The Stable Unit Treatment Value Assumption (SUTVA) and Its Implications for Social Science RCTs

Alan S. Gerber & Donald P. Green
Yale University

From Chapter 8 of *Field Experimentation: Design, Analysis, and Interpretation*

Prepared for Presentation at the Conference on Empirical Legal Studies, Yale Law School, November 5, 2010

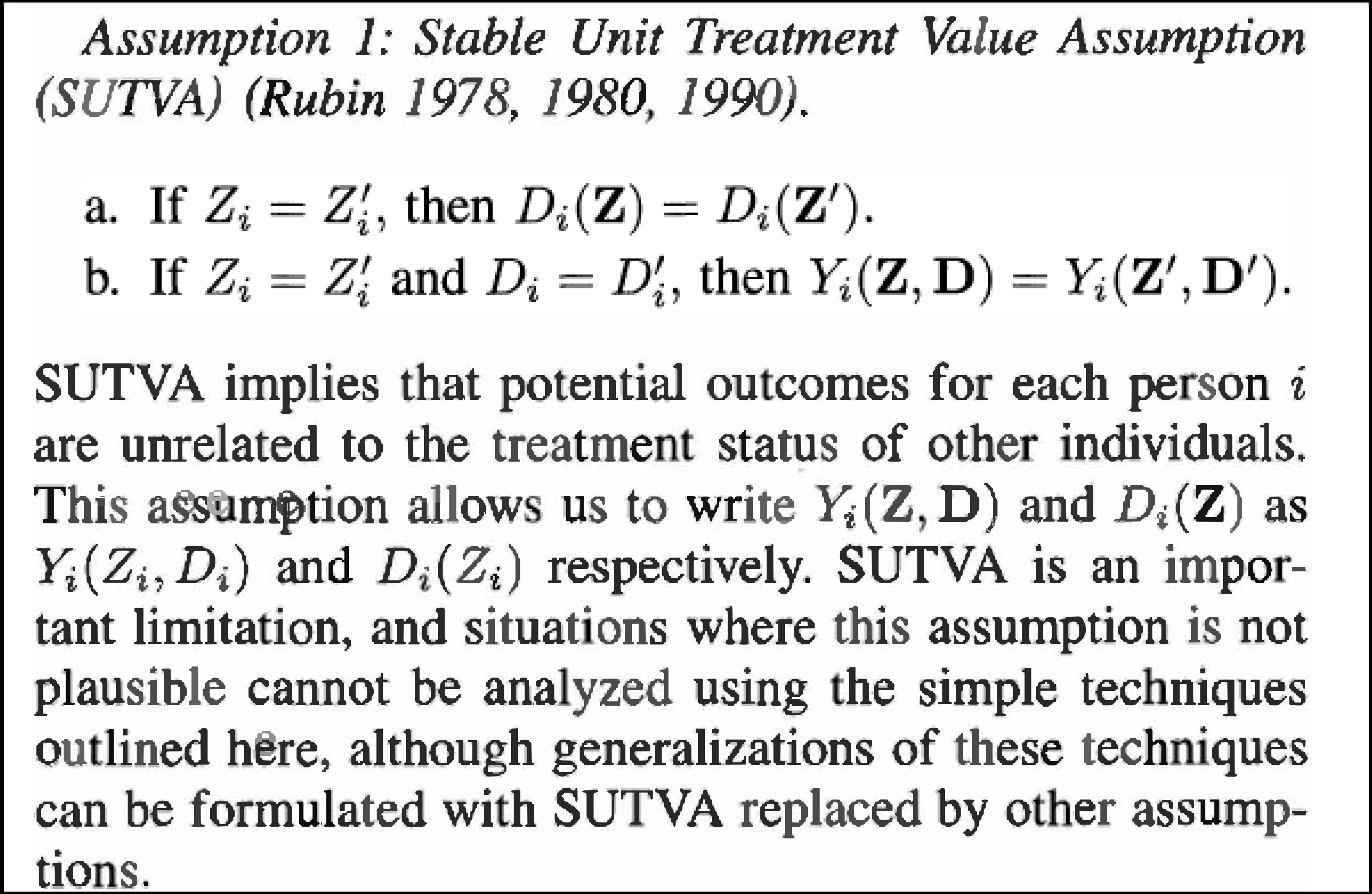
Outline

1. SUTVA defined
2. Consequences of SUTVA violations for estimation
3. Designs for identifying spillover and displacement
4. SUTVA distinguished from spatial or serial correlation

SUTVA defined

- Potential outcomes Y(1) if treated and Y(0) if not treated
- Conventional definition of a causal effect
  - For each observation, the difference in potential outcomes if the unit were treated or not treated
  - T = Y(1) – Y(0)
- SUTVA implies no unmodeled spillovers
  - Under this definition of a causal effect, potential outcomes for a given observation respond only to its own treatment status; potential outcomes are invariant to random assignment of others

