from UK lecture link

Idea of Instrumental Variables attributed to

Philip Wright 1861-1934

interested in working out whether price of butter was demand or supply driven

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 $Cov(X,Z) \neq 0$ empirical assoc. strong, weak insb

AND $cov(Z,u) = 0$ if true ancova also works

cue Dusty S. hope, untestable

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

Two-stage least squares (and IV) [PAF ch 8 proof in general]

predict x by z

$$\hat{x} = \bar{z} + \frac{\hat{cov}(x,z)}{\hat{var}(z)}(z - \bar{z})$$

slope of y on $\hat{x}$

$$\frac{\hat{cov}(y,\hat{x})}{\hat{var}(\hat{x})} = \frac{\left(\frac{\hat{cov}(x,z)}{\hat{var}(z)}\right)\hat{cov}(y,z)}{\hat{var}(z)\left(\frac{\hat{cov}(x,z)}{\hat{var}(z)}\right)^2}$$

(collect terms)

$$= \hat{cov}(y,z)/\hat{cov}(x,z) = \hat{\beta}^{IV}_{yx}$$

cf meas error ex

Is the Wide, Wide World of Economics Too Wide?

Is there anything economists won't study? Should there be?:

> Is an Economist Qualified To Solve Puzzle of Autism?, by Mark Whitehouse, WSJ: In the spring of 2005, Cornell University economist Michael Waldman noticed a strange correlation in Washington, Oregon and California. The more it rained or snowed, the more likely children were to be diagnosed with autism. ...
>
> [This] soon led Prof. Waldman to conclude that something children do more during rain or snow -- perhaps watching television -- must influence autism. Last October, Cornell announced the resulting paper in a news release headlined, "Early childhood TV viewing may trigger autism, data analysis suggests."
>
> Prof. Waldman's willingness to hazard an opinion on a delicate matter of science reflects the growing ambition of economists -- and also their growing hubris, in the view of critics. Academic economists are increasingly venturing beyond their traditional stomping ground, a wanderlust that has produced some powerful results but also has raised concerns about whether they're sometimes going too far. ...
>
> Such debates are likely to grow as economists delve into issues in education, politics, history and even epidemiology. Prof. Waldman's use of precipitation illustrates one of the tools that has emboldened them: the instrumental variable, a statistical method that, by introducing some random or natural influence, helps economists sort out questions of cause and effect. Using the technique, they can create "natural experiments" that seek to approximate the rigor of randomized trials -- the traditional gold standard of ... research. ...
>
> But as enthusiasm for the approach has grown, so too have questions. One concern: When economists use one variable as a proxy for another -- rainfall patterns instead of TV viewing, for example -- it's not always clear what the results actually measure. Also, the experiments on their own offer little insight into why one thing affects another.
>
> "There's a saying that ignorance is bliss," says James Heckman ... at the University of Chicago who won a Nobel Prize in 2000... "I think that characterizes a lot of the enthusiasm for these instruments." Says MIT economist Jerry Hausman, "If your instruments aren't perfect, you could go seriously wrong." ...
>
> In principle, the best way to figure out whether television triggers autism would be to do what medical researchers do: randomly select a group of susceptible babies at birth to refrain from television, then compare their autism rate to a similar control group that watched normal amounts of TV. If the abstaining group proved less likely to develop autism, that would point to TV as a culprit.
>
> Economists usually ...[cannot] perform that kind of experiment. ... Instead, economists look for instruments -- natural forces or government policies that do the random selection for them. First developed in the 1920s, the technique helps them separate cause and effect. Establishing whether A causes B can be difficult, because often it could go either way. If television watching were shown to be unusually prevalent among autistic children, it could mean either that television makes them autistic or that something about being autistic makes them more interested in TV. ...
>
> Prof. Waldman and his colleagues had such [techniques]... in mind when they approached autism and TV. By putting together weather data and government time-use studies, they found that children tended to spend more time in front of the television when it rained or snowed. Precipitation became the group's instrumental variable, because it randomly selected some children to watch more TV than others.
>
> The researchers looked at detailed precipitation and autism data from Washington, Oregon and California -- states where rain and snowfall tend to vary a lot. They found that children who grew up during periods of unusually high precipitation proved more likely to be diagnosed with autism. A second instrument for TV-watching, the percentage of

Table 1

Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments

Outcome Variable	*Endogenous Variable*	*Source of Instrumental Variable(s)*	*Reference*
		1. Natural Experiments	
Labor supply	Disability insurance replacement rates	Region and time variation in benefit rules	Gruber (2000)
Labor supply	Fertility	Sibling-Sex composition	Angrist and Evans (1998)
Education, Labor supply	Out-of-wedlock fertility	Occurrence of twin births	Bronars and Grogger (1994)
Wages	Unemployment insurance tax rate	State laws	Anderson and Meyer (2000)
Earnings	Years of schooling	Region and time variation in school construction	Duflo (2001)
Earnings	Years of schooling	Proximity to college	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Earnings	Veteran status	Cohort dummies	Imbens and van der Klaauw (1995)
Earnings	Veteran status	Draft lottery number	Angrist (1990)
Achievement test scores	Class size	Discontinuities in class size due to maximum class-size rule	Angrist and Lavy (1999)
College enrollment	Financial aid	Discontinuities in financial aid formula	van der Klaauw (1996)
Health	Heart attack surgery	Proximity to cardiac care centers	McClellan, McNeil and Newhouse (1994)
Crime	Police	Electoral cycles	Levitt (1997)
Employment and Earnings	Length of prison sentence	Randomly assigned federal judges	Kling (1999)
Birth weight	Maternal smoking	State cigarette taxes	Evans and Ringel (1999)
		2. Randomized Experiments	
Earnings	Participation in job training program	Random assignment of admission to training program	Bloom et al. (1997)
Earnings	Participation in Job Corps program	Random assignment of admission to training program	Burghardt et al. (2001)
Achievement test scores	Enrollment in private school	Randomly selected offer of school voucher	Howell et al. (2000)
Achievement test scores	Class size	Random assignment to a small or normal-size class	Krueger (1999)
Achievement test scores	Hours of study	Random mailing of test preparation materials	Powers and Swinton (1984)
Birth weight	Maternal smoking	Random assignment of free smoker's counseling	Permutt and Hebel (1989)

Mroz87 U.S. Women's Labor Force Participation

Lab 3 data

The Mroz87 data frame contains data on 753 married women.
These data are collected within the "Panel Study of Income
Dynamics" (PSID). Of the 753 observations, the first 428 are for
women with positive hours worked in 1975, while the remaining 325
observations are for women who did not work for pay in 1975.

