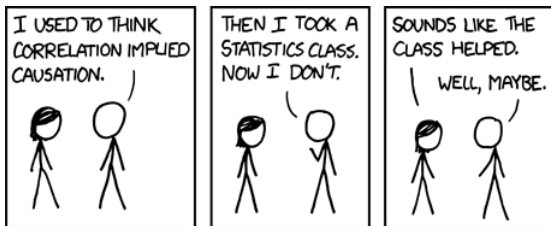


Advanced Applied Econometrics

Teacher: Felix Weinhardt



Slides heavily borrow from Mixed Tape
and Mostly Harmless Econometrics textbooks

Angrist (1990) Veteran Draft Lottery

- Angrist (1990) uses the Vietnam draft lottery as an instrumental variable for military service
- In the 1960s and 1970s, young American men were drafted for military service to serve in Vietnam
- Concerns about the fairness of the conscription policy led to the introduction of a draft *lottery* in 1970
- From 1970 to 1972, random sequence numbers were randomly assigned to each birth date in cohorts of 19-year-olds
- Men with lottery numbers below a cutoff were drafted; in other words
 - Higher numbers were *less* likely to be drafted;
 - Lower numbers were *more* likely to be drafted.
- The draft did not perfectly determine military service:
 - Many draft-eligible men were exempt for health and other reasons
 - Exempt men would sometimes volunteer

Summary of Findings on Vietnam Draft Lottery

- 1 First stage results: Having a low lottery number (i.e., being eligible for the draft) increased veteran status by about 16 percentage points (the mean of veteran status was 27 percent)
- 2 Second stage results: Serving in the army lowers earnings by between \$2,050 and \$2,741 per year

IV with Heterogenous Treatment Effects

- Up to this point, we only considered models where the causal effect was the same for all individuals (i.e., homogenous treatment effects where $Y_i^1 - Y_i^0 = \delta$ for all i units)
- Let's now try to understand what instrumental variables estimation is measuring if treatment effects are *heterogenous* (i.e, $Y_i^1 - Y_i^0 = \delta_i$ which varies across the population)
- Why do we care?
 - 1 We care about internal validity: Does the design successfully uncover causal effects for the population that we are studying?
 - 2 We care about external validity: Does the study's results inform us about different populations?

IV with Heterogenous Treatment Effects

- Similar potential outcomes notation and terminology:
 - Causal chain is $Z_i \rightarrow D_i \rightarrow Y_i$
 - $Y_i(D_i = 0, Z_i = 1)$ is represented as $Y_i(0, 1)$
- “Potential treatment status” versus “observed treatment status”
 - $D_i^1 = i$'s treatment status when $Z_i = 1$
 - $D_i^0 = i$'s treatment status when $Z_i = 0$
 - Observed treatment status switching equation:

$$\begin{aligned} D_i &= D_i^0 + (D_i^1 - D_i^0)Z_i \\ &= \pi_0 + \pi_1 Z_i + \zeta_i \end{aligned}$$

$$\pi_{0i} = E[D_i^0]$$

$$\pi_{1i} = (D_i^1 - D_i^0) \text{ is the heterogenous causal effect of the IV on } D_i.$$

$$E[\pi_{1i}] = \text{The average causal effect of } Z_i \text{ on } D_i$$

Identifying assumptions under heterogenous treatment effects

- 1 Stable Unit Treatment Value Assumption (SUTVA)
- 2 Random Assignment
- 3 Exclusion Restriction
- 4 Nonzero First Stage
- 5 Monotonicity

Stable Unit Treatment Value Assumption (SUTVA)

Stable Unit Treatment Value Assumption (SUTVA)

If $Z_i = Z'_i$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$

If $Z_i = Z'_i$ and $D_i = D'_i$, then $Y_i(\mathbf{D}, \mathbf{Z}) = Y_i(\mathbf{D}', \mathbf{Z}')$

Interpretation Potential outcomes for each person i are unrelated to the treatment status of other individuals.

Example Veteran status of person at risk of being drafted is not affected by the draft status of others at risk of being drafted.

Implication Rewrite $Y_i(\mathbf{D}, \mathbf{Z})$ as $Y_i(D_i, Z_i)$ and $D_i(\mathbf{Z})$ as $D_i(Z_i)$.

Definition 1: Intention-to-treat effects

Causal effect of Z on D is $D_i^1 - D_i^0$

Causal effect of Z on Y is $Y_i(D_i^1, 1) - Y_i(D_i^0, 0)$

Independence assumption

Independence assumption (e.g., “as good as random assignment”)

$$\{Y_i(D_i^1, 1), Y_i(D_i^0, 0), D_i^1, D_i^0\} \perp\!\!\!\perp Z_i$$

Interpretation The IV is independent of the vector of potential outcomes and potential treatment assignments (i.e. “as good as randomly assigned”)

The independence assumption is sufficient for a causal interpretation of the reduced form:

$$\begin{aligned} E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] &= E[Y_i(D_i^1, 1)|Z_i = 1] - E[Y_i(D_i^0, 0)|Z_i = 0] \\ &= E[Y_i(D_i^1, 1)] - E[Y_i(D_i^0, 0)] \end{aligned}$$

Independence means that the first stage measures the causal effect of Z_i on D_i :

$$\begin{aligned} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] &= E[D_i^1|Z_i = 1] - E[D_i^0|Z_i = 0] \\ &= E[D_i^1 - D_i^0] \end{aligned}$$

Example Vietnam conscription for military service was based on randomly generated draft lottery numbers. The assignment of draft lottery number was independent of potential earnings or potential military service – as good as random.

