

Prepared for the “Study Designs and Analytic Strategies for Environmental and Policy Research on Obesity, Physical Activity and Diet” Meeting, April 8, 2008, Washington DC.

Let's get real:

Statistical approaches for environmental and policy research

J. Michael Oakes, PhD

McKnight Presidential Fellow
Associate Professor
Division of Epidemiology & Community Health
Minnesota Population Center
University of Minnesota

oakes007@umn.edu

Themes

- Models substitute assumptions for data
- Always think “experiment”

Topics for today

1. Problems

- a) Statistical fetish
- b) Multiple regression
- c) Large sample surveys
- d) Prospective cohort studies

2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

1. Problems

- a) **Statistical fetish**
- b) Multiple regression
- c) Large sample surveys
- d) Prospective cohort studies

Fancy statistics. Not!

- Jenner (1796) + cow pox
- Semmelweis (1844) + puerperal fever*
- Snow (1854) + cholera*
- McKay (1901) + tooth decay*
- Goldberger (1914) + pellagra*

* Natural experiment

EVALUATION

PETER H. ROSSI
HOWARD E. FREE
MARK W. LIPSEY

A SYSTEMATIC APPROACH

SIXTH EDITION

Springer Series in Statistics

Paul R. Rosenbaum

Observational Studies

Second Edition

 Springer

Regression Analysis Constructive Critique

Richard A. Berk

*Quantitative Techniques
Social Sciences Series*

11

1. Problems

- a) Statistical fetish
- b) Multiple regression**
- c) Large sample surveys
- d) Prospective cohort studies

**As practiced and with rare exception,
MRMs inhibit progress
and are thus responsible for many deaths and
an enormous waste of resources.**

- Regression adjustment makes up data
- Model specification by sampling error, i.e. p-value
- Smooth over sparse cells without knowledge
- Exchangeability
- Bad for ecological data; worse for indiv. social inquiry
- Rote application without theory or insight
- Discourage consideration of analog experiment

Vandenbroucke, J. P. 1987. "Should we abandon statistical modeling altogether?" *Am J Epidemiol* 126:10-3.

D.B. Petitti and D.A. Freedman. 2005 "Invited commentary: How far can epidemiologists get with statistical adjustment?" *Am J of Epidemiol* 162:1–4.

Berk, Richard. 2004. *Regression Analysis. A constructive critique*. Thousand Oaks, CA: Sage Publications.

“Structural Confounding”

Confounding (ie, imbalance) that cannot be solved by collecting more/different data.

‘We can only evaluate sharply distinct treatments that could happen to anyone.’

Paul Rosenbaum (2002)

‘If the differences between groups is large, the average value applied to each group with adjustment may represent “no man’s land”, a place where no actual observations exist. Given this scenario, the interpretation of the estimate becomes speculative rather than soundly based. Heroic modeling assumptions are required.’

William Cochran (1957)

1. Problems

- a) Statistical fetish
- b) Multiple regression
- c) Large sample surveys
- d) Prospective cohort studies

**Increasing heterogeneity
comes from Sir RA Fisher (1935)
and randomization inference**

(...on plants)

**Fisher thought JS Mill's (1864)
insights about eliminating heterogeneity
(ie, method of difference)
was dumb...**

It's not, esp. for observational designs!

Fisher believed, and showed, that reducing heterogeneity by half has same effect on precision as quadrupling sample size.

Thus, why worry about heterogeneity...
just get more subjects with a more heterogeneous sample.

But this is not true in observational designs

More heterogeneity increases probability of “hidden bias”, including residual confounding...

Γ is a measure of the degree of departure of from a study that is free of hidden bias due to unmeasured confounders.

If $\Gamma > 1$, matching on x may not suffice to balance confounders, u .

How big must Γ be to make conclusions suspect? That is, how much “residual” confounding can there be before estimated CI’s that account for bias effects become so *potentially* wide as to be unhelpful?

We can employ (stat power) of sensitivity measure, Γ

Number matched pairs	Treatment effect, τ	σ	VAR (Mean diff)	Power at $\Gamma = 1$	Power at $\Gamma = 1.5$	Power at $\Gamma = 2$
120	0.50	1	1/120	1.00	0.96	0.60
30	0.50	0.50	1/120	1.00	1.00	0.96

Quadruple sample size, twice variance but less power to detect hidden bias!

