Connections between *Experimental Designs* and non-experimental designs

Elizabeth Stuart

Johns Hopkins University
Bloomberg School of Public Health
Departments of Mental Health and Biostatistics

August 2, 2009

My work...the design of non-experiments

- My work has focused on estimating causal effects in non-experimental settings
  - Using methods such as propensity scores to ensure the comparison of similar groups of individuals
- Very applied
  - Education
  - Public policy (including time as a Researcher at Mathematica Policy Research)
  - Public health (now faculty at Johns Hopkins Bloomberg School of Public Health)
  - Mental health (primary appointment)
- Overall theme: the careful design of non-experimental studies

Gertrude Cox

- I sadly know relatively little about her
- Although we are some sort of academic relatives
    - My advisor was Don Rubin, whose advisor was Bill Cochran
- Preparing for this talk prompted me to go back and look through their classic text *Experimental Designs* (Cochran and Cox, 1950/1957)
- Struck by the number of connections and insights that are relevant for non-experimental studies estimating causal effects, even in the first 10 pages
- Will discuss some of those now

The role of statisticians

Cochran & Cox (1957, p. 1):

> *Since statisticians do not usually perform experiments, their claim to attention when they write on this subject requires some explanation. It is true that on many important aspects of experimentation the statistician has no expert knowledge. Nevertheless, in recent years, research workers have turned increasingly to statisticians for help both in planning their experiments and in drawing conclusions from the results. That this has happened is convincing evidence that statistics has something to contribute.*

Cochran & Cox, p. 10:

> *Participation in the initial stages of experiments...leads to a strong conviction that too little time and effort is put into the planning of experiments. The statistician who expects that his contribution to the planning will involve some technical matter in statistical theory finds repeatedly that he makes a much more valuable contribution simply by getting the investigator to explain clearly why he is doing the experiment, to justify the experimental treatments whose effects he proposes to compare, and to defend his claim that the completed experiment will enable its objectives to be realized...These comments are offered with diffidence, because they concern questions on which the statistician has, or should have, no special authority, and because some of the advice is so trite that it would be unnecessary if it were not so often overlooked.*

In non-experimental studies

- These points still an issue
- I often find one of my most valuable roles is to ask basic questions and have people think carefully about what they want to estimate
  - What are the units?
  - What is the treatment of interest (what would you randomize if you could)?
  - What is the control condition?
- e.g., long-term effects of heavy adolescent marijuana use
- e.g., what are the effects of childhood maltreatment?
- e.g., what are the effects in adulthood of childhood obesity?
- This is a key role that statisticians can play...what is the question trying to be answered? Is that question answerable from the data?

More broadly: Study design

In experiments:

- Cochran & Cox (1957) laid out the major components of experimental design
  - Randomized blocks, Latin squares, Lattice squares
- Helped put emphasis on clear thinking about design of experiments

In non-experimental studies:

- "Traditional" methods for observational studies involved simply running many many regressions of the outcome predicted by various sets of covariates
- Recent years has seen increased emphasis on the design of non-experimental studies
- Rubin (2008): "For objective causal inference, design trumps analysis"
- Try to replicate the design of a randomized experiment:
  - Think about a template experiment that could have been done (at least theoretically)
  - Do not use outcome value in setting up design (prevents picking a certain design to get a desired result)
  - Compare treatment and comparison groups that are as similar as possible on the observed covariates
- Matching methods such as propensity scores (Stuart & Rubin, 2007)
- Design elements to limit impact of unobserved confounders (Heller, Rosenbaum, & Small, 2009)

Example: The effects of "heavy" marijuana use

- For most uses of propensity scores, need to have a binary "treatment"
- Here: use "heavy" marijuana use ($>$ 20 times in adolescence)
  - Collected at age 16
  - Measure of level of use (never, 1-2, 3-9, 10-19, 20-39, 40+ times)
- Based on literature and distribution of data
- 26% classified as "heavy users"

Simple 1:1 matching

- Green and Ensminger (2006) use nearest neighbor 1:1 propensity score matching
- "Exact" match on sex
- 137 heavy users matched to 137 non-heavy users

- Fairly good balance obtained
- Effects then estimated using this matched sample

Summary of Balance

Variable	Heavy Users	All Controls	Matched Controls
% Male	67.2	39.9	67.2
Family income	4.66	4.99	4.77
% below poverty	54.7	47.1	52.6
Underachievement	0.61	0.59	0.57
Aggression	0.66	0.41	0.60
Shyness	0.50	0.44	0.45
Immaturity	0.61	0.55	0.56
Inattention	0.67	0.48	0.59
N	137	393	137

The rationale behind this careful design: Confounding

Cochran & Cox (1957, p. 6):

> *It is easy to conduct an experiment [non-experiment] such that no useful inferences can be made...To take a simple example, suppose that in the comparison of the calculating machines each sum of squares had been computed first on machine A and then on machine B. Now it is quite possible that increased familiarity with the data will enable the second computation to be done faster than the first...If an experiment is conducted in this way, the observed difference in speed (B-A) is an estimate of the true difference, plus the unknown difference in speed between a second calculation and a first.*

- Argument for randomization
- Of course need to worry about this in non-experimental studies too...motivation behind propensity scores, instrumental variables, regression discontinuity