Exclusion Restriction

Exclusion Restriction

$Y(D, Z) = Y(D, Z')$ for all Z, Z' , and for all D

Interpretation Any effect of Z on Y must be via the effect of Z on D . In other words, $Y_i(D_i, Z_i)$ is a function of D only. Or formally:

$$Y_i(D_i, 0) = Y_i(D_i, 1) \text{ for } D = 0, 1$$

Example In the Vietnam draft lottery example, an individual's earnings potential as a veteran or a non-veteran are assumed to be the same regardless of draft eligibility status. The exclusion restriction would be violated if low lottery numbers may have affected schooling (e.g., to avoid the draft). If this was the case, the lottery number would be correlated with earnings for at least two cases:

- ① through its effect on military service
- ② through its effect on educational attainment

Implication Random lottery numbers (independence) does not imply that the exclusion restriction is satisfied

Exclusion restriction

- Use the exclusion restriction to define potential outcomes indexed solely against treatment status:

$$Y_i^1 = Y_i(1, 1) = Y_i(1, 0)$$

$$Y_i^0 = Y_i(0, 1) = Y_i(0, 0)$$

- Rewrite the switching equation:

$$Y_i = Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i$$

$$Y_i = Y_i^0 + [Y_i^1 - Y_i^0]D_i$$

- Random coefficients notation for this is:

$$Y_i = \alpha_0 + \rho_i D_i$$

$$\text{with } \alpha_0 = E[Y_i^0] \text{ and } \rho_i = Y_i^1 - Y_i^0$$

First stage

Nonzero Average Causal Effect of Z on D

$$E[D_i^1 - D_i^0] \neq 0$$

Interpretation Z has to have some statistically significant effect on the average probability of treatment

Example Having a low lottery number increases the average probability of service.

Monotonicity

Monotonicity

Either $\pi_{1i} \geq 0$ for all i or $\pi_{1i} \leq 0$ for all $i = 1, \dots, N$

Interpretation Recall that $\pi + 1i$ is the reduced form causal effect of the instrumental variable on an individual i 's treatment status. Monotonicity requires that the instrumental variable (weakly) operate in the same direction on all individual units. In other words, while the instrument may have no effect on some people, all those who are affected are affected *in the same direction* (i.e., positively or negatively, but not both).

Draft example While draft eligibility may have had no effect on the probability of military service for some, monotonicity means the draft lottery either shifted people into service, or it shifted people out of service – but it did not do both.

Schooling example In the quarter of birth example for schooling, this assumption may not be satisfied (see Barua and Lang 2009). Being born in the 4th quarter (which typically increases schooling) may have reduced schooling for some because their school enrollment was held back by their parents

Implication Without monotonicity, IV estimators are not guaranteed to estimate a weighted average of the underlying causal effects of the affected group, $Y_i^1 - Y_i^0$.

Local average treatment effect

If all 1-5 assumptions are satisfied, then IV estimates the **local average treatment effect (LATE)** of D on Y :

$$\delta_{IV,LATE} = \frac{\text{Effect of } Z \text{ on } Y}{\text{Effect of } Z \text{ on } D}$$

Instrumental variables (IV) estimand:

$$\begin{aligned}\delta_{IV,LATE} &= \frac{E[Y_i(D_i^1, 1) - Y_i(D_i^0, 0)]}{E[D_i^1 - D_i^0]} \\ &= E[(Y_i^1 - Y_i^0) | D_i^1 - D_i^0 = 1]\end{aligned}$$

Local Average Treatment Effect

- The LATE parameters is the average causal effect of D on Y for those whose treatment status was changed by the instrument, Z
- Vietnam draft lottery example: IV estimates the average effect of military service on earnings for the subpopulation who enrolled in military service because of the draft but would not have served otherwise.
 - In other words, LATE would not tell us what the causal effect of military service was for volunteers or those who were exempted from military service for medical reasons
- We have reviewed the properties of IV with heterogenous treatment effects using a very simple dummy endogenous variable, dummy IV, and no additional controls example.
 - The intuition of LATE generalizes to most cases where we have continuous endogenous variables and instruments, and additional control variables.

More IV Jargon!

- The LATE framework partitions any population with an instrument into potentially 4 groups:
 - 1 Compliers: The subpopulation with $D_i^1 = 1$ and $D_i^0 = 0$. Their treatment status is affected by the instrument in the “correct direction”.
 - 2 Always takers: The subpopulation with $D_i^1 = D_i^0 = 1$. They always take the treatment independently of Z .
 - 3 Never takers: The subpopulation with $D_i^1 = D_i^0 = 0$. They never take the treatment independently of Z .
 - 4 Defiers: The subpopulation with $D_i^1 = 0$ and $D_i^0 = 1$. Their treatment status is affected by the instrument in the “wrong direction”.
- (You'll notice that increasingly our terms are borrowing from the medical literature where the treatment is taking a pill. Same here.)

Never-Takers

$$D_i^1 - D_i^0 = 0$$

$$Y_i(0,1) - Y_i(0,0) = 0$$

By **Exclusion Restriction**, causal effect of Z on Y is zero.