In observational studies:

Reducing heterogeneity reduces *both*
sampling variability and sensitivity to hidden bias.

Increasing sample size increases precision but does little to reduce bias.

Even if smaller, homogeneous studies are better!

Or, beware of p-values from big fat national sample data.

1. Problems

- a) Statistical fetish
- b) Multiple regression
- c) Large sample surveys
- d) **Prospective cohort studies**

Cohort studies are not going to be useful for ALR/HER research.

“Cohort study” coined by Frost in 1935... referring to cohorts (ie, generations) of men dying from tuberculosis.

Subsequent work

- Doll & Bradford Hill on smoking and lung cancer (British docs) beginning in 1951... as a follow-up to a case-control study.
- Dawber initiated Framingham heart study in 1949.

Advantage over case-control design?

No behavioral effect of disease and no bad controls 😊

Why?

ALR/HER research aim to assess higher-level effect (eg, policy) on some individual outcome (eg, BMI).

One must disentangle person-level causes of outcome (eg, genes, propensity for PA) from macro or policy-level ones (eg, school policy).

Trouble is, multilevel confounding and endogeneity.

Absent cross-over (eg, randomization), there is no data for needed counterfactual substitute ☹️ .

Analysts cannot identify effect of endogenous policy over-and-above characteristics of subjects, even over time.

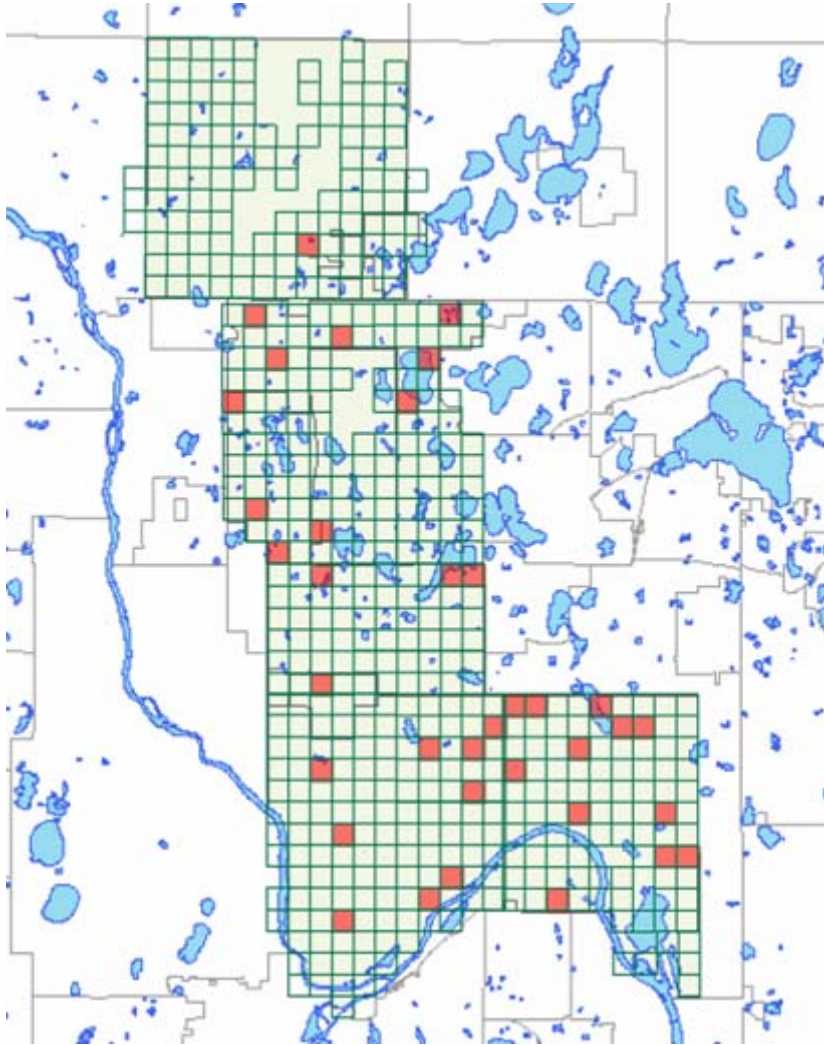
Is the policy working or is there a secular trend?

