Indicators Research

What are the topics appropriate for evaluation research?

- What is a social intervention?
 - An action taken within a social context for the purpose of producing some intended result
- What topics are appropriate for evaluation research?
 - Occurring social intervention
 - Planned social intervention

What are the procedure for evaluation research?

- Formulating the problem
 - Specify the intervention
 - Specify the intended result of such intervention
 - Specify the population
- Measurement issues
 - Measure the outcome (response variable)
 - Measure the context and other related variables - aspects of the context of an experiment that might affect the experiment.
- Mode of observation
 - Classical experimental design
 - Quasi-experimental design
- Evaluating success/failure
 - Cost-benefit analysis
 - Agreement

How to formulate the problem?

- Specify the intervention
 - Ex. DARE (Drug Abuse Resistance Education) program: A drug prevention program that uses uniformed police officers in the schools to warn students away from drugs.
- Specify the intended results of the intervention
 - More than one intended result possible for a specific project
 - Ex. Reduce drug use among school-aged children
- Specifying the population
 - Population affected by the social intervention
 - Ex. Middle school /High school children

What are the issues related to measurements?

- Specify outcomes to be measured
 - Can employ more than one measurement
 - Ex. rate of tobacco use, rate of marijuana use, rate of cocaine use, rate of ecstasy use.
- Measure study contexts and other variables
 - The context within which the study occurs is critical. Hence variables related to the context must be identified and measured as well.
 - Ex. If in the middle of your evaluation research there is a national TV campaign against drug use, then this campaign should be taken into consideration, especially if the mode of observation is time-series (see later slides)
 - Other variables that may affect the outcome
 - Ex. Family background of the school-age children; Local culture.

What are the common modes of observation?

- Classical experiment design
 - The ultimate method to determine causal relationships
- Quasi-experimental designs
 - Time-series design - studying the processes occurring over time. Such analysis is useful when one can only have an experimental group without a control group
 - Nonequivalent control groups: Nonequivalent control group designs can be employed when it is possible to find an alternative control group similar to the experimental group, though this control group is not randomly selected.
 - Multiple Time-Series designs
- Qualitative evaluations
 - Field interviews, unstructured questions

An example of classical experimental design for evaluation research

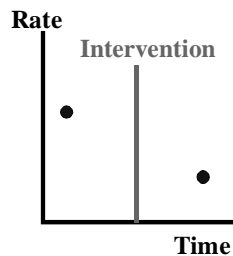
- Suppose we want to evaluate whether DARE is successful in Salt Lake County
- Randomly select a sample of high schools
- Randomly assign half of the schools in the experimental group, the other half in the control group. Pretest the kids in these two groups to make sure they are similar.
- The experimental group will get the DARE program. The control group won't.
- After a certain period of time (maybe a year), measure the usage rate of different drugs and tobacco.
- If the experimental group has significantly lower rate of usage, then the program is successful.

An example of time-series method

- DARE as an example again
- If DARE is implemented in all schools in Salt Lake County, then experimental design is not feasible
- Researchers can collect data on drug use among high school students over time, both before and after DARE was implemented.
- If there is a substantial decrease in the rate of drug use, then the program is considered successful.
- Note in such a design, contextual variables are especially important. Suppose in the midst of your research a national anti-drug use TV campaign started. You may find a lower rate of drug use at the end of your research period. But it could be caused by the TV campaign, not the DARE program.

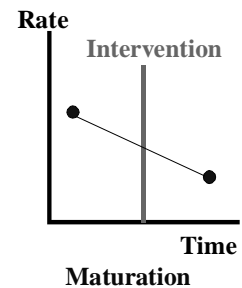
Time-series design: Patterns of rate of drug use in a longer historical period

- Suppose we observe a reduction in rate of drug use after the DARE program is utilized (the Intervention line shows when the program started). We observe the rate of drug use before is higher than the rate of drug use after the Intervention. But can we conclude that the DARE program is successful?



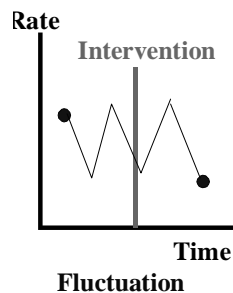
Time series - Maturation

- There are several circumstances under which we might observe these the graph on the previous change.
 - Maturation – meaning that it is possible that over time, the drug use rate will naturally decrease for the same group of students simply because they get more mature and know more. If that is the case, then if more data points are collected, the pattern on drug use rate will show as a decreasing line (see graph). The Intervention, in this case, does not have any effect.



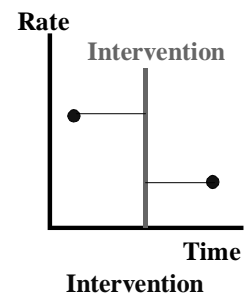
Time series - Fluctuation

- Fluctuation – This is shown on the graph to the right. In this case, the rate of drug use fluctuate over time naturally, may be due to seasonally reasons. If one only collects two data points (the blue ones), one would think the DARE program is effective. But if one collects more data points that shows a fluctuation pattern, then one would know the Intervention is not effective as the two points just happen to be at the high and low points on the fluctuation curve.



Time series - Intervention

- Intervention – This is the pattern where there is evidence that the Intervention is effective. With many data points showing a higher drug use rate before the Intervention and many data points showing a lower drug rate after the Intervention.