Defier

$$D_i^1 - D_i^0 = -1$$

$$Y_i(0,1) - Y_i(1,0) = Y_i(0) - Y_i(1)$$

By **Monotonicity**, no one in this group

Complier

$$D_i^1 - D_i^0 = 1$$

$$Y_i(1,1) - Y_i(0,0) = Y_i(1) - Y_i(0)$$

Average Treatment Effect among Compliers

Always-taker

$$D_i^1 - D_i^0 = 0$$

$$Y_i(1,1) - Y_i(1,0) = 0$$

By **Exclusion Restriction**, causal effect of Z on Y is zero.

Monotonicity Ensures that there are no defiers

- Monotonicity ensures that there are no defiers
- Why is it important to not have defiers?
 - If there were defiers, effects on compliers could be (partly) canceled out by opposite effects on defiers
 - One could then observe a reduced form which is close to zero even though treatment effects are positive for everyone (but the compliers are pushed in one direction by the instrument and the defiers in the other direction)

What Does IV (Not) Estimate?

- As outlined above, with all 5 assumptions satisfied, IV estimates the average treatment effect for *compliers*
 - Contrast this with the traditional IV pedagogy with constant treatment effects (i.e., $\delta_i = \delta$ for all i units).
 - Question: What does IV estimate when treatment effects are assumed to be constant?
- Without further assumptions (e.g., constant causal effects), LATE is not informative about effects on never-takers or always-takers because the instrument does not affect their treatment status
- So what? Well, it matters because in most applications, we would be mostly interested in estimating the average treatment effect on the whole population:

$$ATE = E[Y_i^1 - Y_i^0]$$

- But that's not possible usually with IV

Sensitivity to assumptions: exclusion restriction

Example Someone at risk of draft (low lottery number) changes education plans to retain draft deferments and avoid conscription.

Implication Increased bias to IV estimand through two channels:

- Average direct effect of Z on Y for compliers
- Average direct effect of Z on Y for noncompliers multiplied by odds of being a non-complier

Severity Depends on:

- Odds of noncompliance (smaller \rightarrow less bias)
- “Strength” of instrument (stronger \rightarrow less bias)
- Effect of the alternative channel on Y

Sensitivity to assumptions: Monotonicity violations

Example Someone who would have volunteered for Army when not at risk of draft (high lottery number) chooses to avoid military service when at risk of being drafted (low lottery number)

Implication Bias to IV estimand (multiplication of 2 terms):

- Proportion defiers relative to compliers
- Difference in average causal effects of D on Y for compliers and defiers

Severity Depends on:

- Proportion of defiers (small \rightarrow less bias)
- “Strength” of instrument (stronger \rightarrow less bias)
- Variation in effect of D on Y (less \rightarrow less bias)

Summarizing

- The potential outcomes framework gives a more subtle interpretation of what IV is measuring
 - In the constant coefficients world (i.e., traditional pedagogy), IV measures δ which is “the” causal effect of D_i on Y_i , and assumed to be the same for all i units
 - In the random coefficients world, IV measures instead an *average* of heterogeneous causal effects across a particular population – $E[\delta_i]$ for some group of i units
 - IV, therefore, measures the *local average treatment effect* or LATE parameter, which is the average of causal effects across the subpopulation of *compliers*, or those units whose covariate of interest, D_i , is influenced by the instrument.
- Angrist and Evans (1996) example: not every woman whose first two kids share gender will go on to have a third kid.
 - Under heterogeneous treatment effects, Angrist and Evans (1996) identify the causal effect of the gender composition of the first two kids on labor supply
 - Remark: That is not the same thing as identifying the causal effect of children on labor supply; the former is a LATE whereas the latter might be better described as an ATE
- *Ex post* this is probably obvious, but *ex ante* this was a real breakthrough (see Angrist, Imbens and Rubin 1996; Imbens and Angrist 1994)

Other potentially interesting treatment effects

- Another effect which we may be potentially interested in estimating is the familiar estimand, the average treatment effect on the treatment group (ATT)
- Remark: LATE is *not* the same as ATT, though.

$$\underbrace{E[Y_i^1 - Y_i^0 | D_i = 1]}_{\text{ATT}} = \underbrace{E[Y_i^1 - Y_i^0 | D_i^0 = 1]}_{\text{Effect on always takers}} P[D_i^0 = 1 | D_i = 1] + \underbrace{E[Y_i^1 - Y_i^0 | D_i^1 > D_i^0]}_{\text{Effect on compliers}} P[D_i^1 > D_i^0, Z_i = 1 | D_i = 1]$$

- The average treatment effect on the treated, ATT, is a weighted average of the effects on always-takers and compliers.
- If there are no always takers we can, however, estimate ATT which is equal to LATE in that case.

Discussions and questions

- When might we *not* be interested in the local average treatment effect?
 - Romneycare in Massachusetts examples: If the compliance rate or treatment effects differ in the community than during some quasi-experimental expansion, are we more interested in LATE, ATT or ATE?
 - What might be other examples of this in economics? In education?
 - This has a similar flavor to methodological questions regarding a study's "external" vs. "internal validity"
- What do we make of the fact that LATE is defined for an *unobservable sub-population* (i.e., can't label all units in the population as compliers or noncompliers)?
- What do we make of the fact that IV identification is based on a set of untestable assumptions?
 - For example: colonial settler mortality (Z) influences economic development (Y) *only through* Z 's association with human capital accumulation rather than institutions, D (Acemoglu, Johnson and Robinson, 2001).

