THIS PAGE LEFT BLANK

FOR PAGINATION

THIS PAGE LEFT BLANK

FOR PAGINATION

Stat 209B-- Lectures, Course Files, and Readings

Week 0

Course introduction (slides and audio posted on main page)

Background readings (not required, but of interest if you haven't seen these before)

1. Correlation and Causation: A Comment, Stephen Stigler *Perspectives in Biology and Medicine*, volume 48, number 1 supplement (winter 2005)
2. Secret to Winning a Nobel Prize? Eat More Chocolate (Time)
   Publication: Chocolate Consumption, Cognitive Function, and Nobel Laureates Franz H. Messerli, M.D. N Engl J Med 2012; 367:1562-1564 October 18, 2012
3. *David Freedman chapters*.
   From Association to Causation: Some Remarks on the History of Statistics;
   Statistical Models for Causation: A critical review
   Statistical Models and Shoe Leather, *Sociological Methodology*, Vol. 21. (1991), pp. 291-313. JStor link

Week 1

Lecture slides, week 1 (pdf)
Audio companion, week 1
parta partb partc

1. Encouragement Designs: example of potential outcomes formulation.

Lecture Topics

Illustration using encouragement design representation in Holland (1988). copies of selected overheads.
Encouragement Designs. Potential outcomes formulation and IV parameter estimation in Holland (1988). Estimation handout
Do regression methods (path analysis) identify causal effects? Demonstrations of failure for Holland's encouragement design. class handout Encouragement design slides

Primary Readings

Paul Holland, Causal Effects and Encouragement Designs. Causal Inference, Path Analysis, and Recursive Structural Equations Models
Paul W. Holland Sociological Methodology, Vol. 18. (1988), pp. 449-484. (Encouragement design results; sections 3-5)
Holland Appendix (esp pp. 475-480) presents the potential outcomes formulation.

Abstract Rubin's model for causal inference in experiments and observational studies is enlarged to analyze the problem of "causes causing causes" and is compared to path analysis and recursive structural equations models.
A special quasi-experimental design, the encouragement design, is used to give concreteness to the discussion by focusing on the simplest problem that involves both direct and indirect causation.
It is shown that Rubin's model extends easily to this situation and specifies conditions under which the parameters of path analysis and recursive structural equations models have causal interpretations.

Encouragement Design research examples:

Sesamee Street evaluation
Gelman-Hill text sec 10.5; Data Analysis Using Regression and Multilevel/Hierarchical Models
Salt and Blood Pressure clinical trial
Publication: Feasibility and efficacy of sodium reduction in the Trials of Hypertension Prevention, phase I Trials of Hypertension Prevention Collaborative Research Group. S K Kumanyika, P R Hebert, J A Cutler, V I Lasser, C P Sugars, L Steffen-Batey, A A Brewer, MI. Hypertension doi: 10.1161/01.HYP.22.4.5021993;22:502-512

2. Mediating (process) variables

Lecture Topics

Historical (Barron-Kenny) methods David Kenny web page
R-implementations: mediating variables data analysis example data file
Barron-Kenny method via Sobel function in the multilevel package.
More extensive implementation (incl BCa bootstrapping) function `mediation` in package `MBESS` Ken Kelley;
power and sample size calculations in package `powerMediation`
`mediation` package. takes the topic up a large level of complexity/capabilities

Primary Readings

Vignette for `mediation` package Causal Mediation Analysis Using R
Mediation Analysis David P. MacKinnon, Amanda J. Fairchild, and Matthew S. Fritz Department of Psychology, Arizona State University, Tempe, Arizona 85287-1104; Annu. Rev. Psychol. 2007. 58:593-614

Mediation research examples:

Framing experiment
Brader T, Valentino NA, Suhat E (2008). What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration." American Journal of Political Science, 52(4), 959-978. jstor link
Data in `mediation` package; data description and analyses in `mediation` package vignette (linked below)
Bench Science vs Path Analysis: Exercise and Alzheimers
The irisin bench-science mediation example is discussed at the beginning of Week 2 lecture for recap and because I couldn't find it at the time.
NYTimes:How Exercise May Help Keep Our Memory Sharp .
Publication: Exercise-linked FNDC5/irisin rescues synaptic plasticity and memory defects in Alzheimer's models *Nature Medicine* volume 25, pages165-175 (2019)
Mediated moderation?
Stanford Medicine Common opioids less effective for patients on SSRI antidepressants Publication: Predicting inadequate postoperative pain management in depressed patients: A machine learning approach Arjun Parthipan,Imon Banerjee,Keith Humphreys,Steven M. Asch,Catherine Curtin,Ian Carroll ,Tina Hernandez-Boussard Published: February 6, 2019https://doi.org/10.1371/journal.pone.0210575
New Yorker. December 23, 2013. The Power of the Hoodie-Wearing C.E.O. Publication: The Red Sneakers Effect: Inferring Status and Competence from Signals of Nonconformity Author(s): Silvia Bellezza, Francesca Gino, and Anat Keinan Source: *Journal of Consumer Research*

Additional Resources

Mediators and Moderators of Treatment Effects in Randomized Clinical Trials Helena Chmura Kraemer; G. Terence Wilson; Christopher G. Fairburn; W. Stewart Agras Arch Gen Psychiatry. 2002;59:877-883
additional technical papers. Causal Mediation Analysis Using R K. Imai, L. Keele, D. Tingley, and T. Yamamoto American Political Science Review Vol. 105, No. 4 November 2011
Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies
MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M.,West, S. G., Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.
Useful expositions Using R
Chapter 14: Mediation and Moderation Alyssa Blair
Mediation and Moderation Analyses with R - OSF presentation slides

Week 1 Review Questions

Question 1. Mediating Variable Computations: Class example continued
The data set shown in class example ss423 is linked above and in the legacy directory http://web.stanford.edu/~rag/stat209/ss423
for predictor (IV) 'belong' outcome 'depress' and (potential) mediating variable 'master' The class example showed you the Baron-Kenny analysis using functions from the multilevel and MBESS packages.
Here just use 'lm' basic regression and the recipees from the class handout to recreate point estimates and asymptotic standard errors, significance tests for the mediating variable effect.
Compare your result with the class example posting.
Extra: also try out the more 'sophisticated' functions in the mediation package.

Solution for question 1

Question 2. Potential Outcomes, Encouragement Design Estimation and (Causal) Mediation
Task 1. Create a potential Outcomes dataset following the first ALICE specification in the posted slides (week 3) ## ALICE example beta = 3 rho = 3 tau = 1, delta = 3 (I did n=400; larger would be better so I redid with n = 6400)

Task 2. Use the artificial data to show the results for the mediation (indirect) effect by hand doing the 3 regressions using multilevel package (sobel) using MBESS package using the causal mediation estimation ACME from the mediation package and compare with rho*beta

Task 3 estimate beta by the Wald estimator (assuming tau = 0) and estimate mediation effect

Solution for question 2

Question 3. Sesame Street: Encouragement Design research example
Sesame Street research setting and data description given pdf p.30 of Lecture 1 (also Gelman text).
For this exercise use `postnumb` : posttest on numbers (0-54), along with the measures encour and regular from the class example in Lecture 1.
Use the encouragement design formulation to estimate the effect on child cognitive development (postnumb here) of watching more Sesame Street.
What assumption is necessary for the IV estimation in this design?
Obtain a point and interval estimate for the effect of viewing (use `ivreg` as in class example).
From simple descriptives reproduce this instrumental variables estimate (Wald estimator).
The second approach (path analysis) analyzed by Holland requires what assumption?
Obtain the path analyses (regression) estimate for the effect on child cognitive development (postnumb here) of watching more Sesame Street.
Compare with the IV estimate (which employs different assumptions).

Solution for question 3

Week 2

Moderating Variables in experimental studies (heterogeneous treatment effects)

Lecture slides, week 2 (pdf)
Audio companion, week 2
parta partb partc

Lecture topics
0. Moderation, mediation recap slide
1. Review: formulation and purposes of analysis of covariance
basic (old) ancova exposition slides ancova and extensions, math notes
High School and Beyond (observational study) school means data example HSB ancova handout (ascii version) data for HSB ancova HSB ancova, scanned pdf
2. Moderating variables, Heterogeneous Treatment Effects (CATE).
Analyzing treatment effects as a function of covariate(s)
CNRL, including Johnson-Neyman technique cnrl data cnrl analysis (extended)

Primary Readings
Ancova and extensions
Rogosa, D. R. (1980). Comparing nonparallel regression lines. *Psychological Bulletin*, 88, 307-321. [a better quality scan from the APA site]
R resources (below).

Moderation research examples:
Gender differences in effectiveness of aspirin.
Aspirin may be less effective heart treatment for women than men
Publication: Aspirin Resistance in Patients with Stable Coronary Artery Disease, in the *Annals of Pharmacotherapy* April 2007
Moderating variables can be your friend (statistics is the only friend you need) music: I've got friends in low places
Wash Post: Why smart people are better off with fewer friends .
Publication: Country roads, take me home... to my friends: How intelligence, population density, and friendship affect modern happiness. *British Journal of Psychology* 2016
ATI research
Snow R.E. (1978) Aptitude-Treatment Interactions in Educational Research. In: Pervin L.A., Lewis M. (eds) Perspectives in Interactional Psychology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-3997-7_10
Family SES as a moderating variable in nature/nuture:
Why Rich Parents Don't Matter UTexas press release: Being Poor Can Suppress Children's Genetic Potentials Publication: Emergence of a Gene x Socioeconomic Status Interaction on Infant Mental Ability Between 10 Months and 2 years DOI: 10.1177/0956797610392926 Psychological Science published online 17 December 2010 Elliot M. Tucker-Drob, Mijke Rhemtulla, K. Paige Harden, Eric Turkheimer and David Fask

R implementations and Resources
package `probemod` manual
package `interactions` intro vignette: Exploring interactions with continuous predictors in regression models manual

Additional Resources, Ancova and extensions
Improving Present Practices in the Visual Display of Interactions Advances in Methods and Practices in Psychological Science
analysis of covariance: Background/historical papers:
Covariance Adjustment in Randomized Experiments and Observational Studies Paul R. Rosenbaum *Statistical Science*, Vol. 17, No. 3. (Aug., 2002), pp. 286-304. Jstor
Some Aspects of Analysis of Covariance, A Biometrics Invited Paper with Discussion. D. R. Cox; P. McCullagh *Biometrics*, Vol. 38, No. 3, (Sep., 1982), pp. 541-561. Jstor
Analysis of Covariance: Its Nature and Uses William G. Cochran *Biometrics*, Vol. 13, No. 3, Special Issue on the Analysis of Covariance. (Sep., 1957), pp. 261-281. Jstor
The Use of Covariance in Observational Studies W. G. Cochran *Applied Statistics*, Vol. 18, No. 3. (1969), pp. 270-275. Jstor
Estimation of the Slope and Analysis of Covariance when the Concomitant Variable is Measured with Error James S. Degracie; Wayne A. Fuller *Journal of the American Statistical Association*, Vol. 67, No. 340. (Dec., 1972), pp. 930-937. Jstor
Deep background Neter-Wasserman text (Applied linear statistical models. Neter, Kutner, Nachtsheim and Wasserman 1996. Fifth edition. Homewood IL: Irwin, Inc.) chapters 22 and 8.
Johnson-Neyman technique and aptitude-treatment interaction (ATI)
Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. Irvington
Regions of Significant Criterion Differences in Aptitude-Treatment-Interaction Research Leonard S. Cahen; Robert L. Linn *American Educational Research Journal*, Vol. 8, No. 3. (May, 1971), pp. 521-530. Jstor
Identifying Regions of Significance in Aptitude-by-Treatment-Interaction Research Ronald C. Serlin; Joel R. Levin *American Educational Research Journa*l, Vol. 17, No. 3. (Autumn, 1980),

pp. 389-399. Jstor
Defining Johnson-Neyman Regions of Significance in the Three-Covariate ANCOVA Using Mathematica Steve Hunka; Jacqueline Leighton *Journal of Educational and Behavioral Statistics*, Vol. 22, No. 4. (Winter, 1997), pp. 361-387. Jstor
discussion of substantive issues: Trait-Treatment Interaction and Learning David C. Berliner; Leonard S. Cahen *Review of Research in Education*, Vol. 1. (1973), pp. 58-94. Jstor

Week 2 Review Questions

Question 1. Background: standard analysis of covariance.(no moderating variable)

A researcher is studying the effect of an incentive on the retention of subject matter and is also interested in the role of time devoted to study.
Subjects are randomly assigned to two groups, one receiving (C3 = 1) and the other not receiving (C3 = 0) an incentive. Within these groups, subjects are randomly assigned to 5, 10, 15, or 20 minutes of study (C2) of a passage specifically prepared for the experiment. At the end of the study period, a test of retention is administered.
Treat the study time as a covariate for investigating the differential effects of the incentive. Does using the covariate improve precision in estimating the effect of incentive?
Does the ancova assumption of a constant treatment effect at levels of StudyMin appear reasonable?
full data are in file retention.dat formerly located at http://statweb.stanford.edu/~rag/stat209/retention.dat
***note*:** As of January 2022 Statistics Dept. servers eliminated--files linked at statweb.stanford.edu/~rag/stat209/[file] or www-stat.stanford.edu/~rag/stat209/[file] now reside at rag.su.domains/stat209/[file].
Linked materials resolve to rag.su.domains seemlessly but to read in data files to R requires using the new file location.
update: statweb file locations will `read.table` in R sucessfully; the older equivalent www-stat almost surely will not.
Solution for question 1

Question 2. Revisit High School and Beyond ancova from Week 2 lecture

In the class example we used school level (mean, gradient) outcomes and used school mean ses as a covariate. Investigate the usefulness of that covariate by comparing the ancova in class example with just a simple t-test (sector) on these school level outcomes. What is the difference in precision between using the covariate or not? As this is not an RCT (revisit in Unit 2), also look at differences in the estimate of the sector effect (bias?).
Solution for question 2

Question 3. Comparing Regressions (demonstration data, not an RCT)

Let's give recognition to the guys who made S (and R) and take some data from Venables, W. N. and Ripley, B. D. (1999) Modern Applied Statistics with S-PLUS. Third Edition. Springer (now up to 4th edition). Chap 6 section 1 considers analysis of the data set whiteside (available as part of MASS subset of VR package) to access
> library(MASS) # do need to load library, MASS is part of base R > data(whiteside) > ?whiteside
Description
Mr Derek Whiteside of the UK Building Research Station recorded the weekly gas consumption and average external temperature at his own house in south-east England for two heating seasons, one of 26 weeks before, and one of 30 weeks after cavity-wall insulation was installed. The object of the exercise was to assess the effect of the insulation on gas consumption.
Format The whiteside data frame has 56 rows and 3 columns.:
Insul A factor, before or after insulation.
Temp Purportedly the average outside temperature in degrees Celsius. (These values is far too low for any 56-week period in the 1960s in South-East England. It might be the weekly average of daily minima.)
Gas The weekly gas consumption in 1000s of cubic feet.
Source. A data set collected in the 1960s by Mr Derek Whiteside of the UK Building Research Station. Reported by Hand, D. J., Daly, F., McConway, K., Lunn, D. and Ostrowski, E. eds (1993) A Handbook of Small Data Sets. Chapman & Hall, p. 69.

carry out a comparing regressions analysis with Insul as the group variable, Gas as outcome, and Temp as within-group predictor.
construct a 95% confidence interval for the effect of insul on on gas with temp = 4 (pick-a-point procedure)
for what values of temp does there appear to be an effect of Insul on Gas (simultaneous region of significance)
Solution for question 3

Question 4. R packages `interactions` and `probemod`
In lecture there was short mention of these two R-packages that whose main functions are to carry out the pick-a-point and Johnson-Neyman claculations, which are developed in Rogosa(1980). Try out these functions using the cnrl dataset (also from Rogosa,1980) which we worked out in the lecture materials.
Solutions spoiler alert: no joy from these packages.
Solution for question 4

Week 3

[redacted]
Lecture slides, week 3 (pdf)
week 3, part a (pdf)
week 3, part b (pdf)
Audio companion, week 3
parta partb

1. Compliance in RCT

Lecture topics
1. Compliance background: Intent-to-treat analyses, CACE estimators, research examples
2. Compliance and Dose-response data analysis (Efron-Feldman)
3. Rubin-Holland approach via Booil Jo presentation: Potential Outcomes Approach: A Brief Introduction
Class handouts: Compliance examples Compliance overview Compliance math notes Little-Rubin Ann Rev Pub Health formulation

Primary Readings
Compliance Background: Intent-to-Treat (ITT), the FDA mandate. simple definitions: wiki Encyclopedia of epidemiology, Volume 1 (google books)
Potential outcomes formulation (CACE): Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes: Concepts and Analytical Approaches Roderick J. Little and and Donald B. Rubin Vol. Annual Review of Public Health, 21: 121-145, May 2000.
Epidemiology exposition: An introduction to instrumental variables for epidemiologists, Sander Greenland, *International Journal of Epidemiology* 2000;29:722-729

Compliance research examples.
Clofibrate in Coronary Drug Project
Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New England Journal of Medicine* Volume 303:1038-1041 October 30, 1980 Number 18
Vitamin A in Central America
An introduction to instrumental variables for epidemiologists, Sander Greenland, *International Journal of Epidemiology* 2000;29:722-729
Cholestyramine in Cholesterol trial (measured compliance)
Compliance as an Explanatory Variable in Clinical Trials. B. Efron; D. Feldman *Journal of the American Statistical Association*, Vol. 86, No. 413. (Mar., 1991), pp. 9-17. Jstor