I took these data from an R package and placed in file
http://www-stat.stanford.edu/~rag/stat209/Mroz87.dat
or you can obtain from installing the micEcon package.

Format

This data frame contains the following columns:

lfp Dummy variable for labor-force participation.
hours Wife's hours of work in 1975.
kids5 Number of children 5 years old or younger.
kids618 Number of children 6 to 18 years old.
age Wife's age.
educ Wife's educational attainment, in years.
wage Wife's average hourly earnings, in 1975 dollars.
repwage Wife's wage reported at the time of the 1976 interview.
hushrs Husband's hours worked in 1975.
husage Husband's age.
huseduc Husband's educational attainment, in years.
huswage Husband's wage, in 1975 dollars.
faminc Family income, in 1975 dollars.
mtr Marginal tax rate facing the wife.
motheduc Wife's mother's educational attainment, in years.
fatheduc Wife's father's educational attainment, in years.
unem Unemployment rate in county of residence, in percentage points.
city Dummy variable = 1 if live in large city, else 0.
exper Actual years of wife's previous labor market experience.
nwifeinc Non-wife income.
wifecoll Dummy variable for wife's college attendance.
huscoll Dummy variable for husband's college attendance.

Package 'AER'

January 27, 2015

Version 1.2-2

Date 2014-01-28

Title Applied Econometrics with R

Description Functions, data sets, examples, demos, and vignettes for the book
Christian Kleiber and Achim Zeileis (2008),
Applied Econometrics with R, Springer-Verlag, New York.
ISBN 978-0-387-77316-2. (See the vignette for a package overview.)

LazyLoad yes

Depends R (>= 2.13.0), car (>= 2.0-1), lmtest, sandwich, survival, zoo

Suggests boot, dynlm, effects, foreign, ineq, KernSmooth, lattice,
MASS, mlogit, nlme, nnet, np, plm, pscl, quantreg, ROCR,
sampleSelection, scatterplot3d, strucchange, systemfit, rgl,
truncreg, tseries, urca

Imports stats, Formula (>= 0.2-0)

License GPL-2

Author Christian Kleiber [aut],
Achim Zeileis [aut, cre]

Maintainer Achim Zeileis <Achim.Zeileis@R-project.org>

NeedsCompilation no

Repository CRAN

Date/Publication 2014-01-28 17:50:48

R topics documented:

```
A. Observational data (adapted from Lab 3 Stat209)    Mroz data, returns to schooling
##### Examples from Wooldridge, Introductory Econometrics #######  Chapters 15 and 16
## stata results available from
http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge15.html

> lm.posWage = lm(logWage ~ educ)  > summary(lm.posWage)#session has full output
Coefficients:  #year increase in educ, fit increases 11 cents hourly wage
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1852     0.1852  -1.000    0.318
educ          0.1086     0.0144   7.545 2.76e-13 ***
Multiple R-Squared: 0.1179,     Adjusted R-squared: 0.1158
F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13
> # we get a highly significant slope (but not big Rsq),

> install.packages("AER") > library(AER) > help(ivreg)
> ivreg1 = ivreg(logWage ~ educ| fatheduc)
####### use the diagnostics option of ivreg (recent)
>  summary(ivreg1, diagnostics = TRUE)
Call: ivreg(formula = logWage ~ educ | fatheduc)
Coefficients: # matches tsls Task1
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.44110    0.44610   0.989   0.3233
educ         0.05917    0.03514   1.684   0.0929 .

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1 426     88.84  <2e-16 ***
Wu-Hausman         1 425      2.47   0.117
Sargan             0  NA        NA      NA
---
Residual standard error: 0.6894 on 426 degrees of freedom
Multiple R-Squared: 0.09344,    Adjusted R-squared: 0.09131
Wald test: 2.835 on 1 and 426 DF,  p-value: 0.09294
> confint(ivreg1)   # parameter confidence intervals as usual
                  2.5 %    97.5 %
(Intercept) -0.433239996 1.3154468
educ        -0.009703131 0.1280501

##Investigate the 'Weak instruments' entry, matches cor.test
> cor(educ, fatheduc)
[1] 0.415403

## just a test (like in the main lab of correlation between predictor and instrument

# the Hausman test is often used to show there's a difference between OLS and IV,
# but what does that tell you?
#  ivreg diagnostics finds discrepancy non-signicance even though IV est cuts the OLS
#  estimate in half  and IV estimate now non-significant
## to see the details of the Hausman test (and do it by hand) see pdf page 46 onward from
http://personal.rhul.ac.uk/uhte/006/ec2203/Lecture%2015_IVestimation.pdf
also the Basel class notes linked in week 6

## The Sargan test is a statistical test used for testing over-identifying restrictions
## relevant to simultaneous eqs, which we investigated in identifiability in lab script

# some package ivmodel display in week9 RQ

##############################################################################################
in Rogosa session    B. Encouragement Design   C. Compliance Adjustments
```