IV in Randomized Trials

- The use of IV methods may be helpful when evaluating a randomized trial
- In many randomized trials, participation is nonetheless voluntary among those randomly assigned to treatment
- On the other hand, persons in the control group usually do not have access to treatment
 - only those who are particularly likely to benefit from treatment therefore will probably take up treatment which almost always leads to positive selection bias
 - if you just compare means between treated and untreated individuals using OLS, you will obtain biased treatment effects *even for the randomized trial* due non-compliance
- Solution: instrument for treatment with whether you were offered treatment and estimate LATE

Get some intuition for subgroup populations: binary encouragement designs

- Four types of people by potential outcomes
 - **Compliers** - treatment if encouraged, control if not
 - **Always-takers** - treatment whether encouraged or not
 - **Never-takers** - control whether encouraged or not
 - **Defiers** - control if encouraged, treatment if not encouraged
- Not all of these may exist in a particular study
- In a trial of a new drug or offering (removing) a new (existing) feature, there are neither always-takers nor defiers

Binary encouragement designs

	D=0	D=1
Z=0	Compliers and Never-takers	Defiers and Always-takers
Z=1	Defiers and Never-takers	Compliers and Always-takers

Heterogenous treatment effects

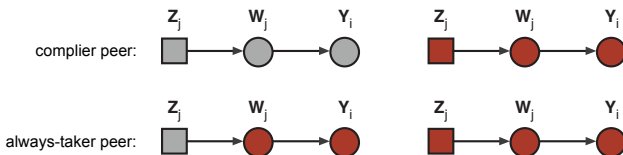
- *Monotonicity*: With probability 1, $D_i^z \geq D_i^{z'}$ for all $z \geq z'$ and all i
- Then local average treatment effect (LATE) is identified
- In binary Z , D case, LATE is the average treatment effect for the population of compliers
- We always have to ask ourselves: are we interested in the LATE?

Peer encouragement designs with a single behavior of interest

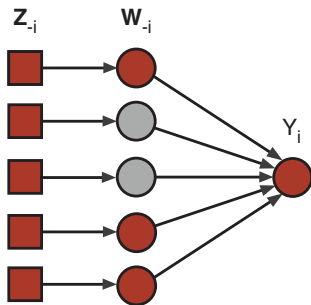
- Assign to encouragement to:
 - enroll in a retirement savings account (Duflo and Saez 2003)
 - post a thankful status update on Thanksgiving Day
- Summarize peer assignments (e.g., number of peers assigned)
- Summarize peer behaviors (e.g., number of adopter peers)
- Compute average ego behaviors as a function of these

Peer encouragement with dyads

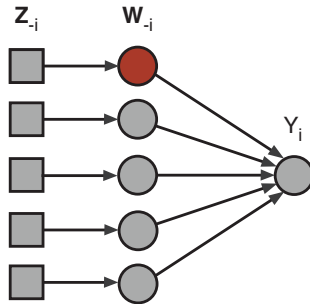
- 1 Randomly encourage j or not
- 2 Observe j 's behavior (endogenous treatment for i)
- 3 Observe i 's behavior (outcome)



Peer encouragement with multiple peers

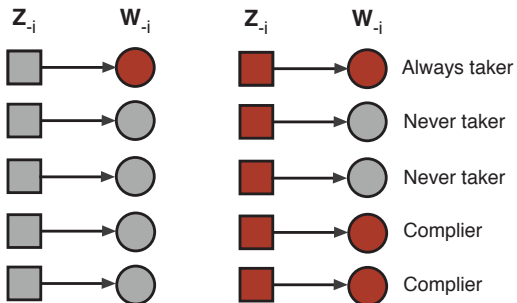


Encourage all peers



Encourage no peers

Noncompliance with multiple peers



Reduced form: Intent-to-treat (ITT)

- One option is forego the IV analysis and instead estimate the ITT
- In other words, analyze ego outcome Y as a function of the number of peers assigned Z
- What else?
 - Compute probabilities of assignment, $Pr(Z)$
 - Conduct complier analysis

Complier analysis: Intuition

- We cannot identify the correct group at the individual level
- But we can compare compliers against other populations at the group level
- This will give us a better sense for the types of individuals who are driving the LATE effect sizes

Complier analysis: size of complier population

- In case of a binary IV and treatment the size of the complier population is given by the WALD- first stage
- How many individuals got moved into treatment from having the different instrument assigned (taking the pill)?
- This will give us a better sense for the types of individuals who are driving the LATE effect sizes

Size of the complier group

$$\begin{aligned} P[D_i^1 > D_i^0] &= E[D_i^1 - D_i^0] \\ &= E[D_i^1] - E[D_i^0] \\ &= E[D^i | Z_i = 1] - E[D^i | Z_i = 0] \end{aligned}$$

Complier analysis: proportion of complier population among treated

- Note for compliers treatment status is completely determined by Z_i

Proportion of the complier group (among treated)

$$\begin{aligned} & P[D_i^1 > D_i^0 | D_i = 1] \\ &= \frac{P[D_i^1 = 1 | D_i^1 > D_i^0] P[D_i^1 > D_i^0]}{P[D_i = 1]} \\ &= E[D_i^1] - E[D_i^0] \\ &= E[D^i | Z_i = 1] - E[D^i | Z_i = 0] \end{aligned}$$

- The proportion of the treated who are compliers is given by the first stage, times the probability of the instrument switched on, divided by the proportion treated.
- Tough to understand intuitively but this can be calculated!