Draft Lottery and Vietnam Service
Joshua D. Angrist; Guido W. Imbens; Donald B. Rubin "Identification of Causal Effects Using Instrumental Variables" *Journal of the American Statistical Association*, Vol. 91, No. 434. (Jun., 1996), pp. 444-455. JStor

Additional resources
Compliance as an Explanatory Variable in Clinical Trials. B. Efron; D. Feldman *Journal of the American Statistical Association*, Vol. 86, No. 413. (Mar., 1991), pp. 9-17. Jstor
David Freedman on Compliance Adjustments: Statistical Models for Causation: What Inferential Leverage Do They Provide? Evaluation Review 2006; 30: 691-713. On regression adjustments to experimental data Advances in Applied Mathematics vol. 40 (2008) pp. 180-93.
Intent-to-treat Analysis of Randomized Clinical Trials Michael P. LaValley Boston University ACR/ARHP Annual Scientific Meeting Orlando 10/27/2003
Intention to treat--who should use ITT? J. A. Lewis and D. Machin Br J Cancer. 1993 October; 68(4): 647-650.
Compliance analyses, R-implementations: Imai `experiment` package Package icsw, Inverse Compliance Score Weighting
What is meant by intention to treat analysis? Survey of published randomised controlled trials Sally Hollis and Fiona Campbell *British Medical Journal* 1999;319;670-674
Booil Jo, Dept of Psychiatry Estimation of Intervention Effects with Noncompliance *Journal of Educational and Behavioral Statistics*
Compliance Publications based on Neyman-Rubin causal models:
Direct and Indirect Causal Effects via Potential Outcomes Donald B. Rubin *Scandinavian Journal of Statistics* Volume 31, Issue 2, Page 161-170, Jun 2004 .
Imbens GW and Rubin DB (1997) Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance The Annals of Statistics, 25, 305-327.
Principal Stratification in Causal Inference Constantine E. Frangakis and Donald B. Rubin, *Biometrics*, 2002, 58, 21- 29.
Addressing Complications of Intention-to-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncompliance and Subsequent Missing Outcomes. Constantine E. Frangakis; Donald B. Rubin *Biometrika*, Vol. 86, No. 2. (Jun., 1999), pp. 365-379. Jstor link
Additional Case Studies
Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City Barnard, Frangakis, Hill, and Rubin *Journal of the American Statistical Association* June 2003, Vol. 98, No. 462, Applications and Case Studies
The British Journal of Psychiatry (2003) 183: 323-331 Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up graham dunn, and mohammad maracy

2. Regression Discontinuity Designs (systematic assignment)

Lecture Topics
Non-random assignment on the basis of the covariate, such as regression discontinuity designs.
Regression Discontinuity handout Example from rdd manual ascii version

Primary Readings
Regression Discontinuity Designs Useful primers by Wm Trochin: William Trochim's Knowledge Base
Rubin, D. B., (1977), "Assignment to a Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 2, 1-26. Jstor link

Regression Discontinuity Research Examples
The original: PSAT and National Merit
Thistlewaite, D., and D. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," Journal of Educational Psychology, 51, 309-317.
Class size, Maimonides' Rule
In Rosenbaum, *Design of Observational Studies* (linked on main page). sections 1.3, 3.2, 5.2.3, 5.3 DOS text
Angrist-Lavy Maimondes (class size) data Angrist and Lavy, 1999. read data `ang = read.dta("http://stats.idre.ucla.edu/stat/stata/examples/methods_matter/chapter9/angrist.dta")`

R implementations and Resources
R-package--rdd; Regression Discontinuity Estimation Author Drew Dimmery
Also Package `rdrobust` Title Robust data-driven statistical inference in Regression-Discontinuity designs
RJournal for rdrobust, rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs

Additional Resources: Regression Discontinuity Designs
Journal of Econometrics (special issue) Volume 142, Issue 2, February 2008, The regression discontinuity design: Theory and applications Regression discontinuity designs: A guide to practice, Guido W. Imbens, Thomas Lemieux
Also from Journal of Econometrics (special issue) Volume 142, Issue 2, February 2008, The regression discontinuity design: Theory and applications Waiting for Life to Arrive: A history of the regression-discontinuity design in Psychology, Statistics and Economics, Thomas D Cook
the original paper: Thistlewaite, D., and D. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," Journal of Educational Psychology, 51, 309-317.
Trochim W.M. & Cappelleri J.C. (1992). "Cutoff assignment strategies for enhancing randomized clinical trials." Controlled Clinical Trials, 13, 190-212. pubmed link
Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program Richard A. Berk; David Rauma *Journal of the American Statistical Association*, Vol. 78, No. 381. (Mar., 1983), pp. 21-27. Jstor
Berk, R.A. & de Leeuw, J. (1999). "An evaluation of California's inmate classification system using a generalized regression discontinuity design." *Journal of the American Statistical Association*, 94(448), 1045-1052. Jstor
another econometric treatment

Week 3 Review Questions

Regression Discontinuity

Question 1. Regression Discontinuity, classic "Sharp" design.

Replicate the package rdd toy example: cutpoint = 0, sharp design, with treatment effect of 3 units (instead of 10). Try out the analysis of covariance (Rubin 1977) estimate and compare with rdd output and plot. Pick off the observations used in the Half-BW estimate and verify using t-test or wilcoxon.
Extra: try out also the `rdrobust` package for this sharp design.

Solution for Review Question 1

Question 2. Systematic Assignment, "fuzzy design". Probabilistic assignment on the basis of the covariate.

i. Create artificial data with the following specification. 10,000 observations; premeasure (Y_uc in my session) gaussian mean 10 variance 1. Effect of intervention (rho) if in the treatment group is 2 (or close to 2) and uncorrelated with Y_uc. Probability of being in the treatment group depends on Y_uc but is not a deterministic step-function ("sharp design"): `Pr(treatment|Y_uc) = pnorm(Y_uc, 10,1)` . Plot that function.
ii. Try out analysis of covariance with Y_uc as covariate. Obtain a confidence interval for the effect of the treatment.
iii. Try out the fancy econometric estimators (using finite support) as in the rdd package. See if you find that they work poorly in this very basic fuzzy design example.
Extra: try out also the `rdrobust` package for this fuzzy design.

Solution for Review Question 2

Question 3. Controlled Assignment (class example)

From Rubin, D. B., (1977), "Assignment to a Treatment Group on the Basis of a Covariate", linked on course page

From page 16 Rubin

7. A SIMPLE EXAMPLE

Table I presents the raw data from an evaluation of a computer-aided program designed to teach mathematics to children in fourth grade. There were 25 children in Program 1 (the computer-aided program) and 47 children in Program 2 (the regular program). All children took a Pretest and Posttest, each test consisting of 20 problems, a child's score being the number of problems correctly solved. These data will be used to illustrate the estimation methods discussed in Sections 4, 5, and 6. We do not attempt a complete statistical analysis nor do we question the assumption of no interference between units.

TABLE I

Raw Data for 25 Program 1 Children and 47 Program 2 Children

Pretest Scores	Posttest Scores	
	Program 1	Program 2
10	15	6,7
9	16	7,11,12
8	12	5,6,9,12
7	8,11,12	6,6,6,6,7,8
6	9,10,11,13,20	5,5,6,6,6,6,6,6,8,8,8,9,10
5	5,6,7,16	3,5,5,6,6,7,8
4	5,6,6,12	4,4,4,5,7,11
3	4,7,8,9,12	0,5,7
2	4	4
1	-	-
0	-	7

Does assignment appear to be random or is this appear to be Assignment on the Basis of Pretest?
Try to estimate the asignment rule, presuming it is based on pretest How does this differ from a regression discontinuity design (simplest version)?
Assuming that assignment to Program 1 or Program 2 was solely on the basis of pretest (plus perhaps a probabilistic component) estimate the effect of program (new vs regular).
note data in table 1 exist in a more convenient form in file hw5rubin.dat http://statweb.stanford.edu/~rag/stat209/hw5rubin.dat and data file included in the solutions

Solution for Review Question 3

Compliance in RCT

Question 4 Non-compliance. Class example week 3.

Adapted from (linked on class page): An introduction to instrumental variables for epidemiologists, Sander Greenland, International Journal of Epidemiology 2000;29:722-729
Additional Reference: Sommer and Zeger (1991). On Estimating Efficacy from Clinical Trials. Statistics in Medicine

Greenland discusses randomized trials with non-compliance where Z indicates treatment assignment, which is randomized; X indicates treatment received, which is affected but not fully determined by assignment Z.

To illustrate Greenland presents in his Table 1 individual one- year mortality data from a cluster-randomized trial of vitamin A supplementation in childhood. Of 450 villages, 229 were assigned to a treatment in which village children received two oral doses of vitamin A; children in the 221 control villages were assigned none. This protocol resulted in 12,094 children assigned to the treatment (Z = 1) and 11,588 assigned to the control (Z = 0). Only children assigned to treatment received the treatment; that is, no one had Z = 0 and X = 1. Unfortunately, 2419 (20%) of those assigned to the treatment did not receive the treatment (had Z = 1 and X = 0), resulting in only 9675 receiving treatment (X = 1). Class handout has depiction and Greenland's table of results. Use as the outcome measure Y, the Deaths per 100,000 within one year (labeled Risk in Greenland's Table 1).

Part 1, using data summary from class handout
a. Give the ITT (intent-to-treat) estimate of the effect of vitamin A on Risk
b. What is the compliance rate in the treatment group (Z=1)? In the control group (Z=0)?
c. What is the instrumental variables estimate (following Angrist Imbens Rubin) of the effect of vitamin A on Risk?
What interpretation is given to this estimate (c.f. Booil Jo presentation)? Compare with part (a) result and comment.

Don Rubin has a great overview talk For Objective Causal Inference, Design Trumps Analysis Don Rubin, posted at http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/trumps.pdf
Starting pdf page 21 Rubin takes up noncompliance using the Viamin A data (slightly different tabulated values than in the Greenland paper handout)
d. Recreate the calculations (ITT As-treated, Per Protocol) shown on pdf p.23; refer to Booil Jo handout
e. also CACE estimate pdf p.24
The Bayesian estimates (Imbens and Rubin 1997) pdf page 25 onward are implimented in part in the `experiment` package (Imai) mentioned in class and class materials.

Solution for question 4

Question 5
From the Booil Jo presentation slides in lecture, consider the JHU PIRC Intervention Study: N=284
Estimate Intervention Effects With Noncompliance
The Johns Hopkins Public School Preventive Intervention Study was conducted by the Johns Hopkins University Preventive Intervention Research Center (JHU PIRC) in 1993-1994 (Ialongo et al., 1999~ The study was designed to improve academic achievement and to reduce early behavioral problems of school children. Teachers and first-grade children were randomly assigned to intervention conditions. The control condition and the Family-School Partnership Intervention condition are compared in this example. In the intervention condition, parents were asked to implement 66 take-home activities related to literacy and mathematics over a six-month period. One of the major outcome measures in the JHU PIRC preventive trial was the TOCA-R (Teacher Observation of Classroom Adaptation)
• Completed at least 45 activities = compliers.
• Outcome: change score (baseline - followup) of anti-social behavior .
From the means and compliance data given in the class materials (also linked Booil talk) compute treatment effect estimate of change in anti-social behavior: give ITT estimate and CACE estimate

Solution for question 5

Question 6 Broken RCT: Compliance, measured or binary

Compliance as a measured variable. In Stat209 week 3 we examine compliance adjustments; both those based on a dichotomous compliance variable and the much much more common measured compliance (often unwisely dichotomized to match Rubin formulation). The Efron-Feldman study (handout description) used a continuous compliance measure. An artificial data set a data frame containing Compliance, Group, and Outcome for Stat209 is constructed so that ITT for cholesterol reduction is about 20 (compliance .6) and effect of cholestyramine for perfect compliance is about 35.
Try out some IV estimators for CACE. Obtain ITT estimate of group (treatment) effect with a confidence interval. Try using G as an instrument for the Y ~ comp regression. What does that produce?
Alternatively use the Rubin formulation with a dichotomous compliance indicator defined as TRUE for compliance > .8 in these data. What is your CACE estimate. What assumptions did you make? Compare with ITT estimate. In this problem the `ivreg` function from `AER` package is used for IV estimation.

Solution for Review Question 6

More Question 6 1. Compliance data, IV analysis, imitating Efron-Feldman cholestyramine trial. Solution showed you the widely used `ivreg` function from package `AER` package. Redo the ivreg analyses using functions from the `ivmodel` package.

Solution for more Review Question 6

Week 4

THIS PAGE LEFT BLANK

FOR PAGINATION

Randomized Experiments with Noncompliance

David Madigan

Introduction

- "Noncompliance" is an important problem in randomized experiments involving humans

- Includes e.g. switching subjects to standard therapy when experimental therapy fails

- "Intent-to-treat" (ITT) is a standard approach and is endorsed by FDA, journals, etc.

- Analyzes the data "as-randomized"

ITT

"Analyses that include all randomized patients in the groups to which they were randomly assigned, regardless of their compliance with the entry criteria, regardless of the treatment they actually received, and regardless of subsequent withdrawal from treatment or deviation from the protocol."

Workgroup for the Biopharmaceutical Section of the American Statistical Association

Intention-to-Treat Analysis

- Key points
  - Use every subject who was randomized according to randomized treatment assignment
  - Ignore noncompliance, protocol deviations, withdrawal, and anything that happens after randomization
- ***As randomized, so analyzed***

Pragmatic vs. Explanatory Analyses

- The hypothesis that an ITT analysis addresses is pragmatic – the effectiveness of therapy when used in autonomous individuals
- Analyses that focus on the biologic effects of therapy are addressing explanatory hypotheses
  - This is often done by excluding noncompliant subjects from analysis

Summary

- Randomization is of central importance in clinical trials
- ITT analyses try to preserve the randomized groups and address pragmatic hypotheses about the clinical utility of treatment
- Explanatory analyses address interesting hypotheses about the biological effect of treatment, but are more prone to bias

Compliance with Treatment

- Some subjects do not comply with their assigned treatment
- For explanatory analyses these subjects might not be used
  - No biologic effect if no treatment taken
- For ITT analysis they would be used
  - Why?

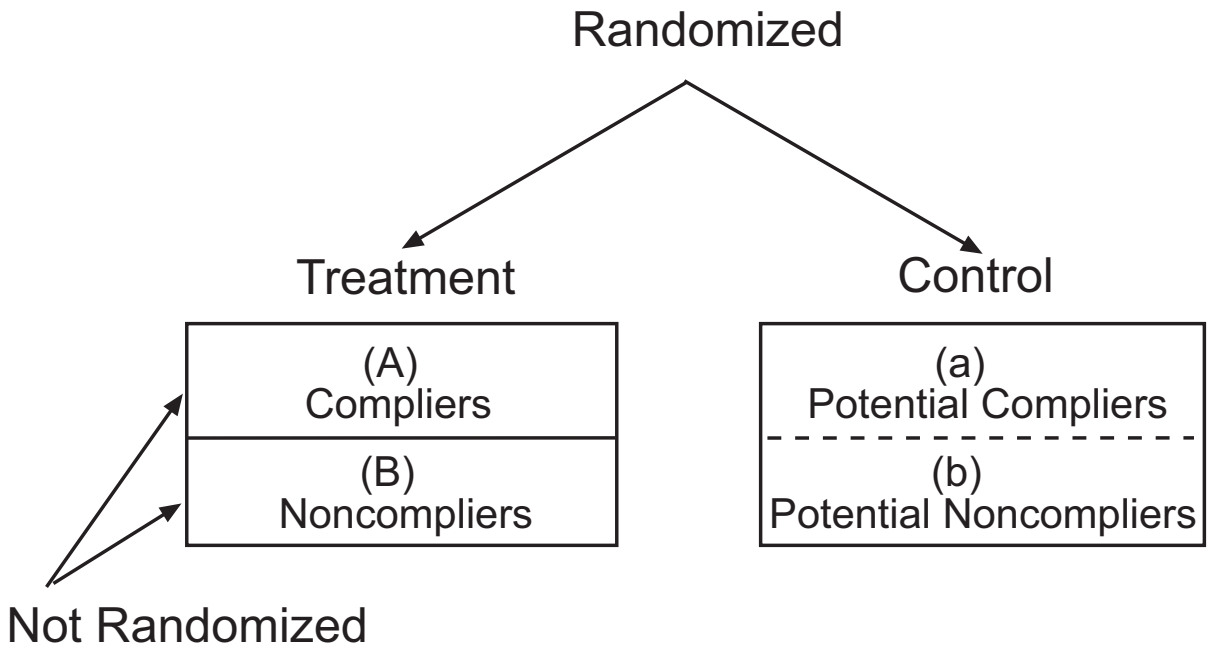
Compliance with Treatment

- Why include noncompliant subjects in ITT analysis? Other considerations are
  - In clinical practice, some patients are not fully compliant
  - Compliant subjects usually have better outcomes than noncompliant subjects, regardless of treatment

from Booil Jo

Note: Compliance is Yes/No

Options in Dealing with Noncompliance



- Intent-to-Treat (ITT) Analysis: (A+B) vs. (a+b)
- As-Treated Analysis: (A) vs. (B+a+b)
- Per-Protocol Analysis: (A) vs. (a+b)
- CACE (Complier Average Causal Effect): (A) vs. (a)

CACE—Complier Average Causal Effect

The **Complier Average Causal Effect**
(CACE) is the average causal effect of a treatment in the subpopulation of compliers (see R.J. Little & D.B. Rubin, "Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytic approaches," *Annual Review of Public Health*, 2000, 21: 121-145).