IV sessions Week 9 Stat266 D Rogosa

```
A. Observational data (adapted from Lab 3 Stat209)
################################################################################
> mroz87 =  read.table( "http://statweb.stanford.edu/~rag/stat209/Mroz87.dat", header = T)
> names(mroz87)
 [1] "lfp"      "hours"    "kids5"    "kids618"  "age"      "educ"
 [7] "wage"     "repwage"  "hushrs"   "husage"   "huseduc"  "huswage"
[13] "faminc"   "mtr"      "motheduc" "fatheduc" "unem"     "city"
[19] "exper"    "nwifeinc" "wifecoll" "huscoll"
> # Variable definition in lab script also linked, main page

# lfp Dummy variable for labor-force participation.
# hours Wife's hours of work in 1975.
# kids5 Number of children 5 years old or younger.
# kids618 Number of children 6 to 18 years old.
# age Wife's age.
# educ Wife's educational attainment, in years.
# wage Wife's average hourly earnings, in 1975 dollars.
# repwage Wife's wage reported at the time of the 1976 interview.
# hushrs Husband's hours worked in 1975.
# husage Husband's age.
# huseduc Husband's educational attainment, in years.
# huswage Husband's wage, in 1975 dollars.
# faminc Family income, in 1975 dollars.
# mtr Marginal tax rate facing the wife.
# motheduc Wife's mother's educational attainment, in years.
# fatheduc Wife's father's educational attainment, in years.
# unem Unemployment rate in county of residence, in percentage points.
# city Dummy variable = 1 if live in large city, else 0.
# exper Actual years of wife's previous labor market experience.
# nwifeinc Non-wife income.
# wifecoll Dummy variable for wife's college attendance.
# huscoll Dummy variable for husband's college attendance.

> poswage = subset(mroz87, wage > 0)  # my new data subset
> poswage$logWage =  log( poswage$wage ) # adding logWage to the data set for session
> attach(poswage)
> length(logWage)
[1] 428
> table(lfp) #all the women in this subset are in the workforce

> # onto fitting regression (predictor educ)
> lm.posWage = lm(logWage ~ educ)
> summary(lm.posWage)

Call:
lm(formula = logWage ~ educ)

Residuals:
     Min       1Q   Median       3Q      Max
-3.10256 -0.31473  0.06434  0.40081  2.10029

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1852     0.1852  -1.000    0.318
educ          0.1086     0.0144   7.545 2.76e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.68 on 426 degrees of freedom
Multiple R-Squared: 0.1179,     Adjusted R-squared: 0.1158
F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13

> # we get a highly significant slope (but not big Rsq),
> # year increase in educ, fit increases 11 cents hourly wage
> cor(logWage,educ)^2  # R-squared for the OLS equation
[1] 0.1178826


> # now the IV fit using fatheduc as instrument (omitted vars concern)
> cor.test(educ, fatheduc)
        Pearson's product-moment correlation
data:  educ and fatheduc
t = 9.4255, df = 426, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3337579 0.4908623
sample estimates:
      cor
0.4154030

##### Examples from Wooldridge, Introductory Econometrics
#######  Chapters 15 and 16
## stata results available from
 http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge15.html

> install.packages("AER")
> library(AER)
> help(ivreg)

> ivreg1 = ivreg(logWage ~ educ| fatheduc)
> summary(ivreg1)

Call:
ivreg(formula = logWage ~ educ | fatheduc)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0870 -0.3393  0.0525  0.4042  2.0677

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.44110    0.44610   0.989    0.323
educ         0.05917    0.03514   1.684    0.093 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6894 on 426 degrees of freedom
Multiple R-Squared: 0.09344,    Adjusted R-squared: 0.09131
Wald test: 2.835 on 1 and 426 DF,  p-value: 0.09294

> # matches tsls Task1

####### use the diagnostics option of ivreg (recent)
>  ivreg1 = ivreg(logWage ~ educ| fatheduc)
>  summary(ivreg1, diagnostics = TRUE)

Call:
```

```
ivreg(formula = logWage ~ educ | fatheduc)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0870 -0.3393  0.0525  0.4042  2.0677

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.44110    0.44610   0.989   0.3233
educ         0.05917    0.03514   1.684   0.0929 .

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1 426     88.84  <2e-16 ***
Wu-Hausman         1 425      2.47   0.117
Sargan             0  NA        NA      NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6894 on 426 degrees of freedom
Multiple R-Squared: 0.09344,    Adjusted R-squared: 0.09131
Wald test: 2.835 on 1 and 426 DF,  p-value: 0.09294

> confint(ivreg1)   # parameter confidence intervals as usual
                  2.5 %    97.5 %
(Intercept) -0.433239996 1.3154468
educ        -0.009703131 0.1280501

##Investigate the 'Weak instruments' entry, matches cor.test
> cor(educ, fatheduc)
[1] 0.415403

## just a test (like in the main lab of correlation between predictor and instrument

# the Hausman test is often used to show there's a difference between OLS and IV,
# but what does that tell you?
#  ivreg diagnostics finds discrepancy non-signicance even though IV est cuts the OLS estima
#  and IV estimate non-significant
## to see the details of the Hausman test (and do it by hand) see pdf page 46 onward from
http://personal.rhul.ac.uk/uhte/006/ec2203/Lecture%2015_IVestimation.pdf
also the Basel class notes linked in week 6

## The Sargan test is a statistical test used for testing over-identifying restrictions
## relevant to simultaneous eqs, which we investigated in identifiability in lab script

# some package ivmodel display in week9 RQ
################################################################################
################################################################################
B. Encouragement Design

> library(foreign)
> ses = read.dta("http://www.stat.columbia.edu/~gelman/arm/examples/sesame/sesame.dta")
> names(ses)
 [1] "rownames"  "id"        "site"      "sex"       "age"       "viewcat"   "setting"   "vi
[18] "postnumb"  "postrelat" "postclasf" "peabody"   "agecat"    "encour"    "_Isite_2"  "_I

> dim(ses)
[1] 240  28
> attach(ses)
> table(encour)
```

Encouragement Designs

Holland (1988) Powers & Swinton (1984)

Intervention: Encouragement to study (for a test)

random assignment to treatment - control
G=1 G=0

Student studies amount R

Student outcome, achievement test score, Y

For each unit observe:

Y, R, G

Other examples: Behavioral Medicine, salt ex

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

Hypertension
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Feasibility and efficacy of sodium reduction in the Trials of Hypertension Prevention, phase I. Trials of Hypertension Prevention Collaborative Research Group

SK Kumanyika, PR Hebert, JA Cutler, VI Lasser, CP Sugars, L Steffen-Batey, AA Brewer, M Cameron, LD Shepek and NR Cook

Hypertension 1993;22;502-512
Hypertension is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514

 Print ISSN: 0194-911X. Online ISSN: 1524-4563

Feasibility and Efficacy of Sodium Reduction in the Trials of Hypertension Prevention, Phase I

Shiriki K. Kumanyika, Patricia R. Hebert, Jeffrey A. Cutler, Vera I. Lasser, Carolyn P. Sugars, Lyn Steffen-Batey, Amy A. Brewer, Mary Cameron, Lana D. Shepek, Nancy R. Cook, Stephen T. Miller
for the Trials of Hypertension Prevention Collaborative Research Group