Examples of quantifying compliers

Source	Endogenous Variable	Instrument	Sample	Size of complier group	Proportion compliers among treated
Angrist (1990)	Veteran Status	Draft Eligibility	White men born 1950	0.534	0.101
			Non-white men born 1950	0.534	0.033
Angrist and Evans (1998)	More than 2 children	Twins as second birth	Married women with 2 or more children	0.603	0.966
Angrist and Krueger (1991)	High school graduate	3rd/4th quarter birth	Men born 1930-39	0.016	0.034

Charaterising compliers

- Despite the fact that we cannot identify individual compliers, we can describe the distribution of complier characteristics
- This can be compared to the general population/sample
- As a result we can get a sense of how the complier population is selected with respect to certain X vars.

Charaterising compliers

Example: Are sex composition compliers more or less likely to be college graduates than other comen with two children?

Charaterising compliers (relative first stages)

$$\frac{P[X_i^1=1|D_i^1>D_i^0]}{P[X_i^1=1]} = \frac{P[D_i^1>D_i^0|X_i^1=1]}{P[D_i^1>D_i^0]}$$
$$= \frac{E[D_i|Z_i=1, X_i^1=1] - E[D_i|Z_i=0, X_i^1=1]}{E[D_i|Z_i=1] - E[D_i|Z_i=1]}$$

- This is the ratio of the first stage for college graduates to the overall first stage.
- This gives the relative likelihood that a complier is a college graduate.

Examples: characterising compliers for twin instruments

Variable	$P[X_i^1 = 1]$	$P[X_i^1 = 1 D_i^1 > D_i^0]$	$\frac{P[X_i^1 = 1 D_i^1 > D_i^0]}{P[X_i^1 = 1]}$
Age 30 or older at first birth	0.0029	0.004	1.39
Black or hispanic	0.125	0.103	0.822
College graduate	0.132	0.151	1.14

- Twin compliers are more likely to be over 30 at first birth compared to sample.
- They are also less likely to be black or hispanic and more likely to have a college degree.

Beyond complier analysis: the value of having multiple instruments

- In the end complier analysis gives us a sense for the population the LATE is estimated for. We still do not know if LATE generalises.
- Imagine you have multiple valid instruments for the same variable but with different complier populations
 - If you find different effects - these are driven by effect heterogeneity
 - If you find similar effects - effect homogeneity assumptions probably holds
 - Remember the Hausman test that we did to check for exogeneity of instruments when you have multiple instruments there? Conditional on all instruments being valid, this equivalent test can now be used for testing against H_0 : homogeneity.

J. Angrist, V. Lavy and A. Schlosser (2010). “Multiple Experiments for the Causal Link between the Quantity and Quality of Children”, Journal of Labor Economics, vol. 28, issue 4.

Where is IV useful? Lottery designs

- A common design is to use a randomized lottery as an instrumental variable for some treatment
- Examples might be randomized lottery for attending charter schools to study effect of charter schools on educational outcomes
- We'll discuss a papers from 2012 evaluating a lottery-based expansion of Medicaid health insurance on Oregon on numerous health and financial outcomes

Overarching question

- What are the effects of expanding access to public health insurance for low income adults?
 - Magnitudes, and even the signs, associated with that question were uncertain
- Limited existing evidence
 - Institute of Medicine review of evidence was suggestive, but a lot of uncertainty
 - Observational studies are confounded by selection into health insurance
 - Quasi-experimental work often focuses on elderly and children
 - Only one randomized experiment in a developed country: the RAND health insurance experiment
 - 1970s experiment on a general population
 - Randomized cost-sharing, not coverage itself

The Oregon Health Insurance Experiment

- Setting: Oregon Health Plan Standard
 - Oregon's Medicaid expansion program for poor adults
 - Eligibility
 - Poor ($<100\%$ federal poverty line) adults 19-64
 - Not eligible for other programs
 - Uninsured > 6 months
 - Legal residents
 - Comprehensive coverage (no dental or vision)
 - Minimum cost-sharing
 - Similar to other states in payments, management
 - Closed to new enrollment in 2004

The Oregon Medicaid Experiment

- Lottery

- Waiver to operate lottery
- 5-week sign-up period, heavy advertising (January to February 2008)
- Low barriers to sign up, no eligibility pre-screening
- Limited information on list
- Randomly drew 30,000 out of 85,000 on list (March-October 2008)
- Those selected given chance to apply
 - Treatment at household level
 - Had to return application within 45 days
 - 60% applied; 50% of those deemed eligible → 10,000 enrollees

Oregon Health Insurance Experiment

- Evaluate effects of Medicaid using lottery as randomized controlled trial (RCT)
 - Intent-to-treat: Reduced form comparison of outcomes between treatment group (lottery selected individuals) and controls (not selected)
 - LATE: IV using lottery as instrument for insurance coverage
 - First stage: about a 25 percentage point increase in insurance coverage
 - Archived analysis plan
 - Massive data collect effort – primary and secondary
- Similar to ACA expansion but limits to generalizability
 - Partial equilibrium vs. General equilibrium
 - Mandate and external validity
 - Oregon vs. other states
 - Short vs. Long-run