The compliance status of an individual is generally unknown, but can be estimated with a number of assumptions, detailed in Little and Rubin, above. An approximately unbiased edstimate of the of the CACE is the estimate of the Intent to Treat effect divided by the difference in the proportions that adopt the new treatment in the new treatment and control groups.

IV estimate

new vocab, for compliance

Instrumental Variable Methods

STAT 266 week 9

Observational Studies

omitted variables: Fixing Broken Regressions

$Y = \beta_0 + \beta_1 D + u$ Dose-response, (returns to schooling) Y = lwage D = educ but $Cov(D,u) \neq 0$

$Y = \beta_0 + \beta_1 G + u$ Group membership effects, t-test but $Cov(G,u) \neq 0$

"Broken" → D, G correlated with omitted variables in u

OLS fails for $Y = \beta_0 + \beta_1 X + u$ when $cov(X,u) \neq 0$

e.g. $\log(wage) = \beta_0 + \beta_1 X + u$ (ability omit, Angrist)

To the rescue? instrument Z such that $Cov(X,Z) \neq 0$

AND $cov(Z,u) = 0$ empirical assoc. strong, weak insb

cue Dusty S. hope, untestable

if true ancova also works

Z "exogenous" no partial effect on Y (even if Z random ass in RCT)

Properties: $Y = \beta_0 + \beta_1 X + u \Rightarrow$

$Cov(Z,Y) = \beta_1 cov(Z,X) + cov(Z,u)$ → 0 by wish upon a star

thus $\beta_1^{IV} = cov(Z,Y)/cov(Z,X)$ (Z replaces X)

$\hat{\beta}_1^{IV} = \frac{S_{YZ}}{S_{XZ}}$ $\hat{\beta}_0^{IV} = \bar{Y} - \hat{\beta}_1^{IV}\bar{X}$, $\widehat{Var}(\hat{\beta}_1^{IV}) = \frac{\hat{\sigma}^2}{SSX \cdot r_{XZ}^2}$ weak instrument?

($r_{XZ} = .1 \Rightarrow n/100$)

RCT IV Random Assignment $G = 1, 0$

Encouragement Designs

Clever, innovative designs for estimating Dose-Response

Dose, binary or measured

$G = 1, 0$ encourage or not (RCT)

D: self-selected dose

Y outcome

IV assume assignment G no effect on Y

$\hat{\beta}_{IV} = \frac{S_{YG}}{S_{DG}} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0}$ wald estimator

salt, sesame st. in session

Compliance Adjustments

Desperate attempt to adjust, salvage Broken protocols in RCT

Compliance seldom binary

Binary Compliance c n (comply not)

no crossover $\pi_c = P(T|G=1)$

ITT $\mu_1 - \mu_0$

CACE $\mu_{c1} - \mu_{c0} = \frac{ITT}{\pi_c}$ iff

$\mu_{n1} - \mu_{n0} = 0$ ER, IV assumption

$\mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0}) + (1-\pi_c)(\mu_{n1} - \mu_{n0}) \to 0$

examples in session

from Booil Jo slides

Common Setting

- Randomized trials, where successful placebo control is unlikely.
- 2 conditions: intervention ($Z = 1$) and control ($Z = 0$)
- 2 compliance types (C_i)

1) complier (c) - receives the intervention treatment if assigned, and does not if not assigned. π_c = compliance rate.

2) noncomplier (n) - does not receive the intervention treatment even if assigned to receive it. $1 - \pi_c = \pi_n$ = noncompliance rate.

- 2 observed average outcomes in $Z = 1$: μ_{c1} and μ_{n1}.
- 2 unobserved average outcomes in $Z = 0$: μ_{c0} and μ_{n0}.
- 1 observed outcome in $Z = 0$: μ_0 $(= \pi_c\,\mu_{c0} + \pi_n\,\mu_{n0})$.
- The estimators of interest are

$$ITT = \mu_1 - \mu_0 = \pi_c\,(\mu_{c1} - \mu_{c0}) + (1 - \pi_c)\,(\mu_{n1} - \mu_{n0}).$$

$$CACE = ITT_c = \mu_{c1} - \mu_{c0}.$$

Single class over

CACE complier average causal effect

If $\mu_{n1} = \mu_{n0}$ CACE = ITT / π_c

IV style assumpt: assignment has no effect on outcome except through medication taken

Week 7

Stat 209

Formulation for Compliance Analyses

Bool Jo version

assume no controls get treatment

IAP: single cross-over Vitamin A

$Z = 1, 0$ T, C $\quad \pi_c = P(T|Z=1)$ compliance

T μ_1, (A) μ_{c1} (B) μ_{n1} $\quad \mu_1 = \pi_c \mu_{c1} + (1-\pi_c)\mu_{n1}$

C μ_0, (a) μ_{c0} (b) μ_{n0} $\quad \mu_0 = \pi_c \mu_{c0} + (1-\pi_c)\mu_{n0}$ (unobserved)

$$ITT = \mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0}) + (1-\pi_c)(\mu_{n1} - \mu_{n0})$$

$$CACE = \mu_{c1} - \mu_{c0} = \frac{\mu_1 - \mu_0}{\pi_c}$$

iff $\mu_{n1} = \mu_{n0}$ no effect of assignment

JHU ex: $\frac{\hat{ITT}}{\hat{\pi}_c} = \frac{.364}{.457} \approx .76$ IV est, AIR c.f AIR handout 2/19

David Freedman Analysis (modelobs talk, OXcauses paper)

Neyman model

Potential outcomes: overall in population

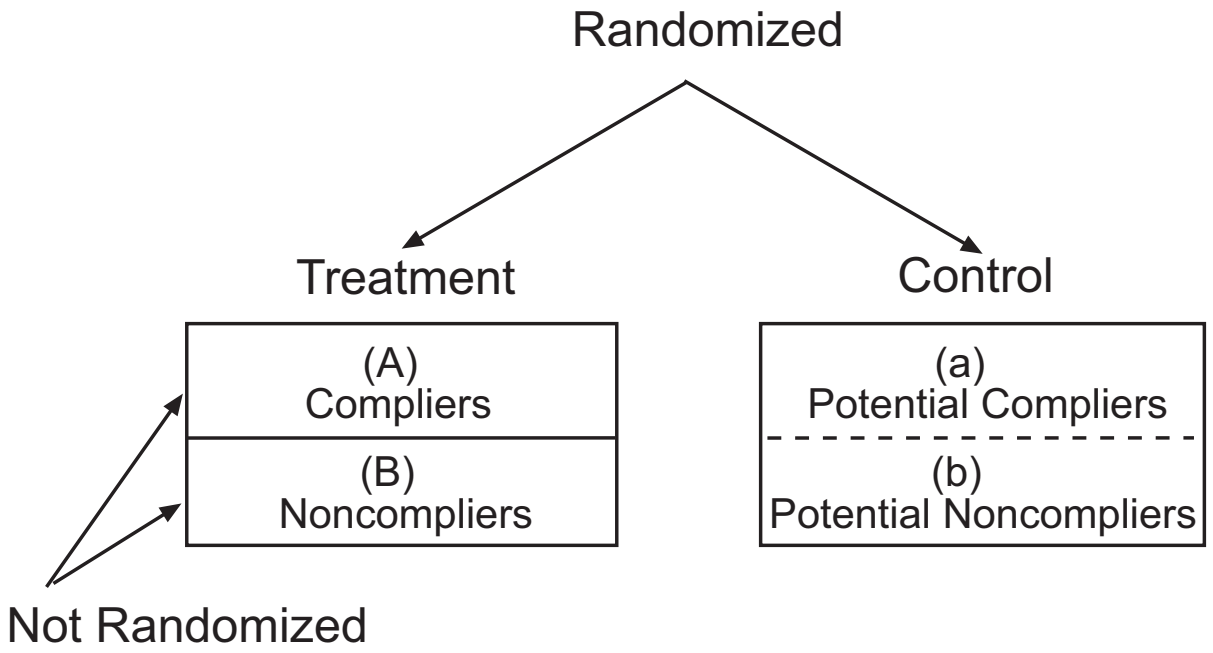
$\bar{T} = Ave(T_i)$ $\bar{C} = Ave(C_i)$

randomize to T, C

	Group	Number	T	C
(Lisa)	Always Treat	αN	a	a
(Marge)	Compliers	βN	τ	c
(Homer)	Never-Treat	γN	η	η
(Bart)	Defiers	δN	τ_D	c_D

IV estimate (pp. 704-5)

Options in Dealing with Noncompliance



- Intent-to-Treat (ITT) Analysis: (A+B) vs. (a+b)
- As-Treated Analysis: (A) vs. (B+a+b)
- Per-Protocol Analysis: (A) vs. (a+b)
- CACE (Complier Average Causal Effect): (A) vs. (a)

Per Protocol

- Estimates the T effect from a group stripped of poorer prognosis patients

- Upward bias in estimation of the T effect

"What are the differences between average T outcomes for patients who choose to adhere to recommended treatment T and outcomes for patients who choose to adhere to recommended treatment C?"

As-Treated

- Assigns the non-compliers to C
- Strips T of poor prognosis patients
- Upward bias in estimation of the T effect

"What are the differences between average outcomes for patients who take T as compared to those who take C, where the C group contains more patients with poor prognosis?"

Two Achilles' Heels

Internal vs external validity: the study population may not be representative.

A threat to internal validity is crossover: some people assigned to treatment decline treatment, some controls insist on treatment.

The intention-to-treat principle is a response to the crossover problem: you measure the effect of assignment, not treatment

Other estimators

(i) per protocol, (ii) treatment received, (iii) IV to estimate effect of treatment

Summary on the other estimators

Per protocol & treatment received. Unless you have very good blinding, these are very bad options.

The IV estimator. Pretty good—if you have a 0–1 response, single crossover, no blocking. With multi-level response, double crossover, or blocking, it's a lot less clear what's being estimated.

"Blocking" means, randomize subjects within (small) strata. It's the least of the issues here.

Calibrate using the Neyman model

Some would say, the Rubin model, but this mistakes the history.

D Dabrowska and TP Speed (1990). On the application of probability theory to agricultural experiments. Essay on principles. English translation of Neyman (1923). *Statistical Science*, 5: 463–80 (with discussion).

Index subjects by i running from 1 to N. If subject i is assigned to treatment, the response is T_i; if assigned to control, the response is C_i. If all subjects are assigned to treatment, the average response is

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i.$$

If all are assigned to control, the average response is

$$\overline{C} = \frac{1}{N}\sum_{i=1}^{N} C_i.$$

The intention-to-treat parameter is $\overline{T} - \overline{C}$. The mean in the treatment group minus the mean in the control group is an unbiased estimate: this is a theorem, not a tautology.

Let's say (i) open-label trial (everybody knows treatment status), (ii) response is 0–1 and so is compliance, (iii) response is to treatment not assignment, (iv) randomize some subjects to T = treatment, rest to C = control.

Group	No.	Ave. response if assigned to T	Ave. response if assigned to C
Always-treat	αN	A	A
Compliers	βN	T	C
Never-treat	γN	N	N
Defiers	θN	$\mathfrak{T}$	$\mathfrak{C}$

N is the number of subjects. The fractions $\alpha, \beta\ \gamma, \theta$ are parameters, constrained to be nonnegative, sum equals 1. The gothic (and very gothic) letters are parameters too. Not all identifiable.

Per-protocol estimand is

$$\frac{\alpha\mathrm{A} + \beta\mathrm{T}}{\alpha + \beta} - \frac{\beta\mathrm{C} + \gamma\mathrm{N}}{\beta + \gamma}$$

Treatment-received estimand is

$$\frac{\alpha\lambda\mathrm{A} + \beta\lambda\mathrm{T} + \alpha\mathrm{A} + \theta\mathfrak{C}}{\alpha\lambda + \beta\lambda + \alpha + \theta} - \frac{\beta\mathrm{C} + \gamma\mathrm{N} + \gamma\lambda\mathrm{N} + \theta\lambda\mathfrak{T}}{\beta + \gamma + \gamma\lambda + \theta\lambda}$$

Do these formulas look useless? Maybe that's because the estimators are useless. . . .

If there are no defiers, e.g., single crossover, IV estimand is T − C

Week 7 Math Notes

Stat 209

Formulation for Compliance Analyses

Bool Jo version — assume no controls get treatment

$Z = 1, 0$ T, C $\quad \pi_c = P(T|Z=1)$ compliance

DAP: single cross-over Greenland Vitamin A

T μ_1, (A) μ_{c1} (B) μ_{n1} $\quad \mu_1 = \pi_c \mu_{c1} + (1-\pi_c)\mu_{n1}$

C μ_0, (a) μ_{c0} (b) μ_{n0} $\quad \mu_0 = \pi_c \mu_{c0} + (1-\pi_c)\mu_{n0}$ (unobserved)

$ITT = \mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0}) + (1-\pi_c)(\mu_{n1} - \mu_{n0})$

$CACE = \mu_{c1} - \mu_{c0} = \frac{\mu_1 - \mu_0}{\pi_c}$ iff $\mu_{n1} = \mu_{n0}$ no effect of assignment

HV ex: $\frac{\hat{ITT}}{\hat{\pi}_c} = \frac{.364}{.457} = .76$ IV est, AIR c.f AIR handout 2/19

David Freedman Analysis (modelobs talk, OXcauses paper)

Neyman model — Potential outcomes: overall in population; $\bar{T} = Ave(T_i)$ $\bar{C} = Ave(C_i)$ randomize to T, C

ITT estimates $= \bar{T} - \bar{C}$

	Group	Number	T	C
(Lisa)	Always Treat	αN	a	a
(Marge)	Compliers	βN	τ	c
(Homer)	Never-Treat	γN	η	η
(Bart)	Defiers	θN	τ_D	c_D

Ave response it in

IV estimate (pp. 704-5) "oxcauser"

assume: single crossover, $\alpha = \theta = 0 \Rightarrow \beta + \gamma = 1$, no Lisas, Barts

Result $\hat{c} = (\bar{Y}^c - \gamma\hat{\eta})/\beta$, $\hat{\tau} = (\bar{Y}^T - \gamma\hat{\eta})/\beta \Rightarrow \hat{\tau} - \hat{c} = \frac{\bar{Y}^T - \bar{Y}^c}{\beta}$ IV

ave response control (sample)

CACE

IV226-7

per-protocol (A vs a+b)

$$\frac{\alpha a + \beta\tau}{\alpha+\beta} - \frac{\beta c + \gamma\eta}{\beta+\gamma}$$

if trial blind may work

As-treated (A vs a+b+B) $\quad \lambda = N_T/N_C$

$$\frac{\alpha\lambda a + \beta\lambda\tau + \alpha a + \theta c_D}{\alpha\lambda + \beta\lambda + \alpha + \theta} - \frac{\beta c + \gamma\eta + \gamma\lambda\eta + \theta\lambda\tau_D}{\beta + \gamma + \gamma\lambda + \theta\lambda}$$

Freedman p. 706.

See scelo "Statistical models for causation"

Evaluation Review see 10

therefore a step forward.

10. TECHNICAL NOTES

Intention-to-treat

The intention-to-treat estimator is the average response in the assigned-to-treatment group, minus the average response in the assigned-to-control group. The estimand is the average response of the study population if all were assigned to treatment, minus the average response if all were assigned to control.

To pursue these ideas, it will be convenient to introduce some mathematical notation. We index subjects by i, running from 1 to N. If subject i is assigned to treatment, the response is T_i; if assigned to control, the response is C_i. If all subjects in the experimental population are assigned to treatment, the average response is

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i.$$

If all are assigned to control, the average response is

$$\overline{C} = \frac{1}{N}\sum_{i=1}^{N} C_i.$$

The intention-to-treat parameter is $\overline{T} - \overline{C}$, which measures the average difference that assignment to treatment would make, in the study population. These quantities are all parameters: they are computed at the level of the population, not the data. (Remember, if you see the treatment response T_i, you don't see the control response C_i.)

The estimators are the obvious ones: $\overline{T}$ is estimated by the average response of the subjects assigned to treatment; $\overline{C}$ is estimated by the average response of the subjects assigned to control; and the difference between these two sample averages estimates the intention-to-treat parameter. The estimators are unbiased, even in finite samples, because the average of a random sample is an unbiased estimator for the average of the parent population.

The version of the model described above is deterministic at the level of individuals. If you assign i to treatment, the response is T_i; if you assign i to control, the response is C_i. But two different subjects i and j may well have different responses to treatment ($T_i \neq T_j$); they may also have different responses to the control regime ($C_i \neq C_j$). Moreover, the model

Johns Hopkins School Intervention Study

(Ialongo et al., 1999)

• Designed to improve academic achievement and to reduce early behavioral problems of school children. Teachers and first-grade children were randomly assigned to control/intervention conditions.