Providing a service to applied researchers

Cochran & Cox (1957, p. ix):

> *At that time requests were received rather frequently from research workers. Some wanted advice on the conduct of a specific experiment: others...asked for a plan or layout that could be followed during the experimental operations.*

- I could say the same thing now about non-experimental studies...applied researchers want guidance, a place to turn for advice

Application areas

- Cochran & Cox: agriculture, biology
- Me: public policy, public health
  - Education, where randomized experiments increasingly used
    - e.g., nationally representative evaluation of Upward Bound
  - Mental health, where interest often in things you can't randomize
    - e.g., long-term effects of adolescent drug use (Stuart & Green, 2008)
  - Policy changes, where want to separate out secular time trends from the effect of interest
    - e.g., effects on quality of care of reporting of nursing home quality measures (Werner et al., 2009)
  - Public policy, where effects not always clearly defined
    - e.g., estimating the effect on relationship quality of a health marriage initative...what to do about couples who break up (McConnell et al., 2008)?

Challenges in these areas

- Randomization often not feasible
- Effects often take a long time to see (e.g., 1st grade interventions to prevent problem behavior 15 years later)
- Effects often relatively small
- Sometimes effects only seen for a subgroup of the population
- But we could probably learn a lot by remembering some of the basic points in Cochran & Cox

Communicating to non-statisticians

Cochran & Cox (1957, p. 3):

> *Obviously it cannot be expected that the solution will provide the exact value of the unknown true difference. As a less ambitious goal we might hope to be able to find 2 limits within which the exact value is certain to lie, but even this cannot quite be attained. What can be done is that for any chosen probability, say .95, two limits are found such that the probability that they enclose the true difference is .95.*

In non-experiments

- Importance of conveying statistical concepts clearly (Stuart, 2007)
- e.g., 1:1 nearest neighbor matching and associated balance measures
  - Very intuitive; can explain to almost anyone
  - For each treated person, try to find a comparison person who looks as similar as possible on the covariates
- e.g., language of instrumental variables, especially for dealing with non-compliance (Angrist, Imbens, & Rubin, 1996)
  - The problem: Estimating the effect of actually taking some treatment of interest (not just of being randomized or told to take that treatment)
  - Assumptions had been expressed in terms of uncorrelated error terms, distributional assumptions
  - AIR re-expressed assumptions in terms of "always-takers," "never-takers", "defiers", "compliers"
  - e.g., no defiers: no one who would take the treatment when in the control group and wouldn't take it when in the treatment group
  - e.g., no effect of being told to take the treatment for always-takers or never-takers

Now a personal note: What the Gertrude Cox award has meant to me

- Confidence
  - I received the Gertrude Cox award in 2000, the summer after my 1st year of graduate school at Harvard
  - I was new to statistics, had taken 1 course in undergrad
  - First year of grad school was understandably tough
  - Receiving the award gave me a confidence boost that I needed then
- Connections and collaborations
  - JSM also always reminds me of that time and of how much the award meant to me
  - My department paid for me to attend JSM to receive the award
  - Helped me connect with the broader statistical community
  - Helped me generate collaborations with other researchers, especially at the Census bureau
  - I have now attended every JSM since then, except last year (my daughter was due August 13)

- Colleagues
  - Being involved in the scholarship, as a recipient, race-runner, and committee member, has helped me connect with other people in the community
  - Helped expose me to COWIS, Caucus on women in statistics
  - Especially helpful to be exposed to other female statisticians, as role models and friends
    - (My graduate school department had no female faculty members at the time I was there; definitely a transition from my college, which was all women's)

Quote from letter by Gertrude Cox to young woman enquiring about statistics:

> *The field of statistics is certainly wide open to women. If you are willing to take the mathematics and science courses and then work very hard to get beyond the junior level, there are all sorts of opportunities to go as far as you wish.*

From http://www.lib.ncsu.edu/exhibits/cox/career.html

Looking forward

- I sincerely hope the scholarship can continue another 20 years
  - (Although I also look forward to the day when we no longer need to encourage women to enter statistical professions)
- The scholarship itself provides encouragement to women just entering the field
  - (Incidentally, may be worthwhile considering adding a travel award component, to encourage recipients to attend JSM)
- The race provides a fun networking and social activity, as an alternative to all the meetings and mixers!
- The combined activities provide visibility and a focal point for women in statistics, and well-deserved attention to Gertrude Cox
- Thank you to all of you who make these things possible!

References

- Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs, 2nd Edition.* New York: John Wiley & Sons, Inc.
- Imai, K., King, G., and Stuart, E.A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A* 171: 481-502.
- Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics* 2(3): 808-840.
- Stuart, E.A., and Green, K.M. (2008). Using Full Matching to Estimate Causal Effects in Non-Experimental Studies: Examining the Relationship between Adolescent Marijuana Use and Adult Outcomes. *Developmental Psychology* 44(2): 395-406.
- Stuart, E.A. and Rubin, D.B. (2007). Best Practices in Quasi-Experimental Designs: Matching methods for causal inference. Chapter 11 (pp. 155-176) in *Best Practices in Quantitative Social Science*. J. Osborne (Ed.). Thousand Oaks, CA: Sage Publications.

- My webpage: http://www.biostat.jhsph.edu/∼estuart

- Me: estuart@jhsph.edu