SUTVA: As defined by Angrist, Imbens, and Rubin 1996



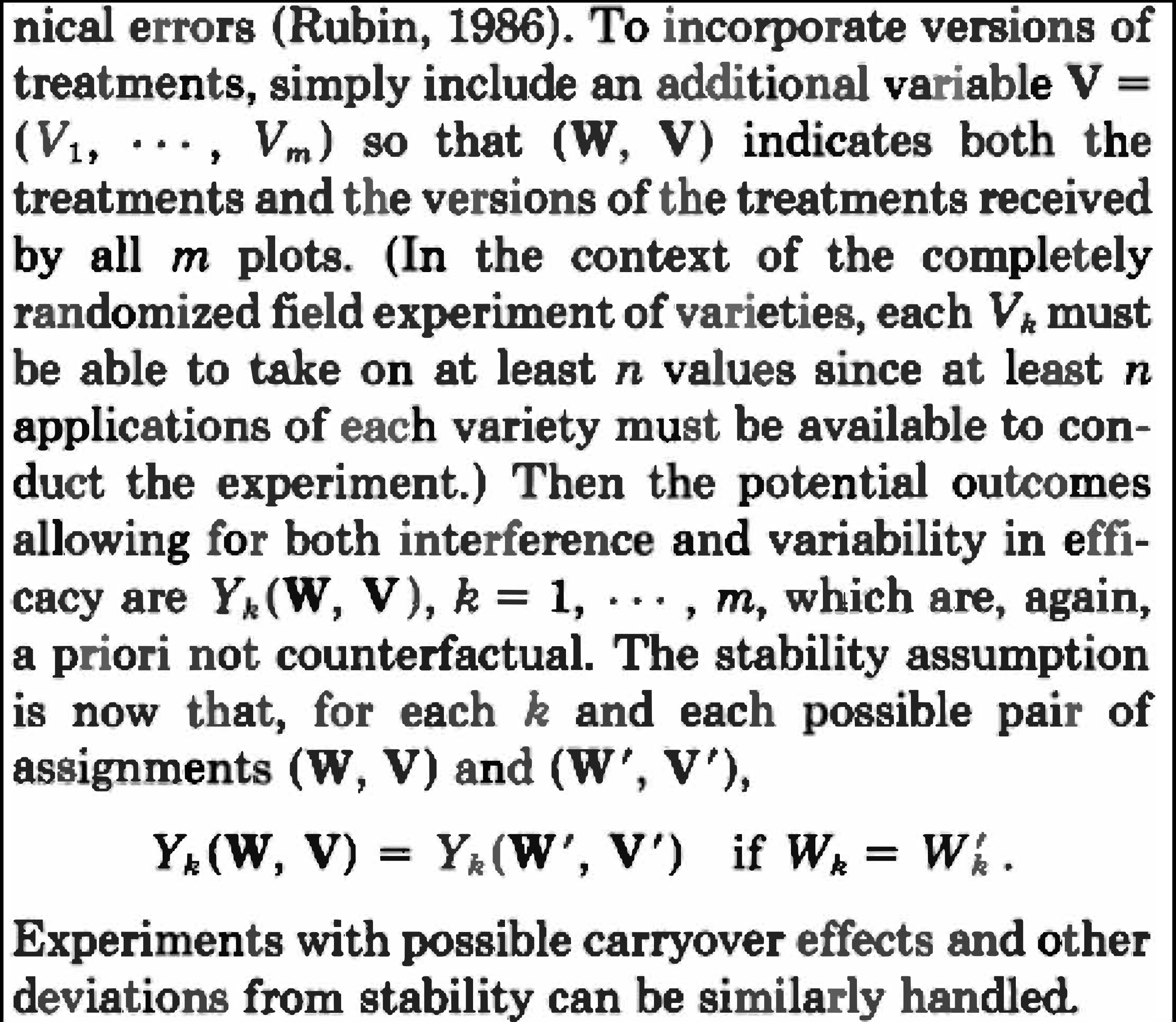
Assumption 1: Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1978, 1980, 1990).

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

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

SUTVA implies that potential outcomes for each person i are unrelated to the treatment status of other individuals. This assumption allows us to write $Y_i(\mathbf{Z}, \mathbf{D})$ and $D_i(\mathbf{Z})$ as $Y_i(Z_i, D_i)$ and $D_i(Z_i)$ respectively. SUTVA is an important limitation, and situations where this assumption is not plausible cannot be analyzed using the simple techniques outlined here, although generalizations of these techniques can be formulated with SUTVA replaced by other assumptions.

SUTVA: As defined by Rubin 1990



nical errors (Rubin, 1986). To incorporate versions of treatments, simply include an additional variable $\mathbf{V} = (V_1, \cdots, V_m)$ so that $(\mathbf{W}, \mathbf{V})$ indicates both the treatments and the versions of the treatments received by all m plots. (In the context of the completely randomized field experiment of varieties, each V_k must be able to take on at least n values since at least n applications of each variety must be available to conduct the experiment.) Then the potential outcomes allowing for both interference and variability in efficacy are $Y_k(\mathbf{W}, \mathbf{V})$, $k = 1, \cdots, m$, which are, again, a priori not counterfactual. The stability assumption is now that, for each k and each possible pair of assignments $(\mathbf{W}, \mathbf{V})$ and $(\mathbf{W}', \mathbf{V}')$,

$$Y_k(\mathbf{W}, \mathbf{V}) = Y_k(\mathbf{W}', \mathbf{V}') \quad \text{if } W_k = W'_k .$$

Experiments with possible carryover effects and other deviations from stability can be similarly handled.