• In Family-School Partnership Intervention condition, parents were asked to implement 66 take-home activities related to literacy and math over a six-month period. Nothing was offered to control condition. **Therefore, compliance with the intervention treatments could not be observed among individuals assigned to the control condition.**

• A large variation in completed activities (ranges 0 to 66), and over-reporting of compliance is also expected. The intervention may not show any desirable effects unless parents report a quite high level of compliance.

• Categorizing individuals into low and high compliers will provide a more meaningful intervention effect estimate than categorizing them into never-takers and compliers (97% reported they completed at least one activity).

JHU PIRC Study: N=284 (listwise deletion)

- Completed at least 45 activities - compliers.
- Outcome: change score (baseline - followup) of anti-social behavior .

>0 good

$\hat{\mu}_0$	$\hat{\mu}_{c1}$	$\hat{\mu}_{n1}$	μ_1	$\hat{\pi}_c$
-0.319 (1.383)	-0.177 (1.214)	0.248 (1.271)	0.045 (1.259)	0.479

- The ITT estimate is

$$\widehat{ITT} = \hat{\mu}_1 - \hat{\mu}_0 = 0.045 - (-0.319) = 0.364.$$

  - From $\mu_0 = \pi_c\,\mu_{c0} + \pi_n\,\mu_{n0}$ and ER ($\hat{\mu}_{n0} = \hat{\mu}_{n1}$), $\hat{\mu}_{c0} = \frac{\hat{\mu}_0 - \hat{\mu}_{n1}\,(1-\hat{\pi}_c)}{\hat{\pi}_c}$.
  - From $\mu_1 = \pi_c\,\mu_{c1} + \pi_n\,\mu_{n1}$, we get $\hat{\mu}_{c1} = \frac{\hat{\mu}_1 - \hat{\mu}_{n1}\,(1-\hat{\pi}_c)}{\hat{\pi}_c}$.

- The $CACE$ estimate is

$$\widehat{CACE} = \hat{\mu}_{c1} - \hat{\mu}_{c0} = \frac{\hat{\mu}_1 - \hat{\mu}_0}{\hat{\pi}_c} = 0.364/0.457 = 0.760.$$

.76 = .364/.479 , comp .479 vs .457 handling of missing data

International Journal of Epidemiology 2000;29:722–729

An introduction to instrumental variables for epidemiologists

Sander Greenland

Instrumental-variable (IV) methods were invented over 70 years ago, but remain uncommon in epidemiology. Over the past decade or so, non-parametric versions of IV methods have appeared that connect IV methods to causal and measurement-error models important in epidemiological applications. This paper provides an introduction to those developments, illustrated by an application of IV methods to non-parametric adjustment for non-compliance in randomized trials.

Keywords Biometry, causal models, compliance, confounding, econometrics, epidemiological methods, instrumental variables, measurement error, misclassification, regression, regression calibration, statistics

Accepted 17 December 1999

Methods for control of confounding and measurement error are central to non-experimental research. One large class of such methods based on instrumental variables (IV) dates back to the 1920s. They have been an integral part of econometrics for decades[1,2] and have appeared in the health sciences,[3–5] yet they remain little known in epidemiology. Their absence from the field may in part be due to the fact that the methods were rarely presented outside of linear-regression contexts until the 1980s. The past two decades have seen extensions of IV methods to non-parametric causal models and to non-linear regression.[4–16] I here provide an elementary introduction to non-parametric IV methods, with a focus on showing how IV assumptions lead to corrections for confounding by non-compliance in randomized trials. This application is especially important because treatment assignment can provide a perfect instrumental variable for confounding control, and IV methods provide an alternative to intent-to-treat analysis. I will also briefly sketch how IV methods for misclassification correction are related to confounding control.

An intuitive basis for the methods discussed here is as follows: Suppose X and Y are the exposure and outcome of interest, and we can observe their relation to a third variable Z, called an *instrumental variable* or *instrument*, that is associated with X but not associated with Y except through its association with X. Then, under certain conditions, we can write the Z-Y association as a product of the Z-X and X-Y associations,

$$\text{Assoc}_{ZY} = \text{Assoc}_{ZX}\text{Assoc}_{XY}$$

and solve this equation for the X-Y association. This equation is of particular use when either (i) the observed X-Y association is confounded by unmeasured covariates, but the Z-X and Z-Y associations are not confounded; or (ii) the X-Y association cannot be observed directly because we cannot observe X directly, but Z is an observed surrogate for X whose association with X is known or estimable, and whose deviation from X is independent of other variables or errors. The precise conditions under which the equation holds vary with the problem, as will be discussed below.

Instrumental variables for confounding control

Let U be the set of all variables that affect X and Y, and suppose Z has the following properties:

1) Z is independent of U;
2) Z is associated with X;
3) Z is independent of Y given X and U.

Note that assumption 3 implies that Z has no direct effect on Y. Figure 1 gives a causal diagram[17] that satisfies these assumptions, with labels from the example below.

The variables in U may be partly or entirely unmeasured or even unimagined. It might then appear that there is no way to estimate the effect of X on Y in an unconfounded manner. A fundamental insight of IV estimation is that the instrument Z provides a means to estimate bounds on the X effect;[6,7,10] with further assumptions, the upper and lower bounds may be narrowed or even equal, in which case IV methods provide a point estimate.[8] This estimate is perhaps most easily understood in the following special case, which is based on a now standard counterfactual (potential-outcomes) model for treatment effects in the presence of non-compliance.[8,18,19]

Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095-1772, USA.

Example 2 from Greenland see handout

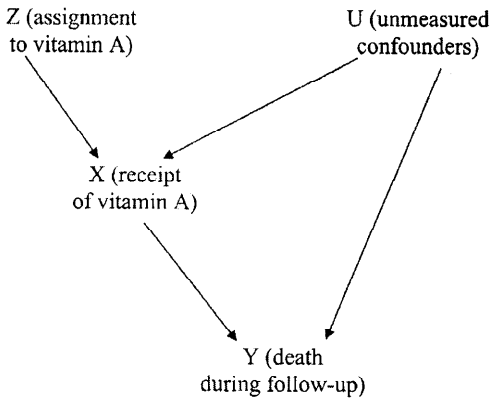


Figure 1

Intrumental variable methods for non-compliance

A paradigmatic example in which the IV conditions 1–3 are often satisfied is in a randomized trial with non-compliance: Z becomes treatment *assignment*, which is randomized and so fulfills assumption 1; X becomes treatment *received*, which is affected but not fully determined by assignment Z. To illustrate these concepts, Table 1 presents individual one-year mortality data from a cluster-randomized trial of vitamin A supplementation in childhood.[18,20] Of 450 villages, 229 were assigned to a treatment in which village children received two oral doses of vitamin A; children in the 221 control villages were assigned none. This protocol resulted in 12 094 children assigned to the treatment (Z = 1) and 11 588 assigned to the control (Z = 0). Only children assigned to treatment received the treatment; that is, no one had Z = 0 and X = 1. Unfortunately, 2419 (20%) of those assigned to the treatment did not receive the treatment (had Z = 1 and X = 0), resulting in only 9675 receiving treatment (X = 1). Nonetheless, assumption 1 is satisfied if the randomization was not subverted, while assumption 2 is supported by the data: Assignment to vitamin A increased the percentage receiving A from 0 to 80%.

Assumption 3 is plausible biologically, but must be reconciled with the fact that, among those both assigned to no vitamin A (Z = 0) and receiving no vitamin A (X = 0), mortality is only 639 per 100 000, versus 1406 per 100 000 for those assigned to vitamin A (Z = 1) but receiving no vitamin A (X = 0). Assuming that this difference is due to confounding by factors U that affect compliance (and hence X) and mortality (Figure 1), this illusory direct effect of assignment Z exemplifies the type of bias that arises when one attempts to estimate direct effects by stratifying on intermediates[21] (X is intermediate between Z and Y). There are many plausible explanations for such confounding. For example, perhaps families that fail to comply tend to be the poorest and so provide high-risk environments (poorer nutrition and sanitation); their low compliance would leave behind a low-risk group of compliers in the X = Z = 1 category, and thus confound an unadjusted comparison of the treated group (X = 1) with the untreated group (X = 0).

Confounding is a threat whenever people fail to comply with their assignment (i.e. have X ≠ Z) for reasons (U) related to their outcome;[4–10,18,19,22,23] this problem is often referred to as one of biased selection for treatment. For example, patients assigned to a complex pill regimen may become lax in following that regimen. These non-compliers are often those who feel less ill and who have a better prognosis with respect to the outcome Y. In such situations, there will be confounding in a comparison of those complying with treatment to the other patients, because those complying are sicker than the others (i.e. there is self-selection for treatment that is related to prognosis).

Concerns of this sort have led to recommendations (often rigid) that *intent-to-treat* analysis be followed. To test and estimate effects, intent-to-treat compares those *assigned* to one treatment against those *assigned* to another treatment without regard to actual treatment received (X). Critics of this approach point out that treatment received is the source of biological efficacy, and that comparison of treatment assigned is biased for the effect of treatment received (furthermore, the bias is not always toward the null,[24] contrary to common lore). By recognizing treatment assignment as an instrument, IV methods provide an alternative to the biased extremes of analysing Z as the treatment (intent-to-treat) and analysing received treatment X in the conventional manner (which is likely to be confounded by determinants of compliance).

To see how the IV concept can be used to control for confounding due to non-compliance, let us refer to people who would always obey their treatment assignment as *co-operative*; among these people, X and Z are always equal. It is crucial to distinguish the concept of *co-operative* people from the concept of *compliance*. Co-operative people are those who will receive their assigned regimen, *no matter which regimen (treatment) they are assigned*. In the example, co-operative children have parents or guardians who will fully co-operate with the researchers, in that they will allow the researchers to give their child the vitamin if assigned to receive it, and will not give their child the vitamin if assigned to not receive it. Non-co-operative people are those who will not receive certain regimens if assigned to them. In the example, some parents may refuse to let their child receive the experimental treatment. These refusers have children who exhibit non-compliance if they are assigned to the vitamin A, but who exhibit compliance if they

Table 1 One-year mortality data from cluster-randomized trial of vitamin A supplementation in children.[20] Z = 1 if assigned A, 0 if not; X = 1 if received A, 0 if not

	Z = 1			Z = 0	
	X = 1	X = 0	Total	X = 1	X = 0
Deaths (Y = 1)	12	34	46	0	74
Total	9675	2419	12 094	0	11 588
Risk[a]	124	1406	380	undefined	639

[a] Deaths per 100 000 within one year.

For *Objective* Causal Inference, Design Trumps Analysis

Donald B. Rubin

Harvard University

Sommer and Zeger Vitamin A Data

Row	True Compliance Type	Treatment Assignment	Treatment Received	Y_{obs}	Number of Children
1	?	0	0	0	11514
2	?	0	0	1	74
3	N	1	0	0	2385
4	N	1	0	1	34
5	C	1	1	0	9663
6	C	1	1	1	12
					23682

Reference: Sommer and Zeger (1991). On Estimating Efficacy from Clinical Trials. Statistics in Medicine.

Results of Three Standard MoM Analyses

Method	Estimate	Calculation	Row Comparison
ITT	-0.0026	$= \frac{12+34}{9663+2385+12+34} - \frac{74}{11514+74}$	3, 4, 5, & 6 vs. 1 & 2
As-treated	-0.0065	$= \frac{12}{9663+12} - \frac{34+74}{11514+2385+34+74}$	5 & 6 vs. 1, 2, 3, &4
Per protocol	-0.0052	$= \frac{12}{9663+12} - \frac{74}{11514+74}$	5 & 6 vs. 1 & 2

Reference: Sommer and Zeger (1991). On Estimating Efficacy from Clinical Trials. Statistics in Medicine.

MoM CACE Analysis

$ACE = p_N \cdot NACE + p_C \cdot CACE$

$-0.0025 = 0.2 \cdot NACE + 0.8 \cdot CACE$

$-0.0025 = 0.8 \cdot CACE \rightarrow CACE = -0.0025/0.8 = -0.0031$

Bayesian Analysis of Sommer & Zeger Data

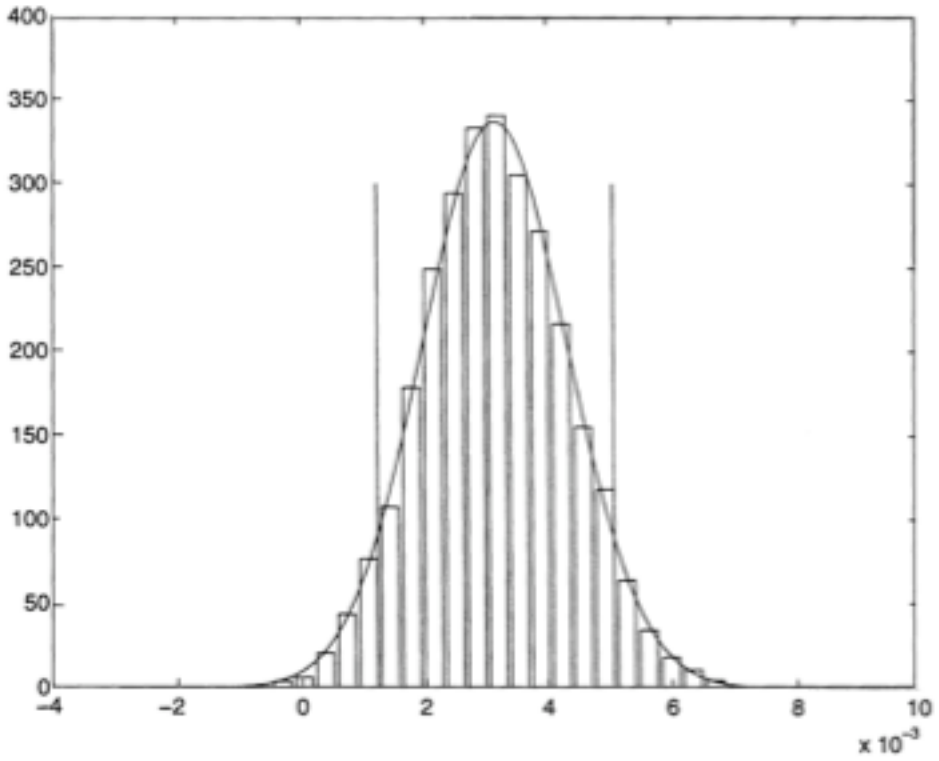


FIG. 3. *Histogram of CACE with exclusion restriction (data from Table 3).*

Imbens G.W. and Rubin D.B. (1997) Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. Annals of Statistics 25(1):305-327.

Bayesian Analysis of Sommer & Zeger Data, Marginal Posterior Distributions with and without Exclusion Restriction

Estimand	Exclusion restriction	Mean	Standard deviation	Median	5^{th} percentile	95^{th} percentile
CACE	No	3.1	2.5	3.2	-0.9	7.0
$ITT_Y^{(n)}$	No	0.5	10.1	0.2	-14.1	17.5
CACE	Yes	3.1	1.2	3.1	1.2	5.1

Imbens G.W. and Rubin D.B. (1997) Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. Annals of Statistics 25(1):305-327.

(2) Compliance -- The IV approach

Stat209

Greenland IJE 2000

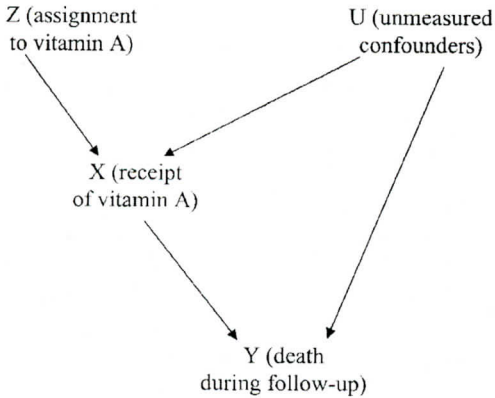


Figure 1

these concepts, Table 1 presents individual one-year mortality data from a cluster-randomized trial of vitamin A supplementation in childhood.[18,20] Of 450 villages, 229 were assigned to a treatment in which village children received two oral doses of vitamin A; children in the 221 control villages were assigned none. This protocol resulted in 12 094 children assigned to the treatment (Z = 1) and 11 588 assigned to the control (Z = 0). Only children assigned to treatment received the treatment; that is, no one had Z = 0 and X = 1. Unfortunately, 2419 (20%) of those assigned to the treatment did not receive the treatment (had Z = 1 and X = 0), resulting in only 9675 receiving treatment (X = 1). Nonetheless, assumption 1 is satisfied if the randomization was not subverted, while assumption 2 is supported by the data: Assignment to vitamin A increased the percentage receiving A from 0 to 80%.