Phase I of the Trials of Hypertension Prevention was a multicenter, randomized trial of the feasibility and efficacy of seven nonpharmacologic interventions, including sodium reduction, in lowering blood pressure in 30- to 54-year-old individuals with a diastolic blood pressure of 80 to 89 mm Hg. Six centers tested an intervention designed to reduce dietary sodium to 80 mmol (1800 mg)/24 h with a total of 327 active intervention and 417 control subjects. The intervention consisted of eight group and two one-to-one meetings during the first 3 months, followed by less-intensive counseling and support for the duration of the study. The mean net decrease in sodium excretion was 43.9 mmol/24 h at 18 months. Women had lower sodium intake at baseline and were therefore more likely to decrease to less than 80 mmol/24 h. Black subjects were less likely to decrease to less than 80 mmol/d, independent of sex or baseline sodium excretion. The mean (95% confidence interval) net decrease associated with treatment was −2.1 (−3.3, −0.8) mm Hg for systolic blood pressure and −1.2 (−2.0, −0.3) mm Hg for diastolic blood pressure at 18 months (both $P<.01$). Multivariate analyses indicated a larger systolic blood pressure effect in women (−4.44 versus −1.23 mm Hg in men), adjusted for age, race, baseline blood pressure, and baseline 24-hour urinary sodium excretion ($P=.02$). Dose-response analyses indicated an adjusted decrease of −1.4 mm Hg for systolic blood pressure and −0.9 mm Hg for diastolic blood pressure for a decrease of 100 mmol/24 h in 18-month sodium excretion. These results support the utility of sodium reduction as a population strategy for hypertension prevention and raise questions about possible differences in dose response associated with gender and initial level of sodium intake. (*Hypertension*. 1993;22:502-512.)

KEY WORDS • **hypertension, sodium-dependent • blood pressure • sodium, dietary • primary prevention • blacks • women**

Primary prevention of hypertension, ie, preventing people at risk of hypertension from developing it, could potentially lower death rates from cardiovascular disease, reduce the need for antihypertensive medications, and reduce hypertension-related medical costs and job absenteeism.[1-4] However, the feasibility and efficacy of sodium reduction in lowering blood pressure among people with normal blood pressure in the general population have not been clearly established. Several lines of evidence suggest that sodium reduction is a logical candidate for incorporation into a primary prevention strategy. The well-established blood pressure–lowering effect of sodium reduction as a component of hypertension treatment is not confined to blood pressures above the physiologically arbitrary cutoffs used to define high blood pressure. Small, controlled trials have demonstrated the efficacy of moderate sodium reduction in reducing blood pressure in normotensive and hypertensive people.[5,6] In addition, data from the INTERSALT study have confirmed that, over a wide range of sodium intake, populations with low sodium consumption have lower blood pressures than those with high sodium consumption.[7,8]

Phase I of the Trials of Hypertension Prevention (TOHP-I) was a multicenter, randomized trial designed to test the short-term efficacy and safety of several nonpharmacologic interventions, including sodium reduction, in reducing blood pressure in a large cohort of men and women without hypertension in the context of their usual patterns of living.[9] TOHP-I attempted to provide an unambiguous answer regarding the potential for lowering the average blood pressure levels in the

Received December 28, 1992; accepted in revised form June 4, 1993.

From the Department of Epidemiology, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md (S.K.K., L.S.-B.); Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass (P.R.H., N.R.C.); Prevention and Demonstration Research Branch, Division of Epidemiology and Clinical Applications, National Heart, Lung, and Blood Institute, Bethesda, Md (J.A.C.); Preventive Cardiology Program, Department of Medicine, New Jersey Medical School, Newark (V.I.L.); Department of Internal Medicine, University of California, Davis (C.P.S.); Department of Preventive Medicine, University of Tennessee, Memphis (A.A.B., S.T.M.); University of Mississippi, Jackson, (M.C.); and St Louis University School of Medicine (Mo) (L.D.S.).

A preliminary version of these results was presented at the 64th Scientific Sessions of the American Heart Association, Anaheim, Calif, November 1991, and the International Heart Health Conference, Vancouver, British Columbia, Canada, May 1992.

Correspondence to Shiriki Kumanyika, Center for Biostatistics and Epidemiology, College of Medicine, Pennsylvania State University, PO Box 850, Hershey, PA 17033.

Other Examples of Encouragement Experiments:

Outcome	Treatment	Instrumental Variable	Reference
Earnings	Participation in federal job training program	Random assignment of admission to training program	Bloom et al. (1997)
Achievement test scores	Enrollment in private school	Randomly selected offer of school voucher	Howell et al. (2000)
Achievement test scores	Class size	Random assignment to a small or normal-size class	Krueger (1999)
Depression Level	Meeting with depression specialist	Random encouragement to meet with depression specialist	Small et al. (2007)
Achievement test scores	Hours of study	Random mailing of test preparation materials	Powers and Swinton (1984)

Causal Effect Being Estimated When Treatment Effect is Not Constant

Suppose that we have an IV that satisfies the exclusion restriction but treatment effects are not constan

$$r_{(z,s)i} = r_{(0,0)i} + \beta_i s$$

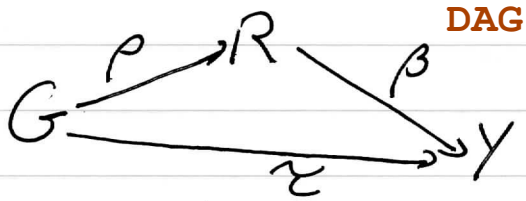
Consider the case of a binary treatment S. We can divide the subjects into four groups:

(1) $s_{(1)i} = 0,\ s_{(0)i} = 0$ (Never takers): subjects who would take not take treatment, regardless whether encouraged to do so

Encouragement Design Estimation (Holland 1988, p471)

Setting
Salt, BP
random assignment
to G=1 workshops, nagging
G=0 none
observe R salt intake
Y BP

Individual Picture
(Holland ALICE)



Individual Potential outcomes

$R_G(u) = R_c(u) + \rho G$

$Y_{Gr}(u) = Y_{c0}(u) + \gamma G + \beta r$

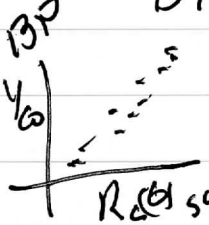
Regression Equations (over persons)

$R = \alpha_R + \gamma_1 G + \varepsilon_R \; ; \; Y = \alpha_y + \gamma_2 R + \gamma_3 G + \varepsilon_y$

Results for regression coefficients (Holland 1988)

$\gamma_1 = \rho \qquad \gamma_2 = \beta + \delta \qquad \gamma_3 = \tau - \rho\delta$

bias $\delta > 0$ association



other healthy habits, attitudes

salt intake when not encouraged

BP outcome when not encouraged and no reduction in salt

IF $\tau \equiv 0$ then algebra says use regression coefficients to solve

$\beta = \gamma_2 + \gamma_3/\gamma_1 = \beta + \delta - \delta = \beta$

Instrumental Variables IV (week 6) requires

$\tau = 0$ (basic IV assumption) G is instrument for R

$$\hat{\beta} = \frac{\hat{\beta}_{YG}}{\hat{\beta}_{RG}} = \frac{\bar{Y}_t - \bar{Y}_c}{\bar{R}_t - \bar{R}_c}$$