What if potential outcomes are affected by the treatment status of others?

- Could write out potential outcomes in a more extensive fashion, taking into account both one’s own treatment status and the treatment status of other types of units
- E.g., housemates, friends, relatives, neighbors, competitors…
- Hypotheses about spillovers or displacement follow from theories about communication, social comparisons, competition, etc.

Hypotheses about spillovers

- Contagion: The effect of being vaccinated on one's probability of contracting a disease depends on whether others have been vaccinated.

- Displacement: Police interventions designed to suppress crime in one location may displace criminal activity to nearby locations.

- Communication: Interventions that convey information about commercial products, entertainment, or political causes may spread from individuals who receive the treatment to others who are nominally untreated.

- Social comparison: An intervention that offers housing assistance to a treatment group may change the way in which those in the control group evaluate their own housing conditions.

- Signaling: Policy interventions are sometimes designed to "send a message" to other units about what the government intends to do or what it has the capacity to do.

- Persistence and memory: Within-subjects experiments, in which outcomes for a given unit are tracked over time, may involve "carryover" or "anticipation."

Expanding the schedule of potential outcomes to satisfy SUTVA

- For example, potential vote outcomes {0,1} may reflect whether you and/or your housemate are encouraged to vote
  - Y(00): no one is treated in the household
  - Y(10): you're untreated, housemate is treated
  - Y(01): you're treated, housemate is not
  - Y(11): you and your housemate are treated

SUTVA now requires no cross-household spillover

Example of potential and observed outcomes

Observation	Y00 No one is treated	Y01 You are treated	Y10 Housemate treated	Y11 Both are treated	T Actual treatment	Y Observed Outcome
Pam	0	0	0	0	You	0
Mary	0	1	1	1	Housemate	1
Peter	0	0	0	1	Both	1
Akhil	0	1	0	1	Neither	0
Ella	0	0	1	1	You	0
Holger	1	1	1	1	Both	1
Barbara	1	0	0	0	Housemate	0

Causal estimands under household spillovers

- Y(01) – Y(00): effect of direct treatment on you, given that your housemate is untreated
- Y(10) – Y(00): spillover effect on you when your housemate is untreated
- Y(11) – Y(10): effect of direct treatment on you, given that your housemate is treated
- Y(11) – Y(01): spillover effect on you, given that you are treated directly

Notice that attentiveness to SUTVA forces us to be clearer about what we seek to estimate

SUTVA violations open Pandora’s Box

- The range of possible spillovers becomes astronomical once we allow spillovers between pairs of units, triples of units, quadruples, etc.
- Clearly, a problem for observational as well as experimental research but also a sobering reminder that experimentation is not an assumption-free endeavor

Intuitively, we sense that SUTVA may be implausible in many applications

- SUTVA implies the following designs will, in expectation, gauge the same estimand:

(1) vaccinations randomly assigned such that 5% of a sample receives them and 95% do not

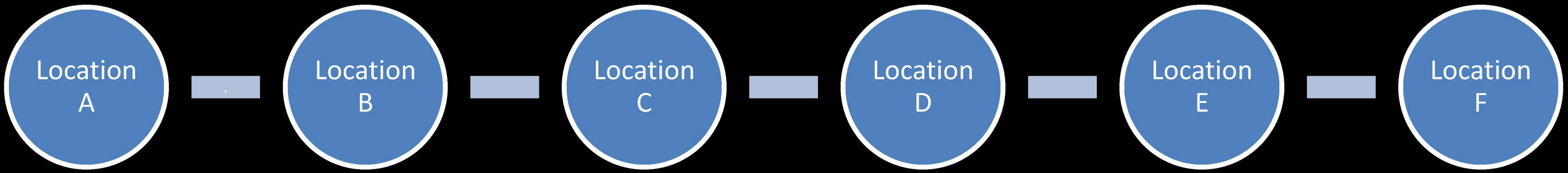
(2) vaccinations randomly assigned such that 95% of a sample receives them and 5% do not

Six Social Science Applications

- Crime displacement: “hot spots” policing
- General deterrence: Brazilian corruption audits
- Recalibration of evaluations: MTO experiments
- Intra- and inter-household spillovers: voter mobilization
- Time-series or within-subjects design
- Lab experiments with dyadic or group interaction