Table 1 One-year mortality data from cluster-randomized trial of vitamin A supplementation in children.[20] Z = 1 if assigned A, 0 if not; X = 1 if received A, 0 if not

	Z = 1			Z = 0	
	X = 1	X = 0	Total	X = 1	X = 0
Deaths (Y = 1)	12	34	46	0	74
Total	9675	2419	12 094	0	11 588
Risk[a]	124	1406	380	undefined	639

[a] Deaths per 100 000 within one year

analysis in HW

Compliance adjustment

Identification of Causal Effects Using Instrumental Variables

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN

We outline a framework for causal inference in settings where assignment to a binary treatment is ignorable, but compliance with the assignment is not perfect so that the receipt of treatment is nonignorable. To address the problems associated with comparing subjects by the ignorable assignment—an "intention-to-treat analysis"—we make use of instrumental variables, which have long been used by economists in the context of regression models with constant treatment effects. We show that the instrumental variables (IV) estimand can be embedded within the Rubin Causal Model (RCM) and that under some simple and easily interpretable assumptions, the IV estimand is the average causal effect for a subgroup of units, the compliers. Without these assumptions, the IV estimand is simply the ratio of intention-to-treat causal estimands with no interpretation as an average causal effect. The advantages of embedding the IV approach in the RCM are that it clarifies the nature of critical assumptions needed for a causal interpretation, and moreover allows us to consider sensitivity of the results to deviations from key assumptions in a straightforward manner. We apply our analysis to estimate the effect of veteran status in the Vietnam era on mortality, using the lottery number that assigned priority for the draft as an instrument, and we use our results to investigate the sensitivity of the conclusions to critical assumptions.

KEY WORDS: Compliers; Intention-to-treat analysis; Local average treatment effect; Noncompliance; Nonignorable treatment assignment; Rubin-Causal-Model; Structural equation models.

1. INTRODUCTION

Economists are typically interested in estimating causal effects rather than mere associations between variables. Potentially interesting causal effects include the effects of education on employment and earnings, the effects of employment training programs on subsequent labor market histories, and the effects of a firm's inputs on its output. The dominant approach to making inferences about causal effects in economics over the last four decades is based on *structural equation models,* which rely on the specification of systems of equations with parameters and variables that attempt to capture behavioral relationships and specify the causal links between variables. Goldberger (1972) and Morgan (1990) provided historical perspectives on these models, which date back to Wright (1928, 1934) and Haavelmo (1943, 1944). Inference in structural equation models often exploits the presence of *instrumental variables* (IV). These are variables that are explicitly excluded from some equations and included in others, and therefore correlated with some outcomes only through their effect on other variables.

Rather than relying on structural equation models, causal inference in statistics, going back at least to work by Fisher (1918, 1925) and Neyman (1923) on agricultural experiments, is fundamentally based on the randomized experiment (see also Kempthorne 1952 and Cox 1958). The basic notion in this formulation, which has been extended by Rubin (1974, 1978) to more complicated situations, including observational studies without randomization, is that of *potential outcomes.* The causal effect of a treatment on a single individual or unit of observation is the comparison (e.g., difference) between the value of the outcome if the unit is treated and the value of the outcome if the unit is not treated. The target of estimation, the estimand, is typically the average causal effect, defined as the average difference between treated and untreated outcomes across all units in a population or in some subpopulation (e.g., males or females). For this definition of causality to be applicable to samples with units already exposed to treatments, we must be able to imagine observing outcomes on a unit in circumstances other than those to which the unit was actually exposed. This approach is now widely used in statistics and epidemiology (e.g., Efron and Feldman 1991 and Greenland and Robins 1986), where it is often referred to as the Rubin Causal Model (RCM; Holland [1986]).

In this article we provide a link between these approaches, capitalizing on the strengths of each. Earlier work combining elements of these approaches includes studies by Hearst, Newman, and Hulley (1986), Holland (1988), Permutt and Hebel (1989), Sommer and Zeger (1991), and Imbens and Angrist (1994). We show how the IV estimand can be given a precise and straightforward causal interpretation in the potential outcomes framework, despite nonignorability of treatment received. This interpretation avoids drawbacks of the standard structural equation framework, such as constant effects for all units, and delineates critical assumptions needed for a causal interpretation. The IV approach provides an alternative to a more conventional intention-to-treat analysis, which focuses solely on the average causal effect of assignment on the outcome (Lee, Ellenberg, Hirtz, and Nelson 1991).

As we show in the context of a specific application, our formulation of these assumptions makes it easier for researchers to judge whether or not a causal interpretation of the instrumental variables estimand is plausible. Standard

Joshua D. Angrist is Associate Professor, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139. Guido W. Imbens is Associate Professor, Department of Economics, and Donald B. Rubin is Professor, Department of Statistics, Harvard University, Cambridge, MA 02138. The authors thank Gary Chamberlain, Clifford Clogg, Sander Greenland, Judy Hellerstein, Paul Rosenbaum, the editor, an associate editor and several referees for comments, the National Science Foundation for financial support, and Tom Newman and Norman Hearst for sharing their data.

Week 6, IV application

JASA 91 (1996)

Identification of Causal Effects Using Instrumental Variables

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN

random assignment as IV

Y_i obs health outcome D_i veteran status (treatment)

effect: veteran status on health outcome (Vietnam era) p445

Z_i "assignment" draft status, lottery: $Z=1$ low lottery number

imperfect compliance: not all $Z=1$ became $D=1$

Z, D positive association not all $Z=0$ became $D=0$, volunteers

---

Y on D regression (t-test) $Y_i = \beta_0 + \beta_1 D_i + \varepsilon$

indicate effect of service on health (self-selection etc)

β_1 biased? eqs p. 445

use Z as instrument uncorr ε, ν

$$\hat{\beta}_1^{IV} = s_{yz}/s_{Dz} = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}}$$ p.447

Results p. 453

Born 1950

service $D=1$: If $Z=1$ then .353 $D=1$; $Z=0$ then .193 $D=1$ — Draft status (random) produced .159 increment

mortality $y=1$: If $Z=1$ .0204 died; $Z=0$ .0195 died — diff .0009; ACE draft on mortality

IV estimator $.0009/.159 = .0056$

causal effect of military service on mortality (population induced by draft)

AIR eqs

Vietnam era

Y = health outcome

D "treatment" serving in military

Z "assignment" draft status lottery (at random birthday)

compliance imperfect

low lottery # $\Rightarrow Z=1$ "drafted"

not all $D=1$ served

high lottery number $Z=0$ exempt from draft.

not all $D=0$ sat out. "volunteers"

Because of self-selection in part to D you D regression (t-test) can't indicate effect of service on health.

$Y_i = \beta_0 + \beta_1 D + \varepsilon$ β_1 biased?

cut-off of continuous endogenous

$D_i^* = \alpha_0 + \alpha_1 Z_i + v_i$ $D_i = \begin{cases} 1 \text{ if } D_i^* > 0 \\ 0 \text{ if } D_i^* \leq 0 \end{cases}$

Z uncorr w/ v, ε

$\hat{\beta}_1^{IV} = S_{YZ}/S_{DZ}$ p. 445

1953 cohort, 2 bins

OCTOBER						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

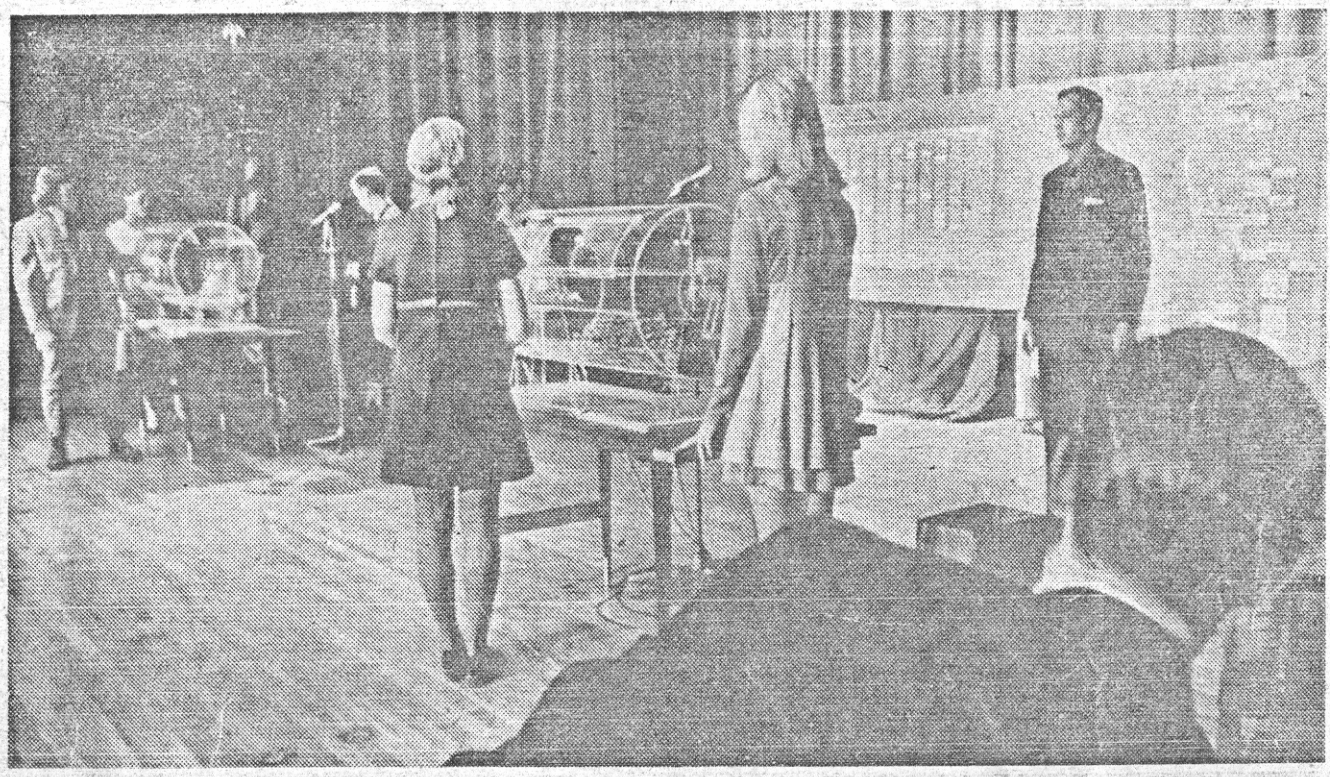
NOVEMBER						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

DECEMBER						
S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

VE Artist

ember
16—209
17—284
18—160
19—270
20—301
21—287
22—102
23—320
24—180
25—25
26—344
27—135
28—130
29—147
30—134

ember
17—294
18—13
19—168
20—149
21—80
22—188
23—252
24—155
25—6
26—351
27—194



[AP Wirephoto]

Draft Director Curtis W. Tarr leans to one side to get a view of board which carried results during Selective Service lottery in Washington.

Capsules containing birth dates and orders of induction were selected from plexiglass drums by student advisers. *(Story on page 1)*

12x3 table 1950 cohort, single basket

9	12	10
7	12	10
5	10	16
8	8	14
9	7	15
11	7	12
12	7	12
13	7	11
10	15	5
9	15	7
12	12	6
17	10	4

Compliance Formulation for AIR

Assignment Draft Lottery Z (Z=1 low lottery number)

people do what they want

D Vietnam Service D=1 serve

? imperfect compliance w/ lottery designation here both directions Z=0 volunteer for service Z=1 dodge, alternative

Y health outcome (death)

Compliance Family Simpsons
Homer nothing
Bart opposite
Marge comply
Lisa protocol, regardless

1950 results

effect of lottery on Viet service

$\hat{\beta}_{DZ} = .353 - .193 = .159$ (Marge − Lisa/Bart)

proportion compliers

effect lottery on death

$\hat{\beta}_{YZ} = .0204 - .0195 = .0009$

ITT draft status = served

consider Z as an instrument for Y on D regression (effect of Viet on health)

IV est $\frac{S_{YZ}}{S_{DZ}} = \hat{\beta}_{YZ}/\hat{\beta}_{DZ} = .0009/.159 = .0057$

CACE (AIR) effect of service on death.

IV est is adjusted for compliance

Little Robin
Ann Rev Pub Health pp132-5

STAT 209
Week 7

p.133 compliers defiers never takers always takers

C=1 compliers (marge)
C=0 all others

usually observed only for Treat.

invoke SUTVA
assignment to Group W, A(W) adopted treatment
Y(W) potential outcome,
unit level causal effect
treatment assignment on treat adopted A(2)-A(1)
" " on outcome Y(2)-Y(1)

ITT

CACE (subpopulation of compliers)
compliers have A(1)=1, A(2)=2
CACE = $E(Y(2)-Y(1) \mid C=1)$

p.134 Estimate CACE use SUTVA, ER and no defiers

p135 $E(Y(2)-Y(1)) = E(Y(2)-Y(1) \mid C=1)\,P(C=1) + E(Y(2)-Y(1) \mid C\neq 1)\,P(C\neq 1)$

assume the second line away

$E(Y(2)-Y(1) \mid C=1) = E(Y(2)-Y(1))/P(C=1)$ (CACE)

$\bar{Y}_2 - \bar{Y}_1$ est ITT, $p_2 - p_1$ estimates $P(C=1)$

p_2: in treat compliers + always takers
p_1: in control always takers

$\frac{\bar{Y}_2 - \bar{Y}_1}{p_2 - p_1}$ IV estimate

$p_2 - p_1 \leftarrow$ 0.1 if single crossover

be observed, particularly when K is large, these methods by necessity need to make relatively strong modeling assumptions to make inferences feasible. Nevertheless, if the assumptions are plausible, answers are likely to be superior to those obtained by methods that fail to take into account the sequential nature of the treatment allocation process.

When the probability of treatment assignment is allowed to depend on covariates or recorded outcomes and hence varies from unit to unit, the individual assignment probabilities $p_i \equiv \boldsymbol{P}(W_i \mid X_i, Y_{obs})$ are called propensity scores (54, 74). For Neymanian inference, they are required to be strictly between 0 and 1, and they are key quantities in the analysis. When the propensity scores are known, the assignment mechanism is essentially known, and simple generalizations of Fisherian and Neymanian modes of inference can be applied. In particular, Horvitz-Thompson estimation (27), in which observations are weighted by the inverse probabilities of their being observed, plays a key role for both modes of inference. An important point is that when there is little or no overlap in the propensity scores in the treatment groups, no causal inference is possible without strong external assumptions (60).

With Bayesian inference, an unconfounded assignment mechanism is ignorable (62), so that, after including the covariates that determine the assignment probabilities or an adequate summary (such as, possibly, the vector of propensity scores in the model), analysis in principle proceeds as in a classical randomized experiment. In this situation, however, there can be much greater sensitivity to model specification than in a classical randomized experiment. This sensitivity is the Bayesians' analog of the increased variance of the Horvitz-Thompson estimator in unbalanced designs and the decreasing power in such designs as they become more unbalanced.

THREATS TO UNCONFOUNDED TREATMENT ASSIGNMENT

Introduction

Randomized studies with full compliance and no missing data are relatively easily analyzed by the methods described above. However, in practice, inference for causal effects is often complicated by lack of compliance and incomplete data. The potential outcome formulation provides conceptual clarity and a set of tools for negotiating these complications. In the next section we consider the issue of noncompliance, and later we consider the issue of missing data.

Noncompliance

Analysis of a randomized study is complicated when subjects do not comply with their assigned treatment. An intention-to-treat (ITT) analysis compares the outcomes of subjects by randomization groups (an as-randomized analysis), ignoring the compliance information. The ITT effect measures the effect of treatment

randomization rather than the effect of treatment for those who actually received it. The ITT estimator is protected from selection bias by randomized treatment assignment, but in general is a distorted measure of the effect of the treatment itself. Alternatively, subjects can be classified by the treatments actually received (an "as-treated" analysis). The problem then is that randomization is violated, and confounding factors associated with switching from the assigned treatments potentially corrupt the causal interpretation of treatment effects.

The potential outcome definition of causal effects provides a useful basis for understanding noncompliance problems and assumptions implied by various estimation strategies (1, 28, 29). Participants are classified as one of four types, compliers, defiers, never-takers, and always-takers, in the context of this experiment. Compliers are people who would adopt whatever treatment is assigned, never-takers are people who would take the control treatment regardless of what they are assigned, always-takers would take the active treatment regardless of what they are assigned, and defiers are those who would adopt the opposite treatment to their assignment. Let C denote the compliance indicator of the participant, where $C = 1$ for compliers and $C = 0$ for all noncompliers: defiers, never-takers, and always-takers.

In practice, we observe compliance status only for the assigned treatment, so the full compliance status of subjects is incompletely observed. Specifically, if a subject is assigned to the new treatment and complies, then that subject may be a complier or an always-taker. If a subject is assigned to the new treatment and fails to comply, then that subject may be a never-taker or a defier. If a subject is assigned to the control and complies, then that subject may be a complier or a never-taker. If a subject is assigned to the control and obtains the new treatment, then that subject may be an always-taker or a defier. Imbens & Rubin (28) treat this partial information about compliance as a missing-data problem.

Let us make the SUTVA assumption, so that individual-level causal effects can be defined without reference to other individuals in the study. For a respondent assigned to group W, let $A(W)$ be the adopted treatment and let $Y(W)$ be the potential outcome. The unit-level causal effect of treatment assignment (W) on treatment adopted (A) is $A(2) - A(1)$, and the unit-level causal effect of treatment assignment (W) on outcome (Y) is $Y(2) - Y(1)$. These unit level effects of W on A and Y are generally not observable, because individuals cannot be assigned to both the experimental and control treatments. However, the average ITT effects on both A and Y can be readily estimated by the averages over groups of respondents in a randomized trial.

Another causal effect of interest in many studies of treatments is the Complier-Average Causal Effect (CACE), which is the average causal effect for the subpopulation of compliers, denoted by $C = 1$. For compliers, $A(1) = 1$ and $A(2) = 2$, and

$$\text{CACE} = E(Y(2) - Y(1) \mid C = 1) \qquad 5.$$

where the conditioning on $C = 1$ is identical to conditioning on $A_{(1)} = 1$ and $A(2) = 2$. The CACE is a valid causal effect because it is a summary measure

of individual-level effects in a subpopulation of interest, namely compliers. The CACE and the ITT effect capture two treatment features of potential interest—the effect of the treatment on those who can be induced to take it in this experiment and the overall impact of a treatment on the whole population, including noncompliers.