Wald estimator

Link between IV (Wald) and Holland estimator (tau = 0)

Barron-Kenney est $\hat{\gamma}_1 \hat{\gamma}_3$ path analysis indirect effect.

for $\tau = 0$

$$\beta = \gamma_2 + \gamma_3/\gamma_1 = \beta_{YR \cdot G} + \beta_{YG \cdot R}/\beta_{RG}$$

weak

Regression recursion

$$\beta_{YG \cdot R} = \beta_{YG} - \beta_{RG}\beta_{YR \cdot G}$$

Substitute

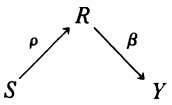
$$\gamma_2 + \gamma_3/\gamma_1 = \beta_{YR \cdot G} + \frac{\beta_{YG} - \beta_{RG}\beta_{YR \cdot G}}{\beta_{RG}}$$

$$= \beta_{YG}/\beta_{RG}$$

estimate by $\dfrac{\bar{Y}_1 - \bar{Y}_0}{\bar{R}_1 - \bar{R}_0}$ Wald estimator

Estimating beta (tau = 0 assume)

FIGURE 7.



S==G
IV for R

Hence,

$$\beta = \frac{\rho\beta}{\rho} = \frac{\text{total effect of } S \text{ on } Y_{SR_S}}{\text{total effect of } S \text{ on } R_S}. \tag{56}$$

This is also easily seen from the definitions of the ACEs and the FACEs. Under the assumption that $\tau = 0$,

$$\text{ACE}_{tc}(Y) = \text{FACE}_{tc}(Y) = \beta\rho, \tag{57}$$

regardless of whether or not $\mu_c(r)$ is linear, and hence

$$\beta = \frac{\text{ACE}_{tc}(Y)}{\text{ACE}_{tc}(R)} = \frac{\text{FACE}_{tc}(Y)}{\text{FACE}_{tc}(R)}. \tag{58}$$

The two FACEs in (58) may be estimated simply by the treatment-control mean difference in Y_{SR_S} and R_S, as mentioned earlier, so that (58) provides an alternative way to estimate β that does not assume that $\mu_c(r)$ is constant. In Powers and Swinton (1984), (58) was used to estimate β.

4.4. *Deriving a Structural Equations Model*

The ALICE model may be used to derive the structural equations model given in (9) and (10). If we substitute $S(u)$ for s in (40) and $S(u)$ for s and $R_S(u)$ for r in (41), we get the following pair of equations that involve the observables, S, R_S, Y_{SR_S}:

$$R_S(u) = R_c(u) + \rho S(u) \tag{59}$$

and

$$Y_{SR_S}(u) = Y_{c0}(u) + \tau S(u) + \beta R_S(u). \tag{60}$$

Now let

$$\eta_1(u) = R_c(u) - E(R_c)$$

and

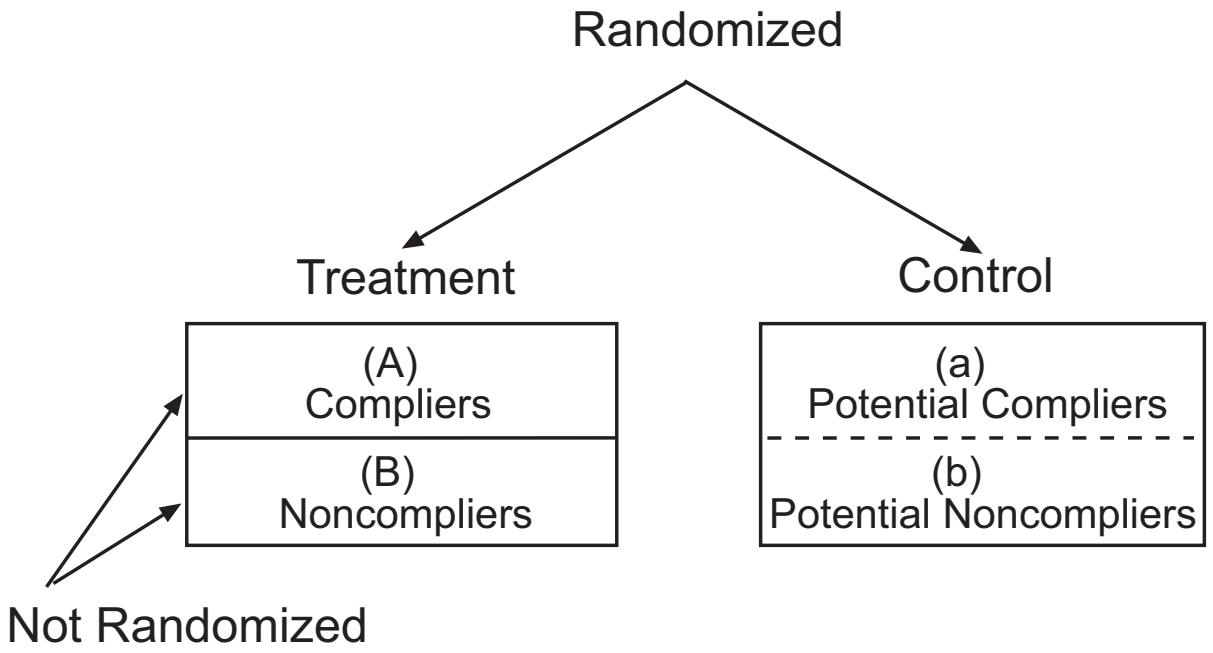
$$\eta_2(u) = Y_{c0}(u) - E(Y_{c0}),$$

```
ivreg(formula = logWage ~ educ | fatheduc)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0870 -0.3393  0.0525  0.4042  2.0677

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.44110    0.44610   0.989   0.3233
educ         0.05917    0.03514   1.684   0.0929 .

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1 426     88.84  <2e-16 ***
Wu-Hausman         1 425      2.47   0.117
Sargan             0  NA        NA      NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6894 on 426 degrees of freedom
Multiple R-Squared: 0.09344,    Adjusted R-squared: 0.09131
Wald test: 2.835 on 1 and 426 DF,  p-value: 0.09294

> confint(ivreg1)   # parameter confidence intervals as usual
                  2.5 %    97.5 %
(Intercept) -0.433239996 1.3154468
educ        -0.009703131 0.1280501

##Investigate the 'Weak instruments' entry, matches cor.test
> cor(educ, fatheduc)
[1] 0.415403

## just a test (like in the main lab of correlation between predictor and instrument

# the Hausman test is often used to show there's a difference between OLS and IV,
# but what does that tell you?
#  ivreg diagnostics finds discrepancy non-signicance even though IV est cuts the OLS estima
#  and IV estimate non-significant
## to see the details of the Hausman test (and do it by hand) see pdf page 46 onward from
http://personal.rhul.ac.uk/uhte/006/ec2203/Lecture%2015_IVestimation.pdf
also the Basel class notes linked in week 6

## The Sargan test is a statistical test used for testing over-identifying restrictions
## relevant to simultaneous eqs, which we investigated in identifiability in lab script

# some package ivmodel display in week9 RQ
################################################################################################
################################################################################################
B. Encouragement Design

> library(foreign)
> ses = read.dta("http://www.stat.columbia.edu/~gelman/arm/examples/sesame/sesame.dta")
> names(ses)
 [1] "rownames"  "id"        "site"      "sex"       "age"       "viewcat"   "setting"   "vi
[18] "postnumb"  "postrelat" "postclasf" "peabody"   "agecat"    "encour"    "_Isite_2"  "_I

> dim(ses)
[1] 240  28
> attach(ses)
> table(encour)
```

```
encour
  0   1
 88 152
> s1 = ivreg(postlet ~ as.numeric(viewcat) | encour, x = TRUE)
> summary(s1, diagnostics = TRUE)

Call:
ivreg(formula = postlet ~ as.numeric(viewcat) | encour, x = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max
-20.136  -8.830  -3.765   8.864  29.864

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          15.394      6.270   2.455   0.0148 *
as.numeric(viewcat)   4.435      2.433   1.823   0.0695 .

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1 238    20.823 8.07e-06 ***
Wu-Hausman         1 237     0.456      0.5
Sargan             0  NA        NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.78 on 238 degrees of freedom
Multiple R-Squared: 0.2282,     Adjusted R-squared: 0.225
Wald test: 3.324 on 1 and 238 DF,  p-value: 0.06954
> confint(s1)
                         2.5 %    97.5 %
(Intercept)          3.1048997 27.683846
as.numeric(viewcat) -0.3328545  9.203703

> install.packages("ivpack")
> library(ivpack)
> robust.se(s1)
[1] "Robust Standard Errors"

t test of coefficients:

                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          15.3944     6.3210  2.4354  0.01561 *
as.numeric(viewcat)   4.4354     2.4437  1.8150  0.07078 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anderson.rubin.ci(s1)
$confidence.interval
[1] "[ -1.28052883742684 , 9.37441745247359 ]"

>
```

```
##############################################################################################
##############################################################################################
C. Compliance Adjustments
# binary compliance or measured compliance, week 9 RQ1
Artificial data in the image of Efron-Feldman
# cholesterol reduction outcome measure
```