Examine Broad Range of Outcomes

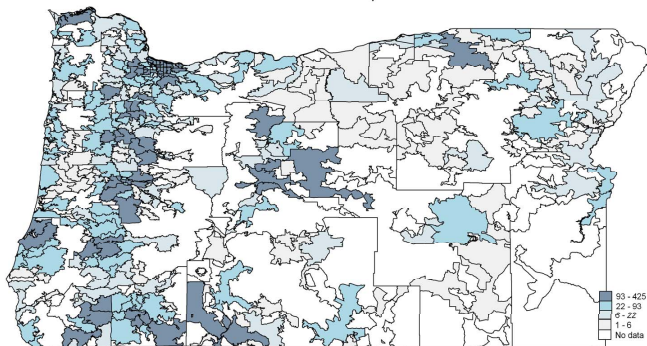
- Costs: Health care utilization
 - Insurance increases resources (income) and lowers price, increasing utilization
 - But improved efficiency (and improved health), decreasing utilization (“offset”)
 - Additional uncertainty when comparing Medicaid to no insurance
- Benefits I: Financial risk exposure
 - Insurance supposed to smooth consumption
 - But for very low income, is most care *de jure* or *de facto* free?
- Benefits II: Health
 - Expected to improve (via increased quantity / quality of care)
 - But could discourage health investments (“*ex ante* moral hazard”)

Data

- Pre-randomization demographic information
 - From lottery sign-up
- State administrative records on Medicaid enrollment
 - Primary measure of first stage (i.e., insurance coverage)
- Outcomes
 - Administrative data (~16 months post-notification): Hospital discharge data, mortality, credit reports
 - Mail surveys (~15 months): some questions ask 6-month look-back; some ask current
 - In-person survey and measurements (~25 months): Detailed questionnaires, blood samples, blood pressure, body mass index

Lottery List

Distribution Across Zip Codes



Empirical Framework

- Reduced form – estimates of the causal effect of lottery selection

$$Y_{ihj} = \beta_0 + \beta_1 LOTTERY_h + X_{ih}\beta_2 + V_{ih}\beta_3 + \varepsilon_{ihj}$$

- Validity of experimental design: randomization; balance on treatment and control
- Instrumental variables – effect of insurance coverage

$$\begin{aligned} INSURANCE_{ihj} &= \delta_0 + \delta_1 LOTTERY_{ih} + X_{ih}\delta_2 + V_{ih}\delta_3 + \mu_{ihj} \\ y_{ihj} &= \pi_0 + \pi_1 INSURANCE_{ih} + X_{ih}\pi_2 + V_{ih}\pi_3 + v_{ihj} \end{aligned}$$

- Effect of lottery on coverage: about 25 percentage points
- Additional assumption for causality: primary pathway
 - Could affect participation in other programs, but actually small
 - “Warm glow” of winning – especially early
- Analysis plan, multiple inference adjustment

Effect of lottery on coverage (first stage)

	Full sample		Credit subsample		Survey respondents	
	Control mean	Estimated FS	Control mean	Estimated FS	Control mean	Estimated FS
Ever on Medicaid	0.141	0.256 (0.004)	0.135	0.255 (0.004)	0.135	0.290 (0.007)
Ever on OHP Standard	0.027	0.264 (0.003)	0.028	0.264 (0.004)	0.026	0.302 (0.005)
# of Months on Medicaid	1.408	3.355 (0.045)	1.352	3.366 (0.055)	1.509	3.943 -0.09
On Medicaid, end of study period	0.106	0.148 (0.003)	0.101	0.151 (0.004)	0.105	0.189 (0.006)
Currently have any insurance (self report)					0.325	0.179 (0.008)
Currently have private ins. (self report)					0.128	-0.008 (0.005)
Currently on Medicaid (self report)					0.117	0.197 (0.006)
Currently on Medicaid					0.093	0.177 (0.006)

Amy Finkelstein, et al. (2012). “The Oregon Health Insurance Experiment: Evidence from the First Year”, Quarterly Journal of Economics, vol. 127, issue 3, August.

Effects of Medicaid

- Use primary and secondary data to gauge 1-year effects
 - Mail surveys: 70,000 surveys at baseline, 12 months
 - Administrative data
 - Medicaid enrollment records
 - Statewide Hospital discharge data, 2007-2010
 - Credit report data, 2007-2010
 - Mortality data, 2007-2010

Mail survey data

- **Fielding protocol**

- ~70,000 people, surveyed at baseline and 12 months later
- Basic protocol: three-stage mail survey protocol, English/Spanish
- Intensive protocol on a 30% subsample included additional tracking, mailings, phone attempts (done to adjust for non-response bias)

- **Response rate**

- Effective response rate = 50%
- Non-response bias always possible, but response rate and pre-randomization measures in administrative data were balanced between treatment and control

Administrative data

- **Medicaid records**

- Pre-randomization demographics from list
- Enrollment records to assess “first stage” (how many of the selected got insurance coverage)