Two common methods of analysis, "as-treated" and "per-protocol" analysis, are generally flawed because they do not estimate the CACE or any other useful summary of individual-level causal effects (77). The as-treated analysis classifies subjects by the adopted treatment $A(W)$ and estimates

$$E(Y(W) \mid A(W) = 2) - E(Y(W) \mid A(W) = 1)$$

This is not an average of individual-level causal effects, because it compares averages of Y for groups with different characteristics. Specifically, it compares the average outcome of those adopting treatment 2 (namely, compliers assigned treatment 2, defiers assigned treatment 1, and always-takers assigned either treatment), with the average outcome of those adopting treatment 1 (namely, compliers assigned treatment 1, defiers assigned treatment 2, and never-takers assigned either treatment). A "per-protocol" analysis compares subjects who actually adopted their assigned treatments and estimates

$$E(Y(2) \mid A(2) = 2) - E(Y(1) \mid A(1) = 1)$$

which is also not an average of individual-level causal effects. It compares the outcomes of those who were assigned and adopted treatment 2 (i.e. compliers and always-takers assigned to 2) with the outcomes of those who were assigned and adopted treatment 1 (that is, compliers and never-takers assigned to 1). The potential outcome formulation of causal effects helps to clarify the deficiencies of these common analysis methods. The crux of the CACE formulation is the distinction between the definition of a "true" complier (compliance under *both* treatments) and an "observed" complier (compliance under the treatment actually assigned).

The CACE in Equation 5 is a valid causal estimand, but it is not immediately obvious how to estimate it because the compliance status C of individuals is generally unknown. However, the CACE can be estimated from the data under SUTVA and random assignment of W, if we make the following additional assumptions (1):

1. Exclusion restriction of treatment assignment given treatment received: for never-takers and always-takers, whose adopted treatment is the same regardless of which treatment is assigned, the outcome Y is the same regardless of which treatment is assigned; that is, $Y(1) = Y(2)$ if $A(1) = A(2)$;
2. Monotonicity of treatment assignment and treatment adopted (there are no defiers);
3. Nonzero denominator (the population includes some compliers).

To derive an estimate of the CACE under these assumptions, note that:

$$E(Y(2) - Y(1)) = E(Y(2) - Y(1) \mid C = 1)\, \boldsymbol{P}(C = 1) + E(Y(2) - Y(1) \mid C \neq 1)\, \boldsymbol{P}(C \neq 1) \quad 6.$$

Under assumption 2, the treatment effect for never-takers and for always-takers is identically zero. If, in addition, assumption 3 holds (there are no defiers), then the second term on the right side of Equation 6 is zero, and

$$E(Y(2) - Y(1) \mid C = 1) = E(Y(2) - Y(1))/\boldsymbol{P}(C = 1) \quad 7.$$

that is, the CACE is the ITT effect divided by the proportion of compliers. Now under randomized treatment allocation, the standard ITT estimator for the effect of W on Y, the difference in sample means $\bar{y}_2 - \bar{y}_1$ is an unbiased estimate of the numerator of Equation 7. Also, let p_2 be the proportion of subjects in the treatment group who adopt the new treatment, and let p_1 be the proportion of subjects in the control group who adopt the new treatment. Then p_2 is an unbiased estimate of the proportion of compliers or always-takers (those who adopt the treatment when assigned it), and p_1 is an unbiased estimate of the proportion of always-takers (those who adopt the treatment when assigned the control). Hence $p_2 - p_1$ is an unbiased estimate of the proportion of compliers, $\boldsymbol{P}(C = 1)$. Thus an approximately unbiased estimate of the CACE is

$$IVE = \frac{\bar{y}_2 - \bar{y}_1}{p_2 - p_1}, \quad 8.$$

which is the estimated ITT effect divided by difference in the proportions that adopt the new treatment in the new treatment and control groups. Assumption 3 assures that the denominator in Equation 8 has a nonzero expectation. Equation 8 is sometimes called the instrumental variable (IV) estimator. See Sommer & Zeger (81) and Ettner (13) for applications of this method. Our derivation of this estimator makes explicit assumptions that are hidden in other formulations (5). In particular, the exclusion restriction, assumption 1, plays a key role, and it is not a consequence of randomization of treatments.

Another advantage of the potential-outcome formulation is that it leads to more efficient estimators of the CACE than the IV estimator (Equation 8). Bayesian formulations for inference about the CACE treat the compliance indicator C as missing data (23, 28), and they make use of iterative simulation techniques such as data augmentation (83). The general topic of causal inference when there is noncompliance is currently an active area of research (2–4, 12, 17, 18, 23, 37).

Missing Data

Missing data (32–34, 65, 76) is a pervasive problem in epidemiological and clinical studies, particularly when they involve repeated measurements over time. Statistical analyses when there are missing values are too often confined to simple and

Week 7 - Intent to Treat and Non-Compliance

Stat 209

EXAMPLES

Coronary Drug Project

- Randomized, multi-center, double-blind, placebo-controlled trial of clofibrate for treatment for coronary heart disease
  - 1103 men on clofibrate
  - 2789 men on placebo
- ITT analysis of 5-year mortality on clofibrate was 20.0%, 20.9% on placebo (p=0.55)

no drug effect

But, in **placebo** subjects, mortality rates at 5 years were

- Compliant: 15.1%
- Non-compliant: 28.2%

- Subjects compliant with placebo had significantly lower mortality ($p < 0.0001$)!
- The explanatory analysis would miss this effect of compliance ?

- In **clofibrate** subjects, mortality rates at 5 years were
  - Compliant: 15.0%
  - Non-compliant: 24.6%
- Subjects compliant with clofibrate had significantly lower mortality (p=0.0001)!
- Explanatory analysis – compare compliant clofibrate subjects to subjects without adequate clofibrate intake (clofibrate noncompliers and placebo subjects) – **significant**!

maybe drug good?

Actual Practice of ITT

Survey of randomized controlled trials published in 1997 in BMJ, Lancet, JAMA, and NEJM (Hollis & Campbell)

Out of 249 trials, 119 (48%) explicitly stated that an ITT analysis was performed

- 15 (13%) clearly did not analyze as randomized
- 65 (55%) appeared to analyze as randomized, but without enough detail for the readers to verify
- No consistent method for handling withdrawal

(2) Compliance -- The IV approach

Stat 209

Greenland IJE 2000

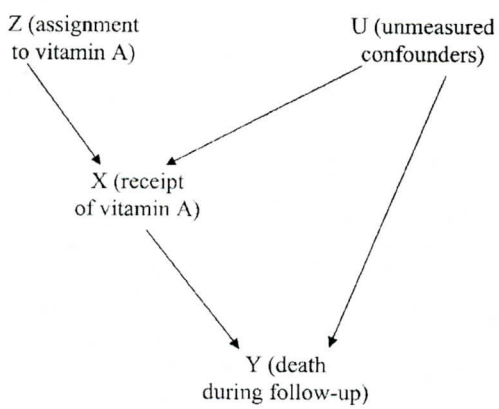


Figure 1

these concepts, Table 1 presents individual one-year mortality data from a cluster-randomized trial of vitamin A supplementation in childhood.[18,20] Of 450 villages, 229 were assigned to a treatment in which village children received two oral doses of vitamin A; children in the 221 control villages were assigned none. This protocol resulted in 12 094 children assigned to the treatment (Z = 1) and 11 588 assigned to the control (Z = 0). Only children assigned to treatment received the treatment; that is, no one had Z = 0 and X = 1. Unfortunately, 2419 (20%) of those assigned to the treatment did not receive the treatment (had Z = 1 and X = 0), resulting in only 9675 receiving treatment (X = 1). Nonetheless, assumption 1 is satisfied if the randomization was not subverted, while assumption 2 is supported by the data: Assignment to vitamin A increased the percentage receiving A from 0 to 80%.

Table 1 One-year mortality data from cluster-randomized trial of vitamin A supplementation in children.[20] Z = 1 if assigned A, 0 if not; X = 1 if received A, 0 if not

analysis in HW

	Z = 1			Z = 0	
	X = 1	X = 0	Total	X = 1	X = 0
Deaths (Y = 1)	12	34	46	0	74
Total	9675	2419	12 094	0	11 588
Risk[a]	124	1406	380	undefined	639

[a] Deaths per 100 000 within one year.

Example 1 Coronary drug see handout front page

Coronary Drug Project

- Randomized, multi-center, double-blind, placebo-controlled trial of clofibrate for treatment for coronary heart disease
  - 1103 men on clofibrate
  - 2789 men on placebo
- ITT analysis of 5-year mortality on clofibrate was 20.0%, 20.9% on placebo (p=0.55)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

◄ Previous | Volume 303:1038-1041 | October 30, 1980 | Number 18 | Next ►

Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project

Abstract

The Coronary Drug Project was carried out to evaluate the efficacy and safety of several lipid-influencing drugs in the long-term treatment of coronary heart disease. The five-year mortality in 1103 men treated with clofibrate was 20.0 per cent, as compared with 20.9 per cent in 2789 men given placebo (P = 0.55). Good adherers to clofibrate, i.e., patients who took 80 per cent of more of the protocol prescription during the five-year follow-up period, had a substantially lower five-year mortality than did poor adherers to clofibrate (15.0 vs. 24.6 per cent; P = 0.00011). However, similar findings were noted in the placebo group, i.e., 15.1 per cent mortality for good adherers and 28.3 per cent for poor adherers (P = 4.7x10-16). These findings and various other analyses of mortality in the clofibrate and placebo groups of the project show the serious difficulty, if not impossibility, of evaluating treatment efficacy in subgroups determined by patient responses (e.g., adherence or cholesterol change) to the treatment protocol after randomization.

This article has been cited by other articles:

- Huss, A. PhD, Scott, P. MSc, Stuck, A. E. MD, Trotter, C. PhD, Egger, M. MD MSc (2009). Efficacy of pneumococcal vaccination in adults: a meta-analysis. *CMAJ* 180: 48-58 [Abstract] [Full Text]
- Kritchevsky, S. B (2008). "Adherence bias" in nutritional epidemiology. *Am. J. Clin. Nutr.* 88: 1448-1449 [Full Text]
- McPherson, K. (2008). Do patients' preferences matter?. *BMJ* 337: a2034-a2034 [Full Text]
- Suissa, S., Ernst, P., Vandemheen, K. L., Aaron, S. D. (2008). Methodological issues in therapeutic trials of COPD. *Eur Respir J* 31: 927-933 [Abstract] [Full Text]
- Jackevicius, C. A., Li, P., Tu, J. V. (2008). Prevalence, Predictors, and Outcomes of Primary Nonadherence After Acute Myocardial Infarction. *Circulation* 117: 1028-1036 [Abstract] [Full Text]
- Ho, P. M., Magid, D. J., Shetterly, S. M., Olson, K. L., Peterson, P. N., Masoudi, F. A., Rumsfeld, J. S. (2008). Importance of Therapy Intensification and Medication Nonadherence for Blood Pressure Control in Patients With Coronary Disease. *Arch Intern Med* 168: 271-276 [Abstract] [Full Text]
- Frenkel, O. (2008). A Phenomenology of the 'Placebo Effect': Taking Meaning from the Mind to the Body. *J Med Philos* 33: 58-79 [Abstract] [Full Text]
- Streiner, D. L (2008). Missing data and the trouble with LOCF. *Evid. Based Ment. Health* 11: 3-5 [Full Text]
- Shumway-Cook, A., Silver, I. F., LeMier, M., York, S., Cummings, P., Koepsell, T. D. (2007). Effectiveness of a Community-Based Multifactorial Intervention on Falls and Fall Risk Factors in Community-Living Older Adults: A Randomized, Controlled Trial. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62: 1420-1427 [Abstract] [Full Text]
- Gehi, A. K., Ali, S., Na, B., Whooley, M. A. (2007). Self-reported Medication Adherence and Cardiovascular Events in Patients With Stable Coronary Heart Disease: The Heart and Soul Study. *Arch Intern Med* 167: 1798-1803 [Abstract] [Full Text]
- Schneeweiss, S., Patrick, A. R., Maclure, M., Dormuth, C. R., Glynn, R. J. (2007). Adherence to Statin Therapy Under Drug Cost Sharing in Patients With and Without Acute Myocardial Infarction: A Population-Based Natural Experiment. *Circulation* 115: 2128-2135 [Abstract] [Full Text]
- Stanley, K. (2007). Evaluation of Randomized Controlled Trials. *Circulation* 115: 1819-1822 [Full Text]
- Cepeda, M. S., Carr, D. B., Sarquis, T., Miranda, N., Garcia, R. J., Zarate, C. (2007). Static Magnetic Therapy Does Not Decrease Pain or Opioid Requirements: A Randomized Double-Blind Trial. *Anesth. Analg.* 104: 290-294 [Abstract] [Full Text]
- Olshansky, B. (2007). Placebo and Nocebo in Cardiovascular Health: Implications for Healthcare, Research, and the Doctor-Patient Relationship. *J Am Coll Cardiol* 49: 415-421 [Abstract] [Full Text]

Coronary Drug Project

- There was speculation that good compliers would show the clofibrate benefit and poor compliers would have mortality similar to placebo subjects
- Good compliance defined as 80% of protocol prescribed treatment taken

Coronary Drug Project

- In **clofibrate** subjects, mortality rates at 5 years were
  - Compliant: 15.0%
  - Non-compliant: 24.6%
- Subjects compliant with clofibrate had significantly lower mortality (p=0.0001)!
- Explanatory analysis – compare compliant clofibrate subjects to subjects without adequate clofibrate intake (clofibrate noncompliers and placebo subjects) – **significant**!

Coronary Drug Project

- But, in **placebo** subjects, mortality rates at 5 years were
  - Compliant: 15.1%
  - Non-compliant: 28.2%
- Subjects compliant with placebo had significantly lower mortality ($p < 0.0001$)!
- The explanatory analysis would miss this effect of compliance

Coronary Drug Project

- Clofibrate wasn't more beneficial than placebo
- Compliance with assigned treatment was beneficial
- The decrease in mortality of subjects complying with clofibrate shouldn't be attributed to clofibrate as would be done in an explanatory analysis
- The ITT result of non-significant clofibrate effect is correct

Compliance as an Explanatory Variable in Clinical Trials

B. EFRON and D. FELDMAN*

The Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) measured the effectiveness of the drug cholestyramine for lowering cholesterol levels. The patients in the study were measured for compliance (the proportion of the intended dose actually taken) and for cholesterol decrease. The compliance-response regression for the Treatment group shows a smooth increasing effect of the drug in cholesterol level with increasing compliance. However, a similar, though less dramatic, compliance-response regression is seen in the Control group. This article investigates the recovery of the true dose-response curve from the Treatment and Control compliance-response curves. A simple model is proposed, analyzed, and applied to the LRC-CPPT data. Under this model, part but not all of the true dose-response curve can be estimated.

1. INTRODUCTION

Figure 1 shows results from the Stanford portion of the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT). This was a placebo-controlled double-blind randomized clinical trial concerning the efficacy of cholestyramine for lowering cholesterol level and thereby reducing coronary heart disease. Lipid Research Clinics Program (1984) describes the complete study.

The left panel depicts the Treatment group, comprising 164 men after removal of the one indicated outlier. The vertical axis is the decrease in total cholesterol level, while the horizontal axis indicates compliance (the proportion of the nominal cholestyramine dose actually taken). A "dose-response" relationship is quite evident: better compliance leads to a greater decrease in cholesterol level, as indicated by the quadratic regression curve. (More careful definitions of compliance and cholesterol differences are given in Section 3.)

The right panel of Figure 1 shows the same plot for the placebo Control group, comprising 171 men after removal of one outlier. Compliance here has been adjusted to match the compliance distribution in the Treatment group, as discussed in Section 3. Interestingly enough, there is a significant dose-response relation between compliance and cholesterol decrease in the placebo Control group, though of smaller magnitude than in the Treatment group.

It seems unfortunate that compliance was so variable in this experiment, but a good argument can be made that the situation is actually more favorable to the investigator than if 100% compliance had been enforced. Figure 1 seems to indicate a nice dose-response relationship, which would not be visible if compliance had been perfect.

The difficulty here of course is that compliance (and hence dose) has not been assigned in a randomized fashion by the investigators, as it would have been in a genuine dose-response experiment. Compliance is an uncontrolled covariate, and it may be that better compliers are better patients to begin with. In fact, the positive dose-response relationship for the Control group indicates that this is certainly the case.

This article concerns the proper interpretation of the data in Figure 1. What would we have seen if 100% compliance had been enforced? To what degree can we estimate the actual dose-response relationship? Section 2 discusses a simple model, which is applied to the LRC-CPPT data in Section 3. Some final remarks are made in Section 4.

Because of the special nature of compliance, the questions considered here are not standard problems of covariate adjustment, as discussed for example in Koch et al (1982). First of all compliance is an adjustable covariate, unlike, say, age. What we would see if 100% compliance were enforced is a different question than what we would see if all patients were 60 years old. The former question concerns *changing* everyone's compliance to 100%, while the latter presumably refers to considering only the population of 60-year-olds.