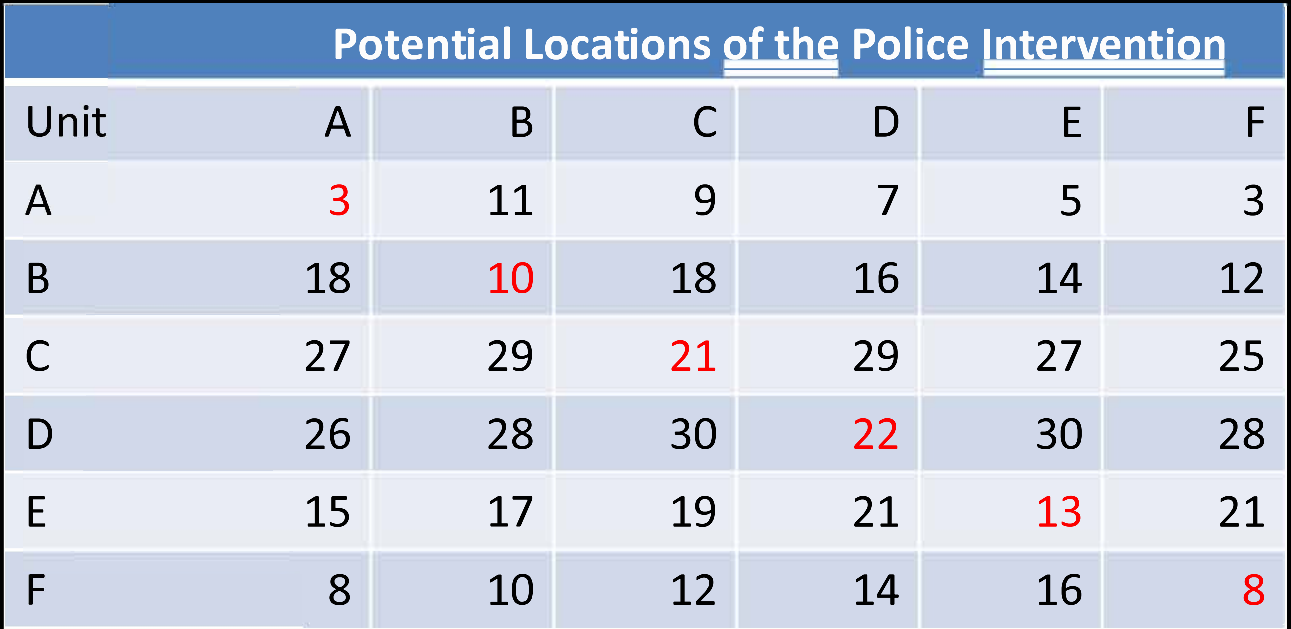
Crime displacement and the perils of naïve data analysis

- Consider a very simple case of policing on one street that stretches for 6 blocks



- The police treat one randomly chosen block while maintaining control tactics elsewhere
- The schedule of potential crime outcomes for each of the units includes the what-if response to all 6 possible assignments

Potential outcomes: crime rates



	Potential Locations of the Police Intervention					
Unit	A	B	C	D	E	F
A	3	11	9	7	5	3
B	18	10	18	16	14	12
C	27	29	21	29	27	25
D	26	28	30	22	30	28
E	15	17	19	21	13	21
F	8	10	12	14	16	8

The true data generation process for this example assumes that direct treatment lowers crime rates by 10 and that crime diminishes by 2 for every unit of distance from the treatment location

Naïve Comparison of Treatment and Control

- Pick one hot spot: Six possible randomizations, each resulting in a comparison between one treated unit and the other five control units
- The six difference-in-means are

  {-15.8, -9.0, 3.4, 4.6, -5.4, -9.8}, which average -5.3
- Due to omitted variable bias (distance is omitted and correlated with treatment), this naïve comparison fails to recover the true effect of direct treatment: -10

Naïve regression

- What happens if one regresses crime rates on treatment and distance from the treated block?

$$Y_i = \alpha + \beta_1 (\text{Treatment}_i) + \beta_2(\text{Distance}_i) + u_i$$

One obtains biased estimates, because distance to the treated block is not fully random despite the fact that the treatment is assigned at random.

Blocks in the middle stretch of the street have a shorter expected distance to potentially treated blocks.

Average estimate of β_1= -17.6, of β_2= -5.5

Spatial experiments are implicit blocked experiments

- Define strata according to which observations share the same array of proximities to all potentially treated units
- The "Pair" variables represent dummy variables for observations {1,6}, {2,5}, {3,4}. Omit one dummy.

$$Y_i = \alpha + \beta_1 (\text{Treatment}_i) + \beta_2(\text{Distance}_i) + \gamma_1(\text{Pair } 1_i) + \gamma_2(\text{Pair } 2_i) + u_i$$

Across all possible randomizations, this regression on average recovers β_1 and β_2.