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)

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

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

Week 7

Stat 209

Formulation for Complicence Analyses

Bool Jo version

assume no controls get treatment

DAP: single cross-over Vitamin A

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

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

David Freedman Analysis (model obs talk, OxCaus 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

Ave response it in

IV estimate (pp. 704-5)

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

Stat 209 [redacted]

Formulation for Compliance Analyses

Bookl Jo version — assume no controls get treatment

DAP: 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: $\frac{\hat{ITT}}{\hat{\pi}_c} = \frac{.364}{.457} = .76$ IV est, AIR

c.f AIR handout 2/19

David Freedman Analysis (modelobs talk, oxcausesp paper)

Neyman model

Potential outcomes: over all in population

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

randomize to T, C

Ave response if in

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

IV estimate (pp. 704-5)

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

result $\hat{c} = (\bar{Y}^c - \hat{\gamma}\hat{\eta})/\hat{\beta}$, $\hat{\tau} = (\bar{Y}^T - \hat{\gamma}\eta)/\hat{\beta} \Rightarrow \underbrace{\hat{\tau} - \hat{c}}_{CACE} = \frac{\bar{Y}^T - \bar{Y}^c}{\hat{\beta}}$ (IV 706-)

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) $\quad \lambda = N_T/N_C$

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

p. 706.

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)