Second, unlike other adjustable covariates (such as exercise level), compliance has a different meaning in the Treatment of Control groups. Compliance determines the amount of active drug taken for Treatment group patients and also indicates something about the patient's psychological status. In the Control group, only the psychological component of compliance applies.

Third, the Treatment and Control groups should behave identically at the 0% end of the compliance scale, since there is no distinction between 0% compliance with an active drug or with a placebo. This seems to be the case in Figure 1, and is in fact the case for experiment LRC-CPPT, as standard hypothesis tests will confirm in Section 4.

The results of the LRC-CPPT focused considerable attention on compliance as an important explanatory variable in clinical trials. Cholestyramine's package labelling now includes a compliance-based dose-response curve; see Urquhart and Chevalley (1988). A recently developed electronic monitor for pill bottles makes compliance data of quality superior to that from LRC-CPPT (which was based on packet counts) available for most clinical trials. See Chevalley and Urquhart (1987).

The goal of this article, beyond analysis of the LRC-CPPT

* B. Efron is Max H. Stein Professor of Humanities and Sciences and Professor of Statistics and Biostatistics, Department of Statistics, Stanford University, Stanford, CA 94305. D. Feldman is an M.D. in Palo Alto, CA 94303. The authors are grateful to the editors and referees for an unusually helpful critique of the original manuscript, and to Dr. John Urquhart for originally suggesting the problem studied here.

continuous (measured) compliance data

p.2 compliance not dichotomous

Compliance Efron-Feldman (JASA 1991)

(cholestyramine grit)

$z(u)$ compliance patient u

$Y_0(u)$ response y (cholest reduction) patient u if Placebo

$Y_x(u)$ response y patient u if given dose X active drug

$$Y_x(u) = G_x + (1+H_x) Y_0(u) + e_x(u) \qquad G_0 = H_0 = 0$$

$$\delta(x) = E(Y_x(u) - Y_0(u)) = G_x + H_x(E Y_0(u)) \quad \text{dose-response diff}$$

Data: ave compliance .601 $\bar{Y}_T = 32.81$, $\bar{Y}_C = 8.29$ (24.52)

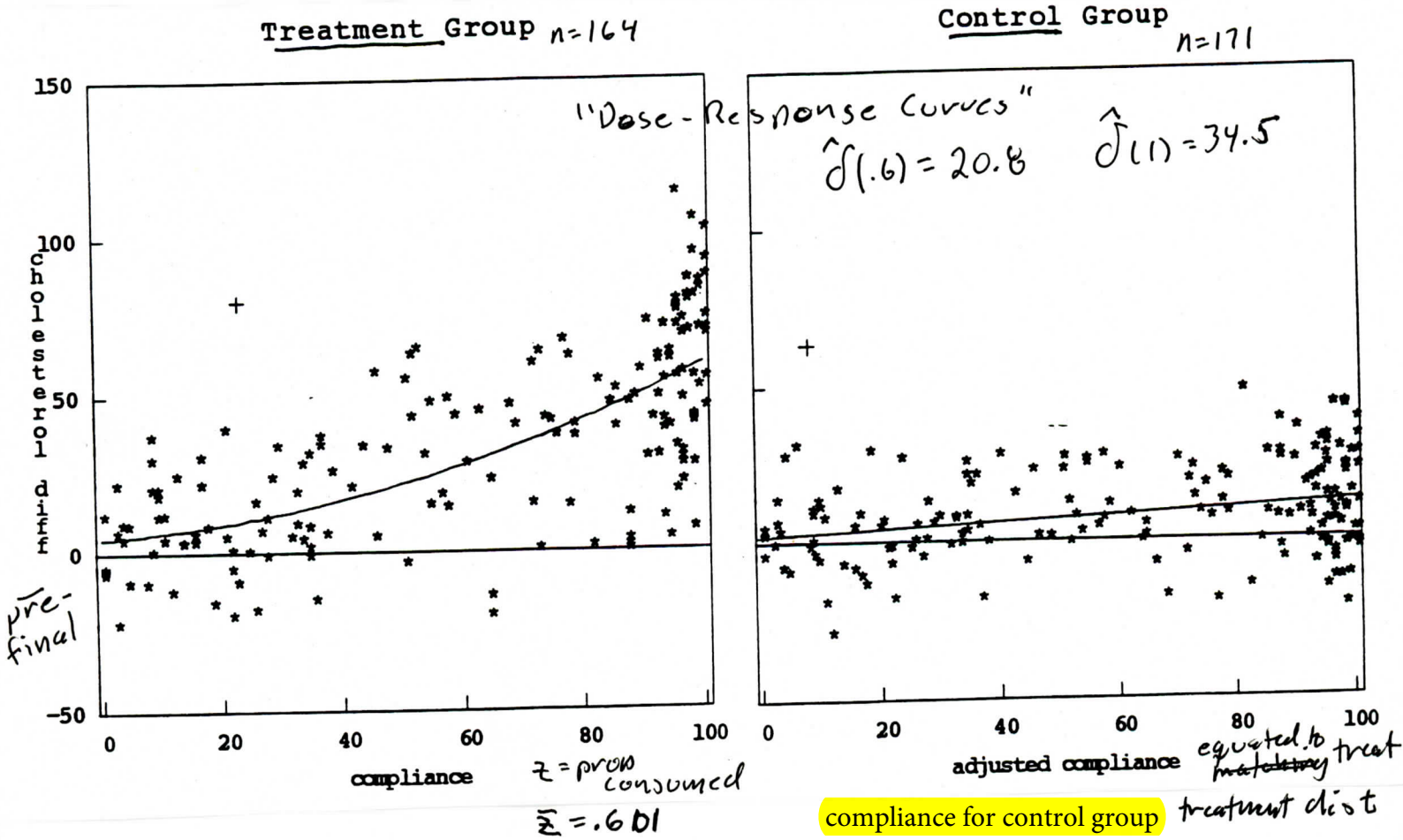
note $24.52/.601 = 40.8$

difference of dose-response curve $\delta(z)$ at z-level

$\hat{\delta}(.6) = 20.8$ $\hat{\delta}(1) = 34.5$

Compliance as an Explanatory Variable in Clinical Trials

B. Efron and D. Feldman



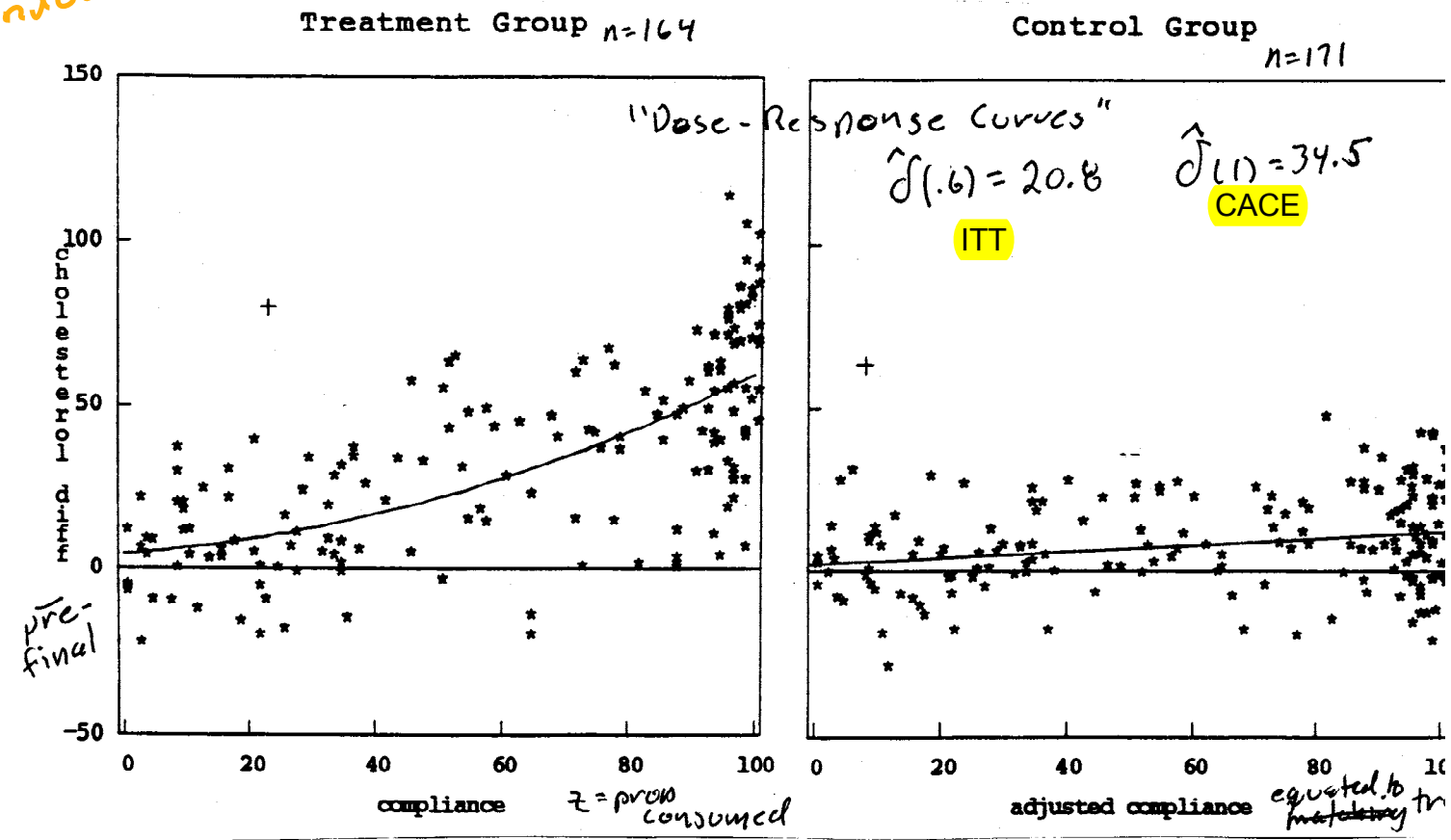
Example 4 Efron Feldman dose response

- No consistent method for handling withdrawal

Compliance as an Explanatory Variable in Clinical Trials

B. Efron and D. Feldman

P2 random



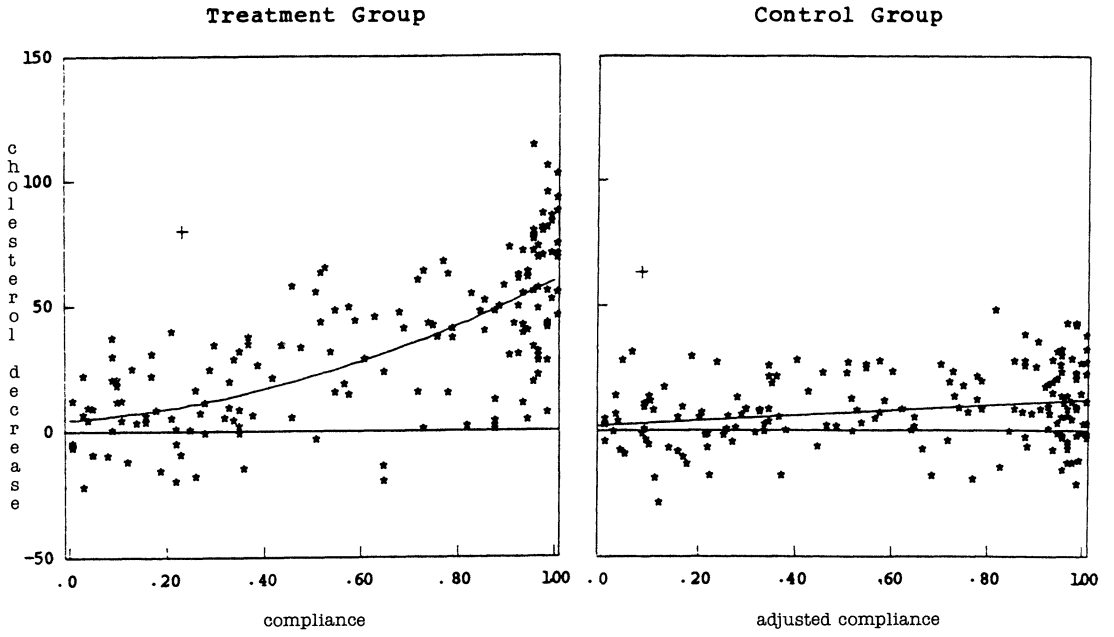


Figure 1. Stanford portion of LRC-CPPT. Left Panel: Treatment group, 164 men (after removal of outlier indicated by +); vertical axis is decrease in total cholesterol; horizontal axis is compliance (the proportion of the nominal choelstyramine dose actually taken). Better compliance leads to larger decreases in total cholesterol, as indicated by the quadratic regression curve. Right Panel: Placebo Control group, 171 men (after removal of outlier +); compliance has been adjusted to match the distribution of compliance in the Treatment group. There is a smaller, but still significant, dose-response relationship between compliance and cholesterol decrease, indicated by the linear regression line.

data, is to establish a moderately general framework for making use of compliance data in clinical trials. The theory presented here is intended to supplement rather than to replace the more conservative intent-to-treat analysis, which in its simplest form is just a two-sample test for equality between the Treatment and Control groups, ignoring covariate information. The LRC-CPPT results are too clear-cut to make such a test interesting, but in more ambiguous situations it can, because of its assumption-free basis, play a decisive role.

2. DOSE–RESPONSE AND COMPLIANCE–RESPONSE

We are interested in recovering the true dose-response curve from the compliance-response curves shown in Figure 1. This section discusses a simple model relating dose, compliance, and response. We will see that it is impossible to recover the full dose-response curve in general, but that we can always estimate certain important aspects of it. In the special situation where the true dose-response curve is linear, it can be estimated in its entirety.

Our theory addresses certain hypothetical questions, such as what results would have been seen if all patients complied perfectly. We will use Rubin's theory of causality, as nicely described in Holland (1986), to phrase such questions and their answers clearly. See also Holland (1988). Rubin's theory begins with a population U of possible patients, u representing an individual member of U. The set of patients actually enrolled in the study, those represented in Figure 1, is obtained by some sampling scheme from U which need not be specified for the purposes here.

Let $Y_0(u)$ represent patient u's placebo-response, in our case the cholesterol decrease experienced by patient u if zero amount of the drug is taken. $Y_0(u)$ is defined for all patients u in U, but only observed for patients in the Control group of the study, or for those in the Treatment group who had zero compliance. Let $z(u)$ be patient u's *compliance*, measured as the proportion of nominal dose actually taken (the horizontal axis in Figure 1); $z(u)$ is observed for all patients in the study. Both $Y_0(u)$ and $z(u)$ are considered to be inherent properties of the patients, what Holland calls "attributes," which may or may not be observed depending on the sampling and treatment assignment schemes employed. (See Remark B.)

We can imagine giving each patient u every possible dose X of the active drug. Here X is measured in the same units as z, as the proportion of nominal dose, so $0 \le X \le 1$. The model we will use relates $Y_X(u)$, the response (cholestral decrease) of patient u, to the amount X of the active drug and the placebo response $Y_0(u)$, as follows:

$$Y_X(u) = G_X + (1 + H_X)Y_0(u) + e_X(u). \tag{2.1}$$

G_X and H_X are continuous functions of X, with

$$G_0 = H_0 = 0; \tag{2.2}$$

$e_X(u)$ is a disturbance term satisfying

$$e_0(u) = 0 \quad \text{and} \quad E\{e_X \mid z\} = 0. \tag{2.3}$$

Model (2.1)–(2.3) is further discussed later (see Remark A).

The true dose-response curve $\delta(X)$ is defined as the average net effect on response of giving all patients amount

X of the active drug rather than amount zero,

$$\delta(X) \equiv E\{Y_X - Y_0\}. \tag{2.4}$$

Expectations and conditional expectations are defined as the appropriate averages over U, so $E\{Y_X - Y_0\} = \Sigma_{u\in U}(Y_X(u) - Y_0(u))/N$, N being the number of elements in U. Expression (2.4) is what Rubin and Holland call the *average causal effect* of X on response. From (2.1)–(2.3) we calculate

$$\delta(X) = G_X + H_X C_0, \tag{2.5}$$

where

$$C_0 \equiv E\{Y_0\}, \tag{2.6}$$

the average placebo-response over all of U.

In a true dose-response experiment, the statistician gets to observe $Y_X(u)$ for different values of X on randomly selected patients u. This makes estimation of $\delta(X)$ straightforward. The situation is more complicated here. Two more definitions are necessary to define what is actually being observed in Figure 1.

Let $s(u)$ indicate patient u's group assignment,

$$\begin{aligned} s(u) &= 0 \quad \text{if} \quad u \quad \text{assigned to Control group} \\ &= 1 \quad \text{if} \quad u \quad \text{assigned to Treatment group.} \end{aligned} \tag{2.7}$$

In a completely randomized experiment like LRC-CPPT, the statistician determines $s(u)$ by independent flips of a fair coin. [In practice $s(u)$ is determined only for patients u selected into the study, but we can think of it as defined over all of U.] Also let $x(u)$ be the amount of active drug actually taken by patient u, so

$$x(u) = s(u) \cdot z(u); \tag{2.8}$$

$x(u) = 0$ in the Control group, and $x(u) = z(u)$ in the Treatment group.