Average estimate of β_1= -10, of β_2= -2

Spatial Spillovers: Summary

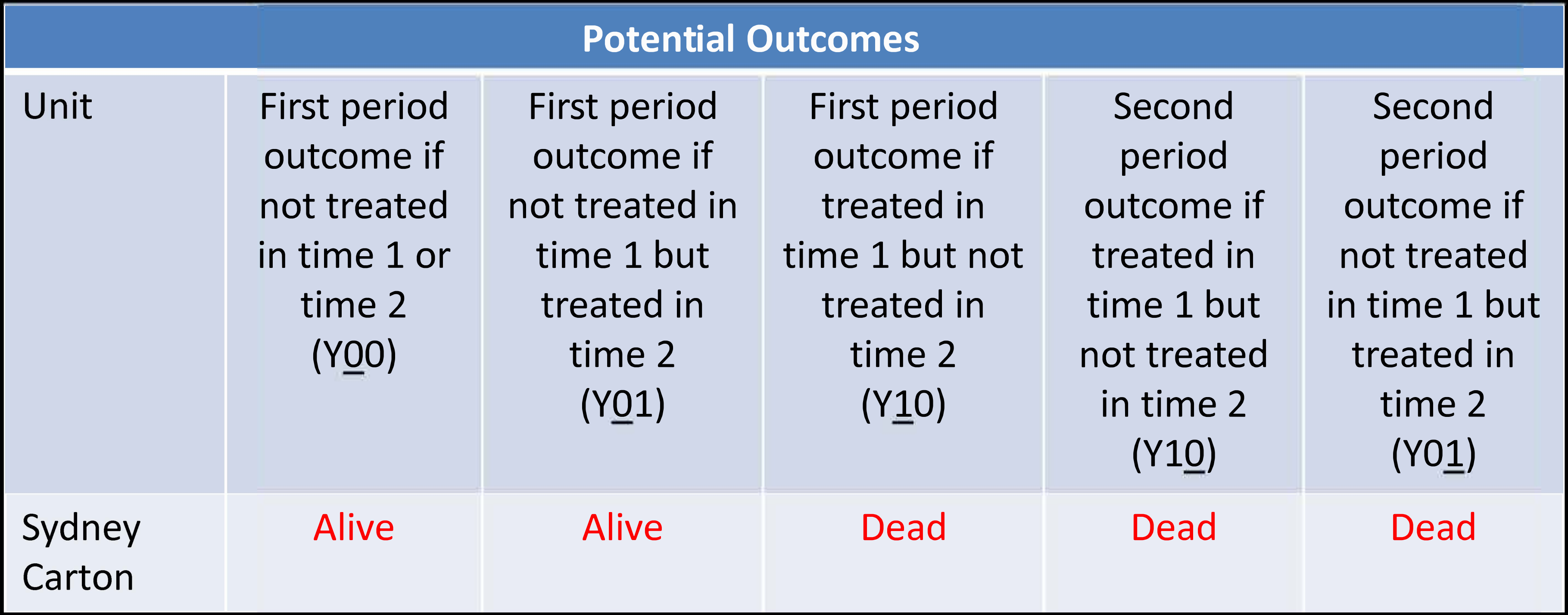
- Delicate matter to estimate treatment effects and spillover/displacement, need to attend to variations in propensity scores (in effect, these are implicitly block-randomized designs where some units may not even have an experimental counterpart)

- Parameterizing the manner in which effects change with distance/dosage invokes substantive assumptions

Potential outcomes: within-subjects design

- Notation becomes complicated because we need to indicate at each period *j* all of the potential outcomes associated with treatments in other periods
- Imagine a two period experiment with binary potential outcomes:
    - An observation is randomly assigned to treatment or control during the first or second period.
    - In the first period, we observe one of the two potential outcomes {Y01, Y10}; in the second period, we observe either {Y01, Y10}.
    - We can also imagine potential outcomes Y00 or Y00, which occur when a subject is untreated in both periods.

Example of potential outcomes for two periods when the treatment is the guillotine



	Potential Outcomes				
Unit	First period outcome if not treated in time 1 or time 2 ($Y\underline{0}0$)	First period outcome if not treated in time 1 but treated in time 2 ($Y\underline{0}1$)	First period outcome if treated in time 1 but not treated in time 2 ($Y\underline{1}0$)	Second period outcome if treated in time 1 but not treated in time 2 ($Y1\underline{0}$)	Second period outcome if not treated in time 1 but treated in time 2 ($Y0\underline{1}$)
Sydney Carton	Alive	Alive	Dead	Dead	Dead

Within-subjects design: What if a treatment is randomly assigned to either period 1 or period 2?

- Random assignment by coin flip generates two pairs of observed outcomes {Y01,Y01} and {Y10,Y10} with equal probability.
- Estimand: In the first period, the causal effect of the treatment is defined as Y10 - Y00
- The outcome Y10 refers to an untreated state that follows a treatment.
- If the treatment’s effects persist, Y10 may be quite different from Y00.

For example, suppose the treatment were the guillotine and the outcome were whether the accused is alive or dead

- The causal effect (Y10 - Y00) in period 1 is clear: $\{Y\underline{1}0=\text{Dead}, Y\underline{0}0=\text{Alive}\}$.
- The over time comparison, however, is distorted by spillover when the treatment is assigned to period 1.
- The person who was executed in period 1 would be dead in period 2 as well: $\{Y\underline{1}0=\text{Dead}, Y1\underline{0}=\text{Dead}\}$. A comparison of $Y\underline{1}0$ -$Y1\underline{0}$ would suggest that the guillotine had no causal effect…!
- SUTVA requires that potential outcomes in one period are unaffected by treatments in another period

What if the treatment were administered in period 2?