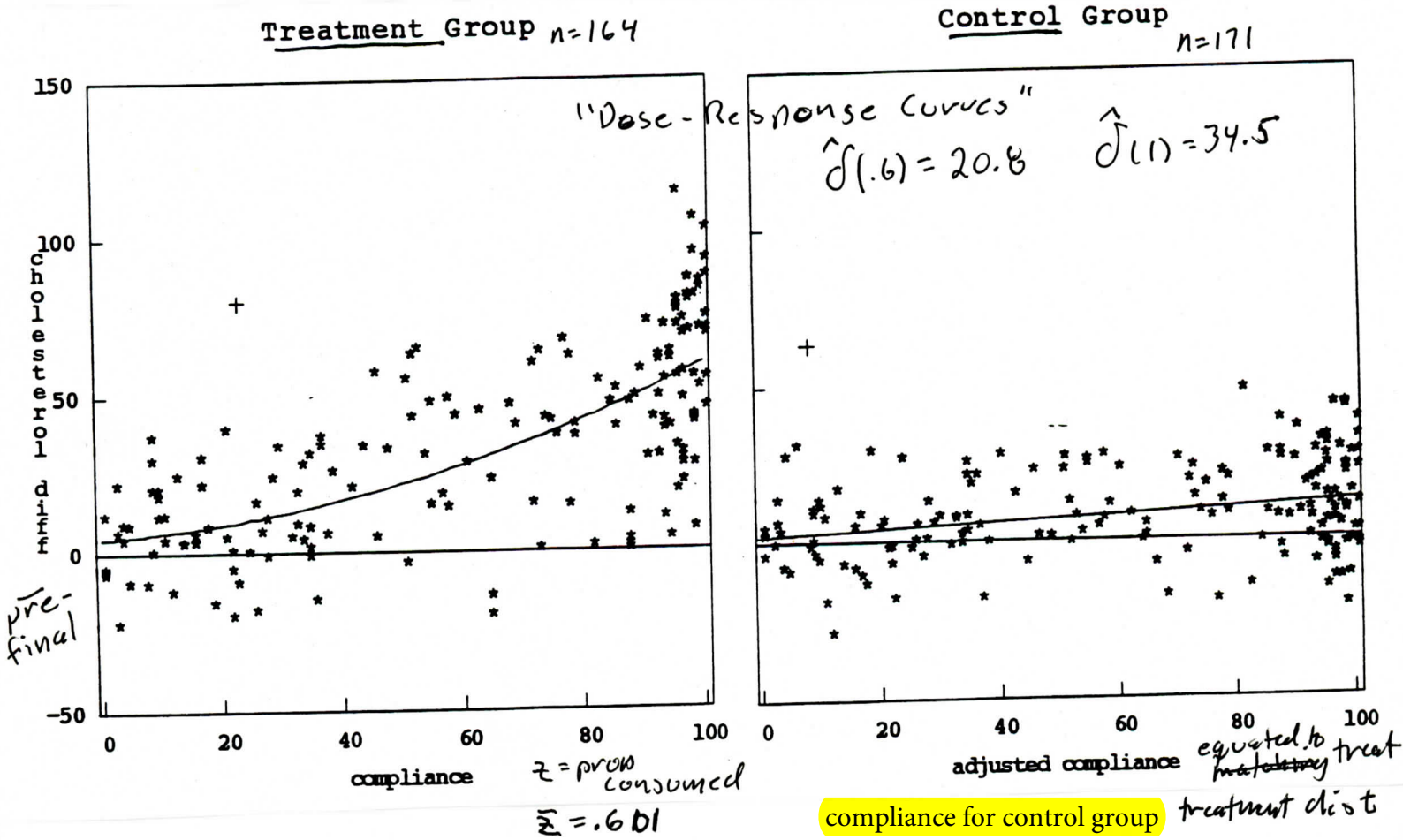
note $24.52/.601 = 40.8$

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



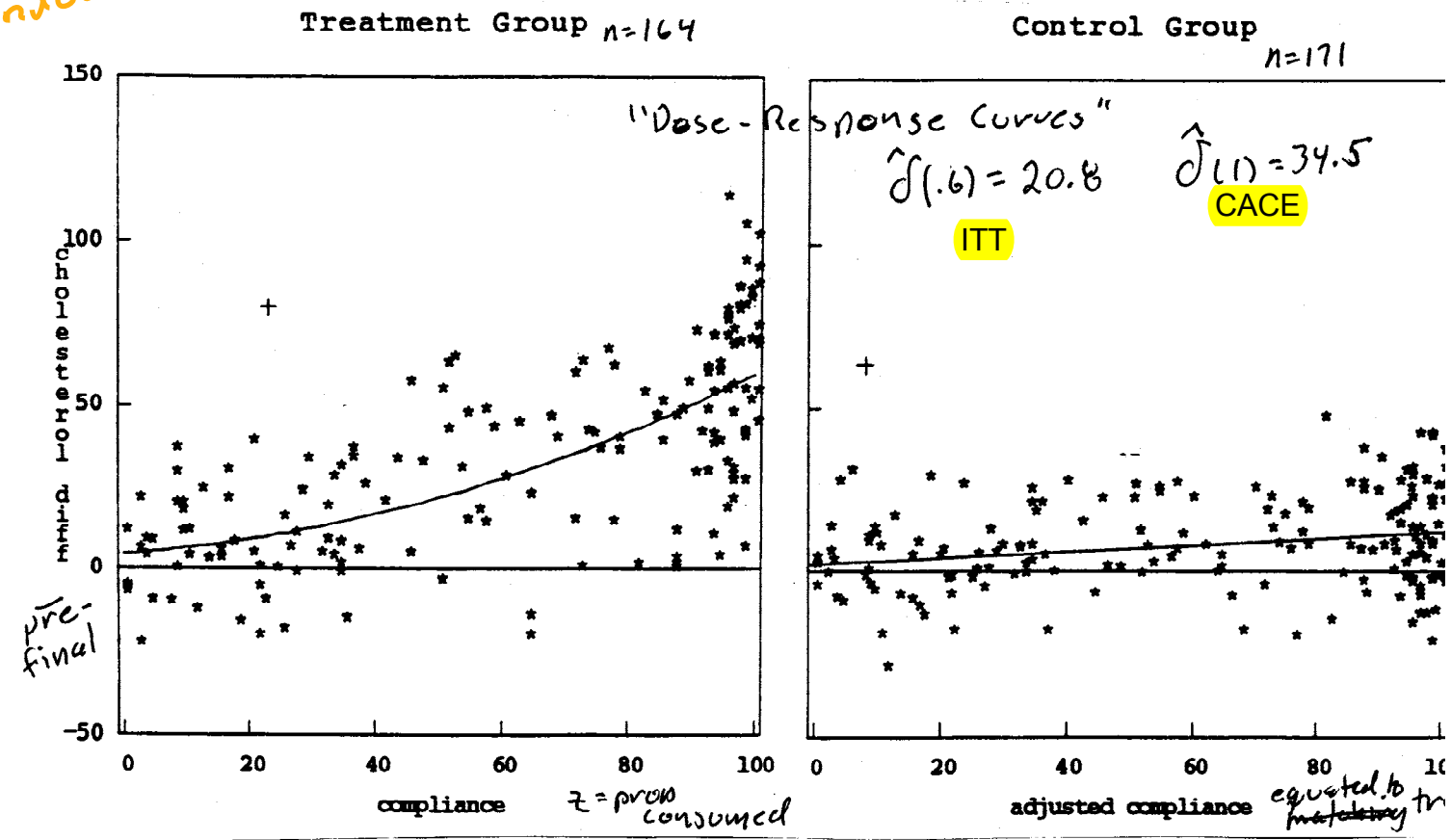
Example 4 Efron Feldman dose response

- No consistent method for handling withdrawal

Compliance as an Explanatory Variable in Clinical Trials

B. Efron and D. Feldman

P2 random



```
encour
  0   1
 88 152
> s1 = ivreg(postlet ~ as.numeric(viewcat) | encour, x = TRUE)
> summary(s1, diagnostics = TRUE)

Call:
ivreg(formula = postlet ~ as.numeric(viewcat) | encour, x = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max
-20.136  -8.830  -3.765   8.864  29.864

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          15.394      6.270   2.455   0.0148 *
as.numeric(viewcat)   4.435      2.433   1.823   0.0695 .

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1 238    20.823 8.07e-06 ***
Wu-Hausman         1 237     0.456      0.5
Sargan             0  NA        NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.78 on 238 degrees of freedom
Multiple R-Squared: 0.2282,     Adjusted R-squared: 0.225
Wald test: 3.324 on 1 and 238 DF,  p-value: 0.06954
> confint(s1)
                         2.5 %    97.5 %
(Intercept)          3.1048997 27.683846
as.numeric(viewcat) -0.3328545  9.203703

> install.packages("ivpack")
> library(ivpack)
> robust.se(s1)
[1] "Robust Standard Errors"

t test of coefficients:

                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          15.3944     6.3210  2.4354  0.01561 *
as.numeric(viewcat)   4.4354     2.4437  1.8150  0.07078 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anderson.rubin.ci(s1)
$confidence.interval
[1] "[ -1.28052883742684 , 9.37441745247359 ]"

>
```

```
##############################################################################################
##############################################################################################
C. Compliance Adjustments
# binary compliance or measured compliance, week 9 RQ1
Artificial data in the image of Efron-Feldman
# cholesterol reduction outcome measure
```

```
> library(AER)
> efdat = read.table(file = "http://web.stanford.edu/~rag/stat209/hw7efdata", header = T)
> attach(efdat)
> head(efdat)
   comp G     Y
1 0.528 0 -9.57
2 0.862 1 58.50
3 0.980 0 10.10
4 0.673 1 33.90
5 0.660 0  3.60
6 0.551 1 29.10
> table(G)
G
  0   1
150 150
###### got data

> t.test(Y ~ G)  # ITT estimate about 20, (not using compliance info)
# approximating E-F paper analysis

        Welch Two Sample t-test

data:  Y by G
t = -14.9477, df = 275.264, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.52955 -17.28584
sample estimates:
mean in group 1 mean in group 2
       9.663234       29.570932

# try an IV with compliance as a measured variable
# and random assignment as instrument (analog to CACE)
# Doesn't seem to work well, but see below

> caceIV = ivreg(Y ~ comp|G)
> confint(caceIV)
                2.5 %   97.5 %
(Intercept) -4262.740 2386.813
comp        -3991.215 7220.521  # pretty wide
> summary(caceIV)

Call:
ivreg(formula = Y ~ comp | G)

Residuals:
    Min      1Q  Median      3Q     Max
-634.86 -202.04   22.26  196.10  840.55

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     -938       1696  -0.553    0.581
comp            1615       2860   0.565    0.573

Residual standard error: 305.4 on 298 degrees of freedom
Multiple R-Squared: -399.6,     Adjusted R-squared: -400.9
Wald test: 0.3187 on 1 and 298 DF,  p-value: 0.5728

# but as you probably presumed that's not really the right analysis
# the research question is, what is the effect of the drug?
```

```