2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

Sampling Problem:

**Maximize environmental variation
while *minimizing*
subject demographic differences
(ie, max exchangeability)**



- 36 areas by density and “block size”
- **718 participants: surveys, travel diaries, motion detectors, measured height and weight**
- **50+ environmental variables, measured at multiple geographies around each participant**

Epidemiologic Perspectives & Innovations



Methodology

Open Access

The effects of neighborhood density and street connectivity on walking behavior: the Twin Cities walking study

J Michael Oakes*¹, Ann Forsyth² and Kathryn H Schmitz³

Address: ¹Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN, USA, ²Department of City and Regional Planning, Cornell University, Ithaca, NY, USA and ³Division of Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Email: J Michael Oakes* - oakes007@umn.edu; Ann Forsyth - af16@cornell.edu; Kathryn H Schmitz - schmitz@mail.med.upenn.edu

* Corresponding author

2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Other stuff

b) Propensity Score Matching

- Minimize threats to inference assoc with regression
 - Mimic the analog RCT
 - Model exposure, not outcome, with regular logistic regression
 - Match on predicted probability of exposure (within caliper)
 - Toss non-matchers (ie, enforce support)
 - Simply subtract outcomes to get ATT effect
-
- Easily implemented in Stata / SAS
 - How best to model exposure?
 - Unmeasured confounders?
 - Do we seek ATT or ACE?
 - Throw away data?
 - Too simple... time and other factors?

2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) **Instrumental variable**
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

c) Instrumental Variable Models

- Addresses unmeasured confounders
- Common in econometrics
- Mimic the analog RCT by exploiting “instrument”
- “Instrument” is variable correlated with exposure but not outcome
 - 1st model exposure with instrument
 - 2nd use predicted values of exposure in regression model
- Note: Randomization is IV, so too is “mother nature” at times
- If assumptions hold, has causal interpretation
 - Easily implemented in Stata / SAS
 - Good instruments? Natural experiments are tough to find...
 - Bad instruments can make inference worse
 - Not viewed as very useful in epidemiology

2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

d) Cross-validation studies

- Take (regression) model seriously
- *Acid test* for any model
 - Fit/build model in one dataset
 - Apply model in new data set
 - Calculate predicted outcomes from model in 2nd data
 - Compare model predicted to observed values
 - If model is “good”, predicted = observed