- **Hospital discharge data**

- Probabilistically matched to list, de-identified at Oregon Health Plan
- Includes dates and source of admissions, diagnoses, procedures, length of stay, hospital identifier
- Includes years before and after randomization

- **Other data**

- Mortality data from Oregon death records
- Credit report data, probabilistically matched, de-identified

Sample

- 89,824 unique individuals on the waiting list
- Sample exclusions (based on pre-randomization data only)
 - Ineligible for OHP Standard (out of state address, age, etc.)
 - Individuals with institutional addresses on list
- Final sample: 79,922 individuals out of 66,385 households
 - 29,834 treated individuals (surveyed 29,589)
 - 40,088 control individuals (surveyed 28,816)

Sample characteristics

Variable	Mean	Variable	Mean
Panel A: Full sample			
% Female	0.56	Average Age	41
Panel B: Survey responders only			
<i>Demographics:</i>		<i>Health Status: Ever diagnosed with:</i>	
% White	0.82	Diabetes	0.18
% Black	0.04	Asthma	0.28
% Spanish/Hispanic/Latino	0.12	High Blood Pressure	0.40
% High school or less	0.67	Emphysema or Chronic Bronchitis	0.13
% don't currently work	0.55	Depression	0.56
<i>Determinants of eligibility:</i>			
Average hh income (2008)	13,050	% with any insurance	0.33
% below Federal poverty line	0.68	% with private insurance	0.13

Outcomes

- **Access and use of care**

- Is access to care improved? Do the insured use more care? Is there a shift in the types of care being used?
- Mail surveys and hospital discharge data

- **Financial strain**

- How much does insurance protect against financial strain?
- What are the out-of-pocket implications?
- Mail surveys and credit reports

- **Health**

- What are the short-term impacts on self-reported physical and mental health?
- Mail surveys and vital statistics (mortality)

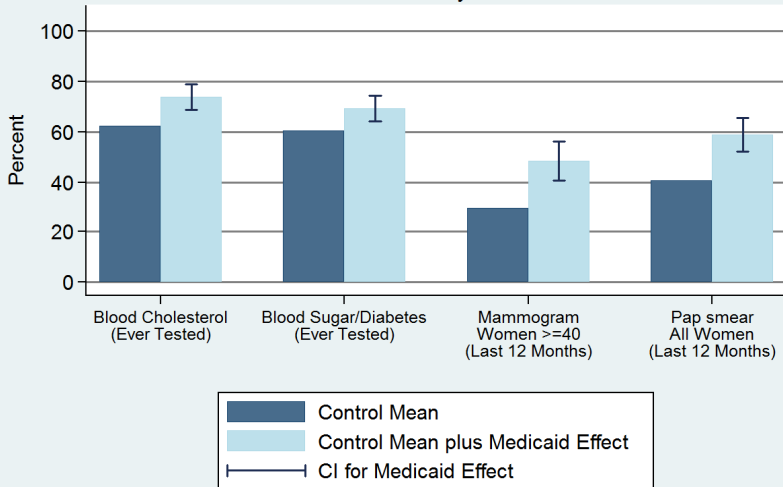
Effect of lottery on coverage (first stage)

- Gaining insurance resulted in better access to care and higher satisfaction with care (conditional on actually getting care)

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Have a usual place of care	49.9%	+9.9%	+33.9%	.0001
Have a personal doctor	49.0%	+8.1%	+28.0%	.0001
Got all needed health care	68.4%	+6.9%	+23.9%	.0001
Got all needed prescriptions	76.5%	+5.6%	+19.5%	.0001
Satisfied with quality of care	70.8%	+4.3%	+14.2%	.001

SOURCE: Survey data

Preventive Care Mail Survey Data



Effect of lottery on coverage (first stage)

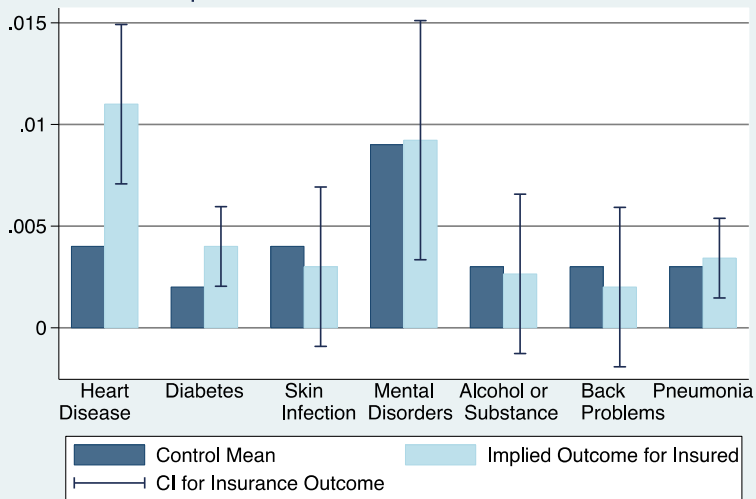
- Gaining insurance resulted in increased probability of hospital admissions, primarily driven by non-emergency department admissions

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Any hospital admission	6.7%	+.50%	+2.1%	.004
--Admits through ED	4.8%	+.2%	+.7%	.265
--Admits NOT through ED	2.9%	+.4%	+1.6%	.002

SOURCE: Hospital Discharge Data

Overall, this represents a 30% higher probability of admission, although admissions are still rare events