Non-equivalent control group design

- Still using DARE program as an example
- Suppose the city of Salt Lake decided not to implement the program, but the city of West Valley decided to do it.
- One can treat Salt Lake City as the control group and West Valley City as the experimental group. But we all know that these two groups are not equivalent, as people living in SLC are likely to be different from people living in WVC.
- Researchers can randomly select some schools from both cities. The drug use rate before and after the DARE program will be measured.
- If WVC shows a significant decrease in drug use rate but SLC does not, then one might be able to conclude that the DARE program is effective.
- Caution: Different demographics should be taken into consideration. The conclusion may not be generalizable, meaning the program might be useful for the people of WVC but not SLC due to demographic differences.

Multiple time-series designs

- Multiple time-series designs are the use of more than one set of data that were collected over time, as in accident rates of time in several states or cities, so that comparison can be made.
- The difference between multiple time-series design and time-series design is the involvement of more than one set of data.

How to evaluate success/failure?

- The most critical aspect of evaluation research is to determine whether the program studied succeeded or failed.
- There are almost never clear-cut answers.
- Two common ways of judgment.
 - Cost-benefit analysis - compare the cost of the program with the benefit of the program.
 - Quite often cost is easier to estimate than benefits. But with this approach one has to estimate how much money is saved by saving one highschooler from using drug, etc, in order to estimate benefits
 - Agreement - Sometimes, the criteria of success and failure are often a matter of agreement. The people responsible for the program may commit themselves in advance to a particular outcome that will be regarded as an indication of success.
 - For example, if people agree that a reduction of 10% in drug use rate attributed to the Intervention would be an indication of a success, then any reduction equal to or greater than 10% would be a success.

What are the common problems with evaluation research?

- Logistical problem - problems with getting subjects to do what they are supposed to do
 - Logistical problems emerge because evaluation researchers often lack sufficient control over the design in real-life contexts.
 - Reluctant participants
 - Administrative control
- Ethical issues
 - Particularly important when determining who receives or does not receive what type of stimulus.
 - Social interventions being evaluated may raise ethical issues.
 - Evaluation research may be a mask for unethical behavior.
- Use of research results
 - Evaluation research is designed to be used to make a difference in the execution of some programs.
 - However, sometimes, research results do not really have an impact.

Why results are often ignored?

- Implications may not be presented in a way that nonresearchers can understand.
- Results sometimes contradict deeply held beliefs.
- Administration has bested interest in a program (pet projects).

Social Indicators Research

- Social indicators are measurements that reflect the quality or nature of social life, such as crime rates, infant mortality rates, number of physicians per 100,000 population, and so forth. Social indicators are often monitored to determine the nature of social change in a society.
- Social indicators research provides an understanding of broader social processes.
- Researchers are developing more refined indicators.
- Research is being devoted to discovering the relationships among variables within whole societies.

An example of evaluation research

- Mittelman, Mary S., Steven H. Ferris, Emma Shulman, Gertrude Steinberg, Abby Ambinder, Joan A. Mackell, & Jacob Cohen (1995). A comprehensive support program: Effect on depression in spouse-caregivers of AD patients. *The Gerontologist*, 35(6), 792-802.
- Evaluating the effectiveness of a multifaceted, structured intervention for spouse-caregivers of Alzheimer's disease patients. Program is at NYU-Aging and Dementia Research Center since 1987.

- Specify the intervention
 - A multifaceted, structured intervention for spouse-caregivers of Alzheimer's disease patients.
- Specify the intended results of the intervention
 - Reduce the depression level of the spouse-caregivers of AD patients
- Specifying the population
 - Those spouses who care for AD patients
- Specify outcomes to be measured
 - Changes in depression level of spouse-caregivers of AD patients
- Measure study contexts and other variables
 - Demographics. The severity of the symptoms, how long the patient has been sick. Caregiver health, other support

- Sample:
 - Spouses who care for AD patients. At least one close relative living in the NYC metropolitan area. Caregivers who had received formal counseling or were participating in a support group were not eligible.
 - Recruitment of spouse-caregivers from 1987-1991
 - Eligible patients seeking treatment at the NYU-Aging Dementia Research center (91 spouse-caregivers)
 - Reference from the Alzheimer's Association of New York City
 - Other day care centers and social services for elderly in the area
 - Final sample = 206 spouse-caregivers

- Evaluation method:
 - Classical experimental design (note it's called randomized treatment/control trial – just a slightly different term)
 - Baseline interview, 4-month follow-up, 8-month follow-up, 12-month follow-up
- Hypotheses:
 - Caregivers in the treatment group would become less depressed or remain stable, while caregivers in the control group would become more depressed

- Measurements:
 - Dep. Variables (change in depression):
 - Caregiver depression: Geriatric Depression scale, a 30-item questionnaire in a yes/no format. (0-30 possible scores)
 - Major independent variable:
 - Treatment group or control group
 - Control variables:
 - See Table 1 for a list: caregiver physical health, reaction to frequency of troublesome patient behaviors, telephone calls from family and friends, visits from family and friends, help from family and friends, family cohesion, satisfaction with social network, formal services used, family income, caregiver gender, severity of patient impairment, amount of paid help, worried about financial future, amount of employment

- Analytical method
 - Multivariate analysis of co-variance (statistical method that can control other variables while isolating out the importance of the treatment variable)
- Result:
 - This intervention (enhances social support) has the potential for alleviating some of the deleterious effects of caregiving on mental health. The effect was not immediate in most cases, so that the difference between treatment and control groups only become significant after 8 months.