FLAWED SOCIAL EXPERIMENTS

David Greenberg and Burt Barnow

Nature of Talk

- All experiments are subject to flaws
- Continuum: Minor flaws to fatal failures
- Talk focuses on more serious end of continuum
- Serious flaws most often occur in underfunded experiments conducted by inexperienced organizations
- Talk focuses on 10 overlapping types of flaws
- Many of the examples taken from the *Digest of Social Experiments*
- Also see SSRN's eJournal on Social Experiments

(1) Few Show Up

- Due to
  - Insufficient marketing and outreach—Can be overcome
    - E.g. UK ERA
  - Target group too small
  - Insufficient benefits perceived by target group—Experiment may provides important information
- The resulting problem: Insufficient sample for hypothesis testing

(2) Failure to Properly Randomize

- Sometimes detected through comparisons of observed control and treatment group characteristics
- Can occur if members of the target group or those administering the experiment have some control over the randomization process
  - Wisconsin Self-Sufficiency First/Pay for Performance
- Can also occur through inadvertent errors
  - UK Supportive Caseloading

(3) Control Cross-Over (Control Contamination)

- Occurs when controls receive the treatment they are suppose to be denied
- Diminishes the size of the estimated impacts
- Larry Orr has suggested a correction when the number of cross-overs is known
- Example: Alternative Schools Demo, in which 13% of the controls in one site and 37% in another attended the alternative schools being tested—corrections for this were made

(4) Adverse Publicity

- Rare, but has caused some experiments to be terminated
- Example: The UK NDDP

(5) Failure to Implement the Treatment as Designed

- Experiments doesn't test what it was suppose to test
- Potentially serious but may not be fatal
- Example: German Targeted Negative Income Tax demonstration
- Example: UK Intensive Gateway Trail Blazer demonstration

(6) Failure to Adequately Communicate the Treatment

- A special type of implementation failure
- To respond to the treatment, persons in the experimental group presumably need to understand what the treatment is
- Not necessarily a flaw if the level of understanding is similar in the experiment to what it would be were the program adopted
- Example: US SIME/DIME experiment
- Example: Maryland Primary Prevention Initiative

(7) Inadequate Sample Size

- Between 1962 and 1996, 19% of all experiments had samples under 200 and 16% had samples between 200 and 500
- True impacts are often small so small sample size may result in a failure to detect impacts when one exists
- On the other hand, large impacts may be suspect (e.g. Perry Preschool Program)
- Small samples are due to: (1) few showing up, (2) attrition, (3) budget constraints

(8) Sample Attrition

- Usually much more serious when survey data, rather than admin data, are used for follow-up
- Resulting problems:
  - Small samples
  - Response bias
- Some examples
  - Food Stamp Empl and Trng demo (50% attrition)
  - Project Hope (24 of 140 interviewed in follow-up)
  - Partnership for Hope only (58 of 109 returned mail survey)

(8) Sample Attrition (continued)

- In the UK ERA demo, data for some outcomes were available from both a survey and admin data
  - Using admin data, findings on these outcomes indicated a strong response bias
  - Thus, the final report relied mainly on the admin data
- Even when attrition low, response bias can still result
- In the Tulsa IDA demo, a 10-year follow-up obtained responses for 855 of 1,103 individuals
  - There was evidence of response bias
  - Attempted corrections made through non-experimental difference-in-differences and propensity matching

(9) Changes in the environment

- Unlike lab experiments, the environment around social experiments cannot be controlled
- Changes can take place in
  - The internal environment of the institution in which an experiment is taking place
  - The external environment
- Example: New Jersey Income Maintenance Experiment

(10) Experiment Designed, But Not Fielded

- Causes
  - Design too complex
  - In a demo, resistance to testing the treatment
  - In an existing program, resistance by stakeholders to evaluating “their program”
- Resistance to random assignment
- Example: UK NDDP
- Example: US Youth Offender Demo