# the correct IV picture is as displayed in the Greenland handout
# in the treatment group dose of the drug is confounded with compliance
# in the control group dose of the drug is 0 (regardless of compliance level)
# so the correct analog to AIR is
> caceIV2 = ivreg(Y ~ I(comp*G)|G)
> summary(caceIV2)

Call:
ivreg(formula = Y ~ I(comp * G) | G)

Residuals:
     Min       1Q   Median       3Q      Max
-28.5645  -6.1078  -0.2812   6.5362  29.6355

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.6645     0.8059   11.99   <2e-16 ***
I(comp * G)  33.2189     1.9021   17.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.871 on 298 degrees of freedom
Multiple R-Squared: 0.5814,     Adjusted R-squared:  0.58
Wald test:   305 on 1 and 298 DF,  p-value: < 2.2e-16

> # not far from the E-F paper estimate of 34.5 for ("CACE") effect (cholesterol reduction)
(these data are an imitation not exact recreation of E-F)
> confint(caceIV2)
                2.5 %   97.5 %
(Intercept)  8.084849 11.24406
I(comp * G) 29.490842 36.94686
> # nice, tight interval

################################################
# Now go back to conventional Rubin approach with compliance as 0,1

> efdat$compT = efdat$comp > .8  # comp indicator
# calculate simpler form of CACE with traditional .8 cutpoint,
# yields with low compliance, 15%, maybe not the best choice of cut
> mean(efdat$compT)
[1] 0.15

> 20/.15  # divide ITT by proportion defined as compliant
[1] 133.3333
> # wow

> caceIVT = ivreg(Y ~ compT|G, data = efdat)
> summary(caceIVT)
Call:
ivreg(formula = Y ~ compT | G, data = efdat)

Residuals:
   Min     1Q Median     3Q    Max
-872.8  131.8  143.8  155.9  185.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -129.7      306.5  -0.423    0.673
compTTRUE      995.3     2039.0   0.488    0.626  # good bit diff than ITT of 20, CACE above
```

Wald estimator

```
Residual standard error: 353.2 on 298 degrees of freedom
Multiple R-Squared: -534.8,     Adjusted R-squared: -536.6
Wald test: 0.2383 on 1 and 298 DF,  p-value: 0.6258


> confint(caceIVT)
                 2.5 %    97.5 %
(Intercept)  -730.4593  471.1126
compTTRUE   -3001.0986 4991.6386       # here shows CACE can be anywhere for IV version

# but as in above the right IV analysis uses dose (which equals compT in treat, 0 in control
> efdat$compT = efdat$comp > .8  # comp indicator
> attach(efdat)
> caceIV2 = ivreg(Y ~ I(compT*G)|G) # pretending dose is (0,1)

> summary(caceIV2)  # gives about usual case, wayoff from E-F effect for perfect compliance

Call:
ivreg(formula = Y ~ I(compT * G) | G)

Residuals:
     Min       1Q   Median       3Q      Max
-114.173   -4.502    5.436   16.561   46.436

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     9.664      2.447   3.950 9.78e-05 ***
I(compT * G)  124.409     21.628   5.752 2.18e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.97 on 298 degrees of freedom
Multiple R-Squared: -2.858,     Adjusted R-squared: -2.871
Wald test: 33.09 on 1 and 298 DF,  p-value: 2.184e-08

> confint(caceIV2)
                 2.5 %    97.5 %
(Intercept)   4.868521  14.46039
I(compT * G) 82.018297 166.79920
>

> cor(efdat$G, efdat$compT)
[1] 0.0280056

# G a pretty weak instrument here
> cor.test(efdat$G, as.numeric(efdat$compT))

        Pearson's product-moment correlation

data:  efdat$G and as.numeric(efdat$compT)
t = 0.4836, df = 298, p-value = 0.629
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.08550641  0.14079991
sample estimates:
      cor
0.0280056

> cor.test(efdat$G, efdat$comp)
```

```
        Pearson's product-moment correlation

data:  efdat$G and efdat$comp
t = 0.5565, df = 298, p-value = 0.5783
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.08131855  0.14493103
sample estimates:
       cor
0.03221898
```