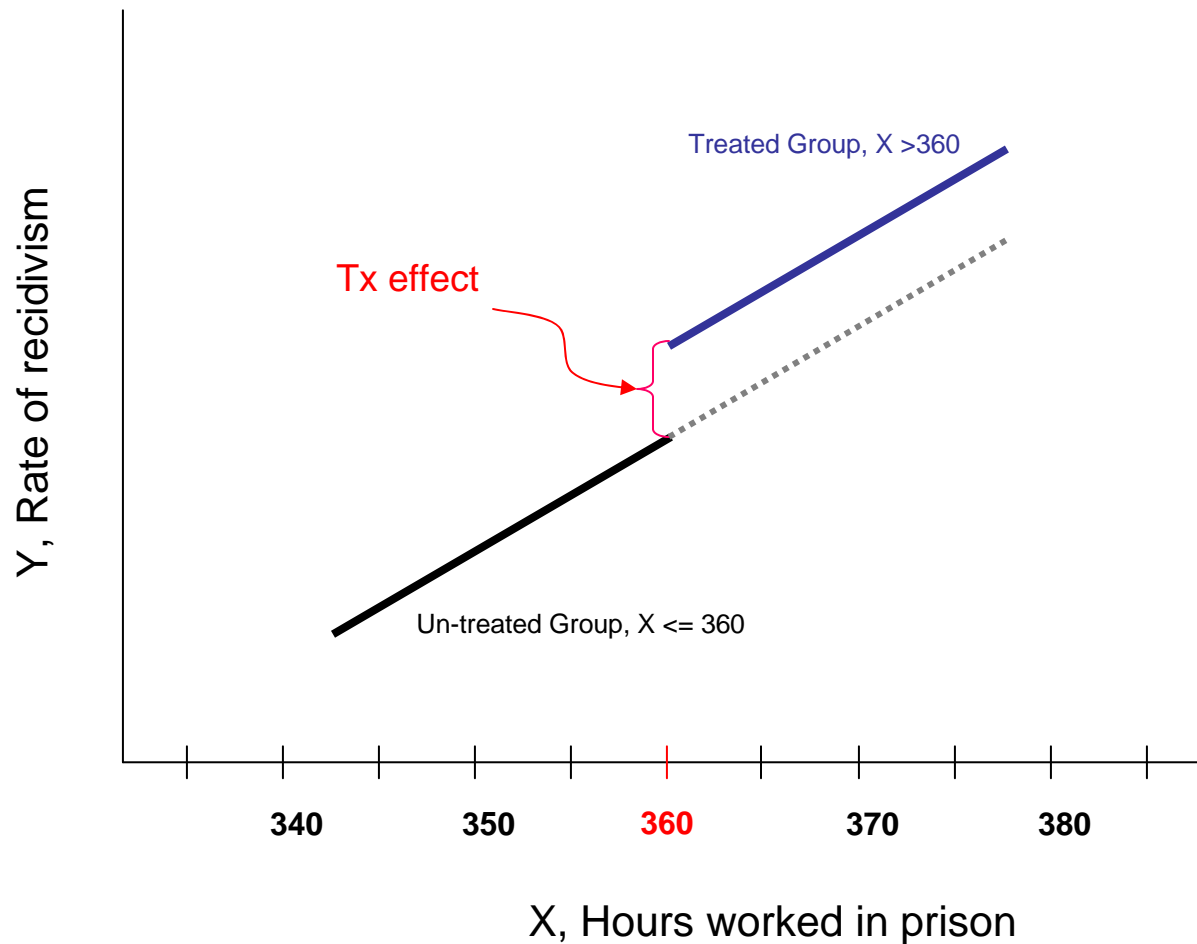
2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

e) Regression discontinuity designs

- Addresses “selection” (eg, 2 populations), by design
- Cx group is unlike Tx wrt some confounder!
- But model yields unbiased estimate of Tx effect
- Focuses attention on “local” matching subjects
- Useful if denial of treatment *to needy* is unethical
- Useful is “natural” cut-off or eligibility for Tx
- Researcher *must* have control over assignment, i.e. cutoff

e) Regression discontinuity designs



2. Recommendations

- a) Matched/Homogeneous sampling
- b) Propensity score matching
- c) Instrumental variable
- d) Cross-validation
- e) Regression discontinuity
- f) Other stuff

Marginal Structural Models

- Time-dependent confounders; multiple treatments over time
- Motivated by AIDS treatments and “selection” for next drug
- When we have a non-randomized treatment that is offered over multiple time points, covariate adjustment will not work because you cannot adjust for confounders affected by treatments without biases desired effect estimates
- Solution? MSM
 - Model probability of Tx (ie, propensity score) given confounder
 - Take inverse of pscore and do stuff to make “weight”
 - Conduct “regular” weighted analysis
 - Effect estimates are unbiased, given assumptions

The Group-randomized trial

Exploit in-tact social groups

- N'hoods
 - Schools
 - Work-sites
 - Clinics/practices
-
- Capitalize on human interaction
 - SUTVA* is much easier to assure
 - Constructive of health improvements
 - Expensive, difficult, ...ineffective?

Agent-based Models

Computer simulation to advance theory/concepts

See Auchinloss & Diez Rouz. *In Press*. American J Epidemiology

- Methodological individualist
- Nonlinear
- SUTVA not an issue
- Different kind of explanation
- But not empirical...

Oh, before I go...

- Models substitute assumptions for data
- All models are wrong, $\frac{1}{100,000}$ are useful
- Seek simple designs with intense content
- To improve health: (quasi)experiments *with groups*
- Observational studies have experimental analogs
- Gell-Mann and simple stuff... like quantum physics