- The causal effect (Y01 – Y00) in period 2 is {Y01=Dead,Y00=Alive}.
- We observe {Y01=Dead,Y01=?}
- What assumptions get us from Y01= Y00?
  - Y01 = Y00: no foresight (e.g., no dying of fright)
  - Y00 = Y00: no trends over time (e.g., no onset of lethal violence or disease)

Within-subjects design is akin to observational research

- Depends on supplementary assumptions that are not related to randomization
  - Randomization of the timing of the intervention reduces (but does not eliminate) risk of foresight and coincidence between treatment and other trends
- Experimental procedures: wash-out periods and efforts to eliminate outside disturbances
- In sum, within-subjects design is jeopardized by SUTVA violations (as well as trends over time)

Designs to detect spillovers

- Random assignment of density of treatments
  - Special complications arise when an experiment involves noncompliance

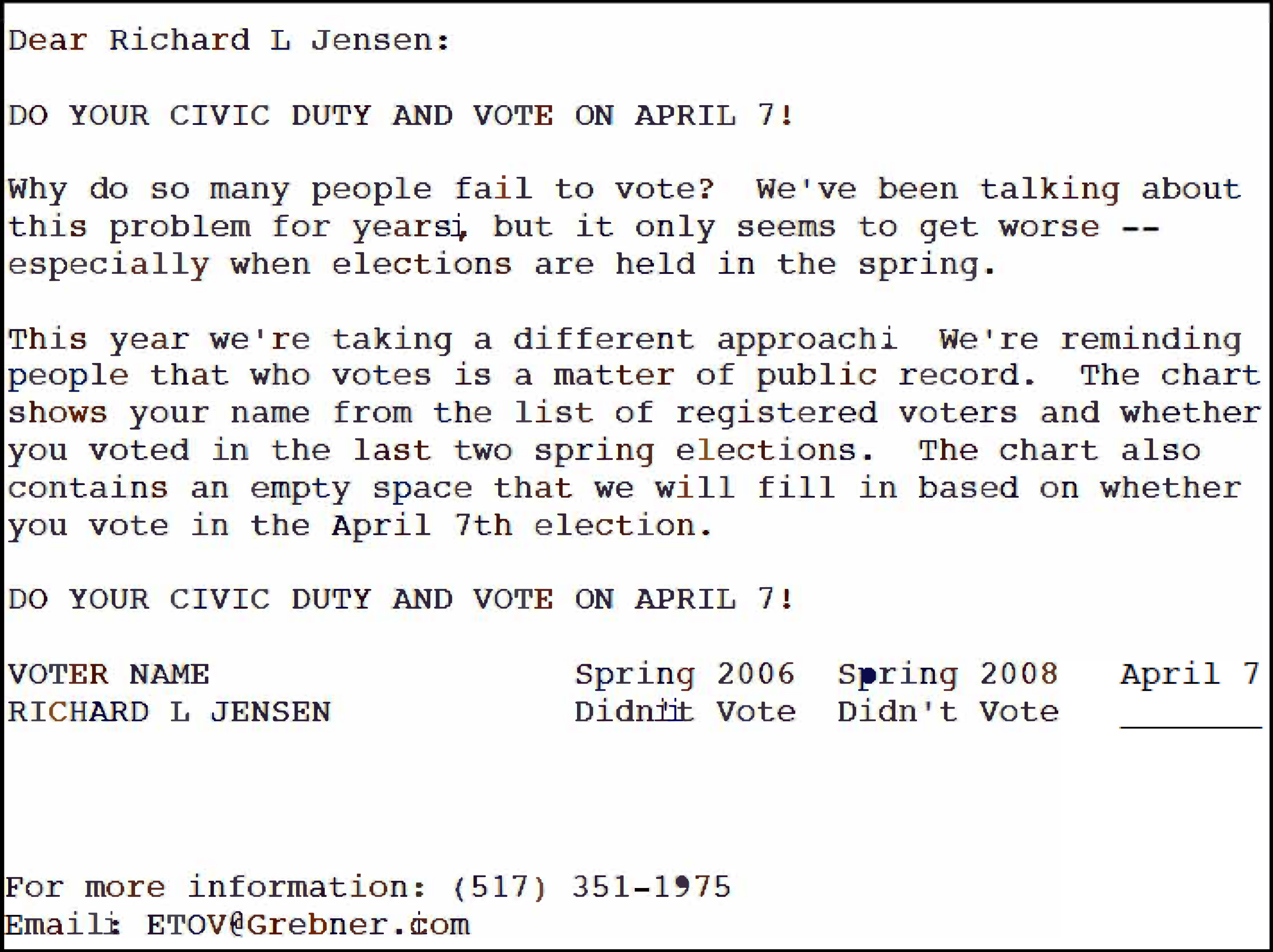
- Random-density design does not allow for all types of spillover but does address the most likely culprits

- Example: looking for within- and across-household spillovers in voter mobilization

Voter mobilization study using direct mail

- Social pressure mail in low salience election
- Design randomly varied density of treatments in 9 digit zip codes, and randomly targeted at most one member of each household
    - Zip code density: {none, one, half, all}
    - Household: {housemate in control, housemate treated}
    - Individual: {control, treatment}
- V{abc} = expected voting rate given (a) your zip code’s level of treatment, (b) whether your housemate was treated, and (c) whether you were treated

Social pressure treatment from Sinclair, McConnell, and Green (2010)



Dear Richard L Jensen:

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

Why do so many people fail to vote? We've been talking about this problem for yearsi, but it only seems to get worse -- especially when elections are held in the spring.