The observed response for patient u is

$$y(u) = G_{x(u)} + (1 + H_{x(u)})Y_0(u) + e_{x(u)}(u) \tag{2.9}$$

according to (2.1), (2.8). This is the quantity plotted along the vertical axis in Figure 1. Let $y_C(u)$ indicate a Control group response, that is $y(u)$ for a patient u having $s(u) = 0$, and likewise write $y_T(u)$ for a Treatment group response. The previous definitions give

$$y_C(u) = Y_0(u) \tag{2.10}$$

and

$$y_T(u) = G_{z(u)} + (1 + H_{z(u)})Y_0(u) + e_{z(u)}(u). \tag{2.11}$$

The compliance-response regression functions $C(z) \equiv E\{y_C \mid z\} = E\{y \mid s = 0, z\}$ and $T(z) \equiv E\{y_T \mid z\} = E\{y \mid s = 1, z\}$ are obtained from (2.10), (2.11),

$$C(z) = E\{Y_0 \mid z\} \tag{2.12}$$

and

$$T(z) = G_z + (1 + H_z)E\{Y_0 \mid z\}. \tag{2.13}$$

These are the functions estimated by the linear and quadratic curves in Figure 1. We will be particularly interested in the difference between the two regressions, the *observed difference* $D(z)$,

$$\begin{aligned} D(z) \equiv T(z) - C(z) &= G_z + H_z E\{Y_0 \mid z\} \\ &= G_z + H_z C(z). \end{aligned} \tag{2.14}$$

The observed difference $D(z)$ is an obvious first guess for the true dose-response function $\delta(z)$. Comparing (2.14) with (2.5), (2.6) gives a simple but important result:

Lemma. Under model (2.1)–(2.3), the observed difference $D(z)$ and the true dose-response $\delta(z)$ satisfy

$$D(z) - \delta(z) = H_z \cdot \{C(z) - C_0\}. \tag{2.15}$$

In practice we can estimate $C(z)$ and $D(z)$ from the compliance-response data, as in Figure 1. But we want to estimate the true dose-response function $\delta(z)$. Relation (2.15) shows the limitations on estimating $\delta(z)$:

1. For any given compliance-response regressions $C(z)$ and $D(z)$, there is a family of possible dose-response functions $\delta(z)$,

$$\delta(z) = D(z) - H_z \cdot \{C(z) - C_0\}, \tag{2.16}$$

corresponding to different choices of H_X in (2.1). In general, $\delta(z)$ is not completely identifiable under model (2.1)–(2.3).

2. If we assume $H_X \equiv 0$, then $\delta(z) \equiv D(z)$. Setting $H_X \equiv 0$ amounts to assuming no interaction between X and $Y_0(u)$ (see Remark A), so the no-interaction assumption makes $\delta(z)$ directly estimable from $D(z)$.

3. Let z_0 be a value of the compliance z such that

$$C(z_0) = C_0 = E\{Y_0\}. \tag{2.17a}$$

Then

$$\delta(z_0) = D(z_0), \tag{2.17b}$$

no matter what the interaction H_X may be. In other words, $\delta(z)$ and $D(z)$ always intersect at $z = z_0$.

4. If $C(z)$ is linear, then the intersection value z_0 is given by

$$z_0 = E\{z\}, \tag{2.18}$$

since a linear regression passes through the point $(z_0, C(z_0))$.

5. The curves $\delta(z)$ and $D(z)$ always intersect at two points in the plane, namely $(0, 0)$ and $(z_0, D(z_0))$ (see Remark C). Therefore, if we assume that $\delta(z)$ is linear, it can be fully estimated from the compliance-response data via the estimation of z_0 and $D(z_0)$.

These points are illustrated by the quadratic model shown in Figure 2, a simple example of model (2.1)–(2.3), which will be used in Section 3 as part of the analysis of the LRC-CPPT data. The quadratic model assumes $C(z)$ and H_z linear, and $D(z)$ quadratic,

$$\begin{aligned} C(z) = c_0 + c_1 z, \qquad D(z) = d_1 z + d_2 z^2, \\ H_z = h_1 z. \end{aligned} \tag{2.19}$$

[This is equivalent to model (2.1), (2.3) with $G_z = g_1 z + g_2 z^2$, $H_z = h_1 z$, and $E\{Y_0 \mid z\} \equiv c_0 + c_1 z$, where $(g_1, g_2) = (d_1 - h_1 c_0, d_2 - h_1 c_1)$.] Since $C(z)$ is linear, (2.18) gives

For the quadratic model (2.19), we can express (2.27) as $c_1 h_1 - d_2 \le Q(z)$,

$$Q(z) \equiv \frac{c_1}{z}\left[\left(\frac{\text{var}\{y_T \mid z\}}{\text{var}\{y_C \mid z\}}\right)^{1/2} - 1\right] - d_2. \quad (2.28)$$

The function $Q(z)$ is estimable from the data in Figure 1. In Section 3 we will use (2.28) to show that h_1 is probably less than d_2/c_1 for the LRC-CPPT study, so that $\delta(1)$ is probably greater than the linear extrapolate in Figure 2.

3. THE LRC–CPPT DATA

The dose-compliance-response model decribed in Section 2 will now be used to analyze the LRC-CPPT data, as it appears in Figure 1. We begin with a more careful description of the variables involved. See Lipid Research Clinic Program (1984) for further details.

1. The patients were men aged 35 to 59 years with high initial cholesterol levels (total plasma cholesterol level greater than 265).
2. Two baseline cholesterol measurements were taken for each patient, one before and one after a low-cholesterol diet was suggested to them.
3. Patients were subsequently observed at two-month intervals, for a period averaging 7.3 years.
4. The nominal dose was six 4-gram packets per day of cholestyramine. This was reduced for some patients who could not tolerate 24 grams per day.

The quantity y, labelled cholesterol decrease in Figure 1, is

$$y = .25 \cdot (\text{prediet baseline cholesterol}) + .75 \cdot (\text{postdiet baseline cholesterol}) - (\text{average of all subsequent cholesterol readings}). \quad (3.1)$$

The weights .25, .75 were chosen as nearly optimal on the basis of a preliminary regression analysis; see Remark F. Patients returned unused packets of cholestyramine or placebo at each visit. Compliance was the proportion of assigned packets not returned, averaged over all visits. There are some obvious weaknesses of this compliance measure compared with the newly available electronic compliance monitoring, so the strength of the regression relationships in Figure 1 is perhaps surprising.

Perfect Blind Assumption. Figure 3 compares compliance in the two groups. It is obvious that compliance was better for the Controls, violating the perfect blind assumption that (Y_0, z) has the same joint distribution in both groups. As a simple corrective, each Control compliance z_C was mapped into $\tilde{z}_C = m(z_C)$, where m was defined in terms of $\hat{F}_T$ and $\hat{F}_C$, the cumulative distribution functions (cdf) for the Treatment and Control groups respectively,

$$\tilde{z}_C = m(z_C) \equiv \hat{F}_T^{-1}\hat{F}_C(z_C). \quad (3.2)$$

Mapping (3.2) makes the 171 $\tilde{z}_C$ values for the Control group have nearly the same empirical distribution as the

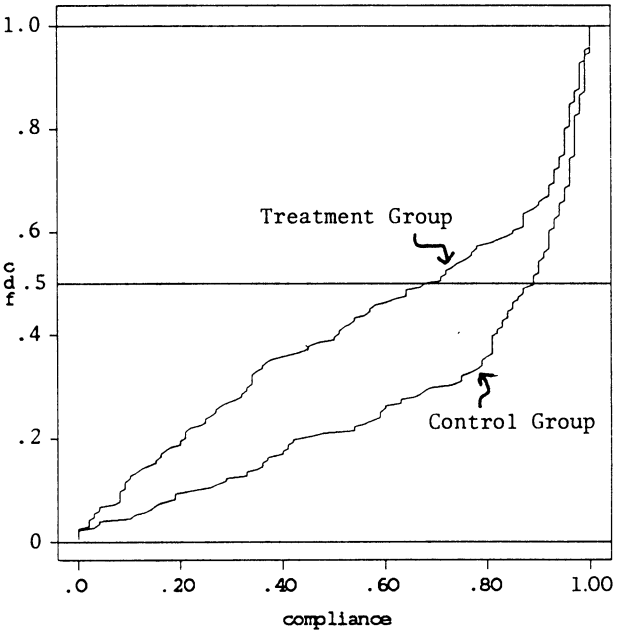


Figure 3. Compliance cdf's in the Treatment and Control groups. Compliance was significantly higher in the Control group. The average compliance for the Treatment group was $z_0 = .601$.

164 z_T values for the Treatment group. In particular, the average value of the $\tilde{z}_C$ is nearly the same as the z_T average, $z_0 = .601$.

In all of our analyses, what is called "compliance" for the Control group is actually the adjusted compliance, (3.2). Accepting result (2.15) amounts to believing that $E\{Y_{0T} \mid z_T = z\} = E\{Y_{0C} \mid \tilde{z}_C = z\}$, in particular for $z = z_0$. Here (Y_{0T}, z_T) is (Y_0, z) for a random patient put into the Treatment group, and likewise $(Y_{0C}, \tilde{z}_C)$ is $(Y_0, m(z))$ for a random patient put into the Control group. See Remark G.

Variance Structure. Figure 4 displays regression percentiles for the Treatment and Control data obtained by the method of asymmetric least squares; see Efron (1988). These are estimates of the conditional percentiles of cholesterol difference y given compliance z. The Treatment group clearly shows increasing variability of y given z as we move toward the high end of the compliance scale: the percentiles are about 2.11 times as far apart of $z = 1.00$ as at $z = 0$. The Control group percentiles, except for the lowest one, show this same effect though less dramatically: the percentiles are about 1.35 times as far apart at $z = 1.00$ as at $z = 0$.

We will use the following values for the conditional variances in the Treatment and Control groups, $v_{Ti} \equiv \text{var}\{y_{Ti} \mid z_i\}$ and $v_{Ci} \equiv \text{var}\{y_{Ci} \mid z_i\}$,

$$v_{Ti} = 471.52 + 485.00(z_i - z_0)$$
$$v_{Ci} = 198.09 + 128.33(z_i - z_0), \quad (3.3)$$

where $z_0 = .601$, the average compliance. (For v_{Ci}, z_i is actually the adjusted value $\tilde{z}_i$, as it will be from now on unless noted otherwise.)

The numerical coefficients in (3.3) were obtained by first running an ordinary least squares regression of cholesterol

---

Problem 5

From the Booil Jo presentation slides in lecture, consider the
JHU PIRC Intervention Study: N=284

Estimate Intervention Effects With Noncompliance

The Johns Hopkins Public School Preventive Intervention Study was conducted by the Johns Hopkins University Preventive Intervention Research Center (JHU PIRC) in 1993-1994 (lalongo et al., 1999~ The study was designed to improve academic achievement and to reduce early behavioral problems of school children. Teachers and first-grade children were randomly assigned to intervention conditions. The control condition and the Family-School Partnership Intervention condition are compared in this example. In the intervention condition, parents were asked to implement 66 take-home activities related to literacy and mathematics over a six-month period. One of the major outcome measures in the JHU PIRC preventive trial was the TOCA-R (Teacher Observation of Classroom Adaptation)

· Completed at least 45 activities = compliers.

· Outcome: change score (baseline - followup) of anti-social behavior .

From the means and compliance data given in the class materials (also linked Booil talk)
compute treatment effect estimate of change in anti-social behavior: give ITT estimate and CACE estimate

---

Problem 6

Artificial data in the image of Efron-Feldman

data frame containing Compliance, Group, and Outcome in file http://web.stanford.edu/~rag/stat209/hw7efdata

Obtain ITT estimate of group (treatment) effect with a confidence interval.
Try using G as an instrument for the Y ~ comp regression.
What does that produce?

Alternatively use the Rubin formulation with a dichotomous compliance indicator defined as TRUE for compliance > .8
in these data. What is your CACE estimate. What assumptions did you make? Compare with ITT estimate.

==========

e[redacted]

Example 6

Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City

John BARNARD, Constantine E. FRANGAKIS, Jennifer L. HILL, and Donald B. RUBIN

The precarious state of the educational system in the inner cities of the United States, as well as its potential causes and solutions, have been popular topics of debate in recent years. Part of the difficulty in resolving this debate is the lack of solid empirical evidence regarding the true impact of educational initiatives. The efficacy of so-called "school choice" programs has been a particularly contentious issue. A current multimillion dollar program, the School Choice Scholarship Foundation Program in New York, randomized the distribution of vouchers in an attempt to shed some light on this issue. This is an important time for school choice, because on June 27, 2002 the U.S. Supreme Court upheld the constitutionality of a voucher program in Cleveland that provides scholarships both to secular and religious private schools. Although this study benefits immensely from a randomized design, it suffers from complications common to such research with human subjects: noncompliance with assigned "treatments" and missing data. Recent work has revealed threats to valid estimates of experimental effects that exist in the presence of noncompliance and missing data, even when the goal is to estimate simple intention-to-treat effects. Our goal was to create a better solution when faced with both noncompliance and missing data. This article presents a model that accommodates these complications that is based on the general framework of "principal stratification" and thus relies on more plausible assumptions than standard methodology. Our analyses revealed positive effects on math scores for children who applied to the program from certain types of schools—those with average test scores below the citywide median. Among these children, the effects are stronger for children who applied in the first grade and for African-American children.

KEY WORDS: Causal inference; Missing data; Noncompliance; Pattern mixture models; Principal stratification; Rubin causal model; School choice.

1. INTRODUCTION

There appears to be a crisis in America's public schools. "More than half of 4th and 8th graders fail to reach the most minimal standard on national tests in reading, math, and science, meaning that they probably have difficulty doing grade-level work" (Education Week 1998). The problem is worse in high poverty urban areas. For instance, although only 43% of urban fourth-graders achieved a basic level of skill on a National Assessment of Educational Progress (NAEP) reading test, a meager 23% of those in high-poverty urban schools met this standard.

One of the most complicated and contentious of educational reforms currently being proposed is school choice. Debates about the equity and potential efficacy of school choice have increased in intensity over the past few years. Authors making a case for school choice include Cobb (1992), Brandl (1998), and Coulson (1999). A collection of essays that report mainly positive school choice effects has been published by Peterson and Hassel (1998). Recent critiques of school choice include those by the Carnegie Foundation for the Advancement of Teaching (1992), Cookson (1994), Fuller and Elmore (1996), and Levin (1998).

In this article we evaluate a randomized experiment conducted in New York City made possible by the privately-funded School Choice Scholarships Foundation (SCSF). The SCSF program provided the first opportunity to examine the question of the potential for improved school performance (as well as parental satisfaction and involvement, school mobility, and racial integration) in private schools versus public schools using a carefully designed and monitored randomized field experiment. Earlier studies were observational in nature and thus subject to selection bias (i.e., nonignorable treatment assignment). Studies finding positive educational benefits from attending private schools include those of Coleman, Hoffer, and Kilgore (1982), Chubb and Moe (1990), and Derek (1997). Critiques of these studies include those of Goldberger and Cain (1982) and Wilms (1985). On June 27, 2002, the U.S. Supreme Court upheld the constitutionality of a voucher program in Cleveland that provides scholarships both to secular and religious private schools.

As occurs in most research involving human subjects, however, our study, although carefully implemented, suffered from complications due to missing background and outcome data and also to noncompliance with the randomly assigned treatment. We focus on describing and addressing these complications in our study using a Bayesian approach with the framework of principal stratification (Frangakis and Rubin 2002).

We describe the study in Section 2 and summarize its data complications in Section 3. In Section 4 we place the study in

John Barnard is Senior Research Statistician, deCODE Genetics, Waltham, MA-02451 (E-mail: *john.barnard@decode.is*). Constantine E. Frangakis is Assistant Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205 (E-mail: *cfrangak@jhsph.edu*). Jennifer L. Hill is Assistant Professor, School of International and Public Affairs, Columbia University, New York, NY 10027 (E-mail: *jh1030@columbia.edu*). Donald B. Rubin is John L. Loeb Professor of Statistics and Chair, Department of Statistics, Harvard University, Cambridge, MA 02138 (E-mail: *rubin@stat.harvard.edu*). The authors thank the editor, an associate editor, and three reviewers for very helpful comments; David Myers and Paul E. Peterson as principal coinvestigators for this evaluation; and the School Choice Scholarships Foundation (SCSF) for cooperating fully with this evaluation. The work was supported in part by National Institutes of Health (NIH) grant RO1 EY 014314-01, National Science Foundation grants SBR 9709359 and DMS 9705158; and by grants from the following foundations: Achelis Foundation, Bodman Foundation, Lynde and Harry Bradley Foundation, Donner Foundation, Milton and Rose D. Friedman Foundation, John M. Olin Foundation, David and Lucile Packard Foundation, Smith Richardson Foundation, and the Spencer Foundation. The authors also thank Kristin Kearns Jordan and other members of the SCSF staff for their cooperation and assistance with data collection, and Daniel Mayer and Julia Kim, from Mathematica Policy Research, for preparing the survey and test score data and answering questions about that data. The methodology, analyses of data, reported findings and interpretations of findings are the sole responsibility of the authors and are not subject to the approval of SCSF or of any foundation providing support for this research.

Journal of the American Statistical Association
June 2003, Vol. 98, No. 462, Applications and Case Studies
DOI 10.1198/016214503000071