Hospital Utilization for Selected Conditions



Summary: Access and use of care

- Overall, utilization and costs went up relative to controls
 - 30% increase in probability of an inpatient admission
 - 35% increase in probability of an outpatient visit
 - 15% increase in probability of taking prescription medications
 - Total \$777 increase in average spending (a 25% increase)
- With this increased spending, those who gained insurance were
 - 35% more likely to get all needed care
 - 25% more likely to get all needed medications
 - Far more likely to follow preventive care guidelines, such as mammograms (60%) and PAP tests (45%)

Results: Financial Strain

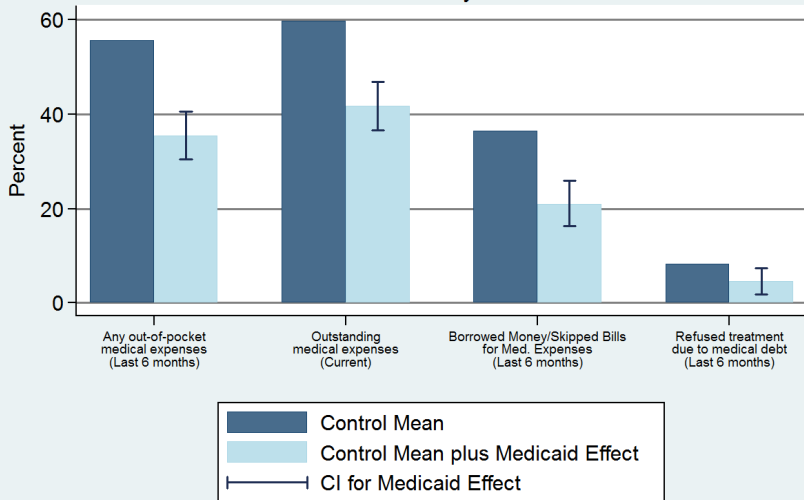
- Gaining insurance resulted in a reduced probability of having medical collections in credit reports, and in lower amounts owed

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Had a bankruptcy	1.4%	+0.2%	+0.9%	.358
Had a collection	50.0%	-1.2%	-4.8%	.013
--Medical collections	28.1%	-1.6%	-6.4%	.0001
--Non-medical collections	39.2%	-0.5	-1.8%	.455
\$ owed medical collections	\$1,999	-\$99	-\$390	.025

Source: Credit report data

Self-reported Financial Strain

Mail Survey Data



Summary: Financial Strain

- Overall, reductions in collections on credit reports were evident
 - 25% decrease in probability of a medical collection
 - Those with a collection owed significantly less
- Household financial strain related to medical costs was mitigated
 - Substantial reduction across all financial strain measures
 - Captures “informal channels” people use to make it work
- Implications for both patients and providers
 - Only 2% of bills sent to collections are ever paid

Results: Self-reported health

- Self-reported measures showed significant improvements one year after randomization

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Health good, v good, excellent	54.8%	+3.9%	+13.3%	.0001
Health stable or improving	71.4%	+3.3%	+11.3%	.0001
Depression screen NEGATIVE	67.1%	+2.3%	+7.8%	.003
CDC Healthy Days (physical)	21.86	+.381	+1.31	.018
CDC Healthy Days (mental)	18.73	+.603	+2.08	.003

Source: Survey data

Summary: Self-reported health

- Overall, big improvements in self-reported physical and mental health
 - 25% increase in probability of good, very good or excellent health
 - 10% decrease in probability of screening for depression
- Physical health measures open to several interpretations
 - Improvements consistent with findings of increased utilization, better access, and improved quality
 - BUT in their baseline surveys, results appeared shortly after coverage ($\sim 2/3$ magnitude of full result)
 - May suggest increase in *perception* of well-being rather than physical health
- Biomarker data can shed light on this issue

Discussion

- At 1 year, found increases in utilization, reductions in financial strain, and improvements in self-reported health
 - Medicaid expansion had benefits and costs – didn't "pay for itself"
 - Confirmed biases inherent in observational studies – would have estimated bigger increases in use and smaller improvements in outcomes
- Policy-makers may have different views on value of different aspects of improved well-being
 - "I have an incredible amount of fear because I don't know if the cancer has spread or not."
 - "A lot of times I wanted to rob a bank so I could pay for the medicine I was just so scared . . . People with cancer either have a good chance or no chance. In my case it's hard to recover from lung cancer but it's possible. Insurance took so long to kick in that I didn't think I would get it. Now there is a big bright light shining on me." (Anecdotes)
- Important to have broad evidence on multifaceted effects of Medicaid expansions

What to do when there are many outcomes?

- Need to pre-register analysis plan with the AEA
 - Forces you to pre-specify analysis plan
 - Many leading journals will not publish RCT results that are not pre-registered
- Problem is that with many different outcomes, some might show effects by chance.
 - If you stick to your analysis plan, you cannot opportunistically select outcomes that show effects ex-post
 - In principle, you could adjust for multiple outcomes testing using statistical methods such as the Bonferroni correction. This does not solve problem of “opportunistic” behaviour.
- Q: what about pre-analysis plans also for non-RCT studies?

Temporary page!

\LaTeX was unable to guess the total number of pages correctly. There was some unprocessed data that should have been added to the document, so this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will disappear, because \LaTeX now knows how many pages to expect for the document.