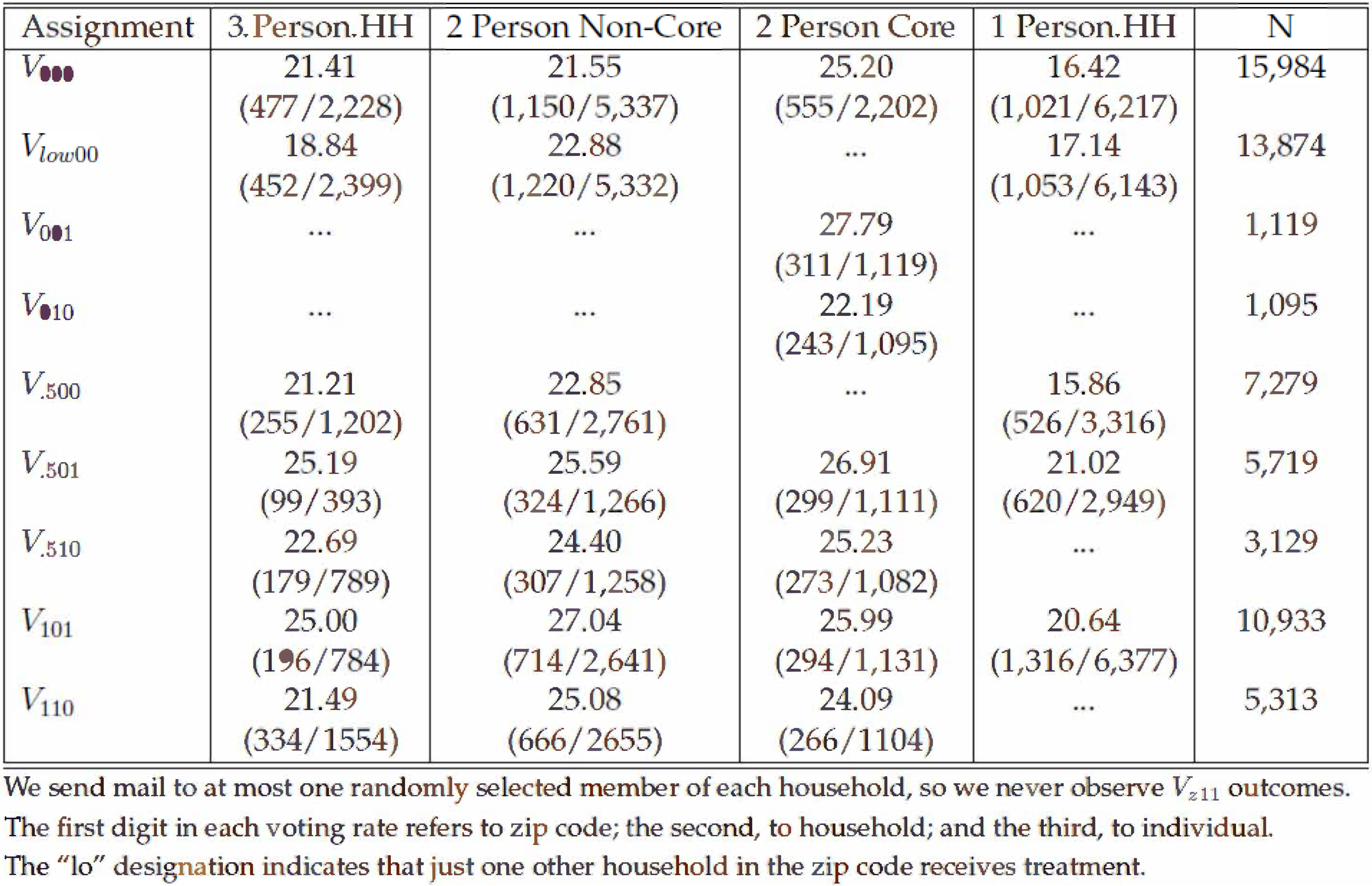
This year we're taking a different approachi We're reminding people that who votes is a matter of public record. The chart shows your name from the list of registered voters and whether you voted in the last two spring elections. The chart also contains an empty space that we will fill in based on whether you vote in the April 7th election.

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

VOTER NAME	Spring 2006	Spring 2008	April 7
RICHARD L JENSEN	Didn't Vote	Didn't Vote	________

For more information: (517) 351-1975
Email: ETOV@Grebner.com

Results from Sinclair, McConnell, and Green (2010)



Assignment	3.Person.HH	2 Person Non-Core	2 Person Core	1 Person.HH	N
V_{000}	21.41 (477/2,228)	21.55 (1,150/5,337)	25.20 (555/2,202)	16.42 (1,021/6,217)	15,984
V_{low00}	18.84 (452/2,399)	22.88 (1,220/5,332)	...	17.14 (1,053/6,143)	13,874
V_{001}	...	...	27.79 (311/1,119)	...	1,119
V_{010}	...	...	22.19 (243/1,095)	...	1,095
$V_{.500}$	21.21 (255/1,202)	22.85 (631/2,761)	...	15.86 (526/3,316)	7,279
$V_{.501}$	25.19 (99/393)	25.59 (324/1,266)	26.91 (299/1,111)	21.02 (620/2,949)	5,719
$V_{.510}$	22.69 (179/789)	24.40 (307/1,258)	25.23 (273/1,082)	...	3,129
V_{101}	25.00 (196/784)	27.04 (714/2,641)	25.99 (294/1,131)	20.64 (1,316/6,377)	10,933
V_{110}	21.49 (334/1554)	25.08 (666/2655)	24.09 (266/1104)	...	5,313

We send mail to at most one randomly selected member of each household, so we never observe V_{z11} outcomes.
The first digit in each voting rate refers to zip code; the second, to household; and the third, to individual.
The "lo" designation indicates that just one other household in the zip code receives treatment.

Results from Sinclair, McConnell, and Green (2010)

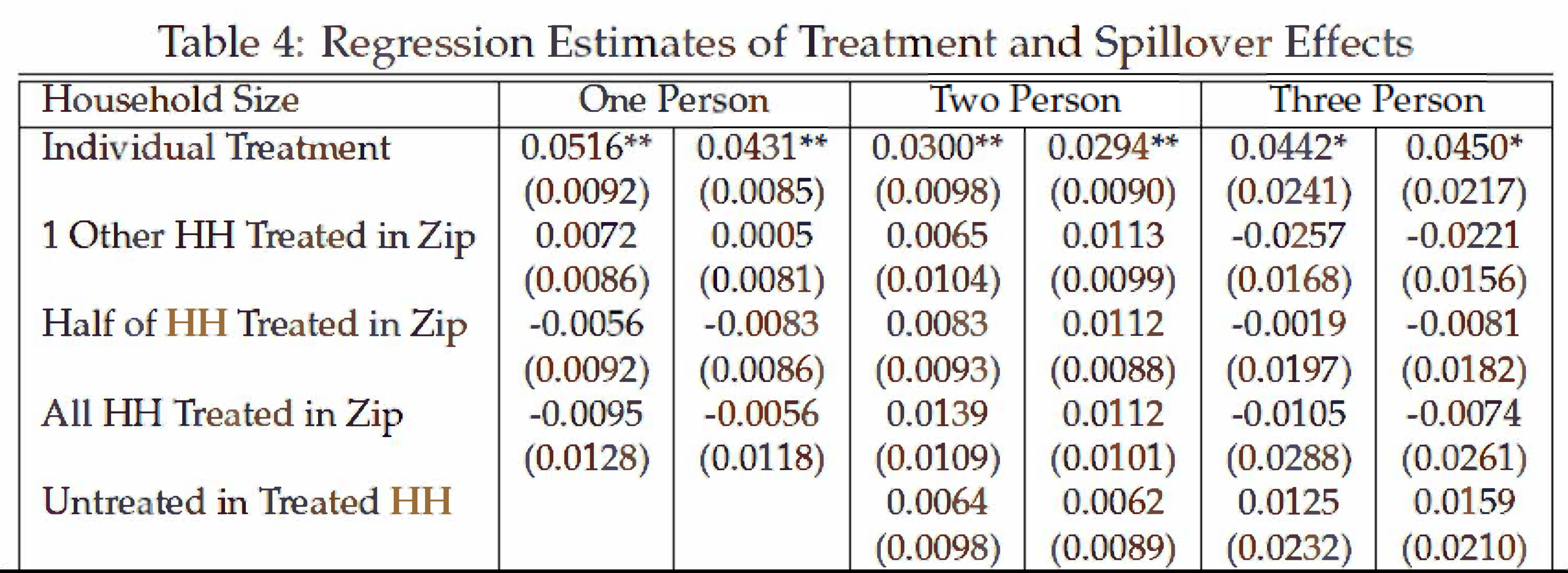


Table 4: Regression Estimates of Treatment and Spillover Effects

Household Size	One Person		Two Person		Three Person	
Individual Treatment	0.0516**	0.0431**	0.0300**	0.0294**	0.0442*	0.0450*
	(0.0092)	(0.0085)	(0.0098)	(0.0090)	(0.0241)	(0.0217)
1 Other HH Treated in Zip	0.0072	0.0005	0.0065	0.0113	-0.0257	-0.0221
	(0.0086)	(0.0081)	(0.0104)	(0.0099)	(0.0168)	(0.0156)
Half of HH Treated in Zip	-0.0056	-0.0083	0.0083	0.0112	-0.0019	-0.0081
	(0.0092)	(0.0086)	(0.0093)	(0.0088)	(0.0197)	(0.0182)
All HH Treated in Zip	-0.0095	-0.0056	0.0139	0.0112	-0.0105	-0.0074
	(0.0128)	(0.0118)	(0.0109)	(0.0101)	(0.0288)	(0.0261)
Untreated in Treated HH			0.0064	0.0062	0.0125	0.0159
			(0.0098)	(0.0089)	(0.0232)	(0.0210)

SUTVA Should Not be Confused with Spatial or Serial Correlation

- Distinction between spillover/displacement and correlated disturbances
- In the direct mail example, zip code level voting rates are highly correlated with your voting rate, but a random zip code level intervention apparently has no effect on you
- Similarly, housemates' voting patterns are highly correlated, but weak spillover effects

Summary

- SUTVA is too often ignored
- Forces us to give more thought to how we define a causal estimand
- Spillover and displacement can lead to bias
- Do not confuse spillover with spatial or serial correlation
- Research has gradually shifted from treating interference between units as a nuisance to treating spillover as a research opportunity
- Good news: can make use of non-experimental units to detect spillovers
- Bad news: proper detection of spillovers requires careful attention to modeling details