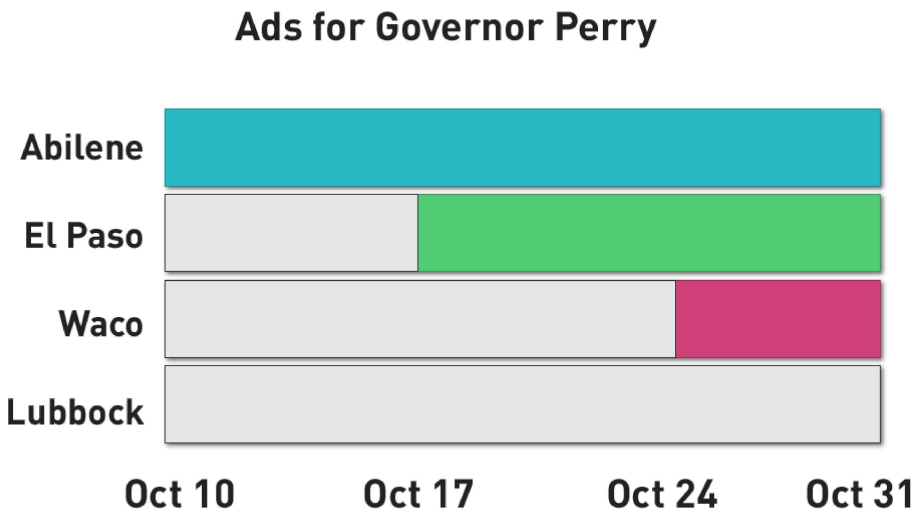
Want a title slide

Waitlist Designs

- Administer randomized treatment to all subjects over time.
- Eventually administer treatment to all subjects by the end of the experiment.
- Roll out treatment to control group in stages.
- Also known as *stepped-wedge* designs.



Benefits of Waitlist Design

- Allows experimentation while allaying concerns about totally withholding treatment
- Delays widespread treatment long enough to see some effects in early treatment groups
- Helps demonstrate long-run treatment effects as well as simultaneous

What to Remember From This Week

- All experiments aim to impose *ceteris paribus*: "all other things being equal" except the treatment effect.
- Spillovers between subjects can bias results:
    - Positive spillovers cause underestimation of treatment effects (e.g., I see ads and take my friend to the store).
    - Negative spillovers cause overestimation of true treatment effects (e.g., extra police divert crime to a different area).
- Clustered treatment assignment can eliminate this bias (e.g., treat all kids in a school or none).
- Within-subjects designs can be valuable if treatment effects are limited in time.
- Stepped-wedge designs get variation both within and between subjects.

Reading Assignment

- Read the first three paragraphs of *Field Experiments*, Section 8.5 on waitlist designs.
- The example in the rest of this section is instructive, but the details are beyond our scope.

Within-Subjects Design inverse slide before this

Within-Subjects Design: Why?

- Variance in *y*reduces precision of estimated treatment effects.
  - Social sciences involve much heterogeneity between subjects.
  - Using a person as a control for himself reduces variance.
    - Ultimate paired-subjects design
- Some real-life questions concern just one person.

Single-Subject Examples

- Will I feel better if I eat fewer carbs?
- Will I want to keep exercising if I can just make myself do it for 30 days?
- How do I know whether avoiding gluten will help me in particular?
- These questions require self-experimentation (within-subject, over time).
- Caution: Statistical inference difficult without multiple repetitions of treatment and control.

Within-Subjects Design: Why Not?

- Benefits of randomization are lost:
  - Independence from other factors that might cause our outcome
    - Before-after experiment on advertising: How do we know the "after" week didn't happen to take place at a time when the subject was more likely to shop anyway?
    - Impossible to predict in advance what other events might influence the outcome.
  - Ample permutation space for statistical inference
    - Best attainable *p*-value is 0.5.
- Including many time periods can solve this problem, giving us room for randomization and enough data for valid standard-error estimates.

Quiz

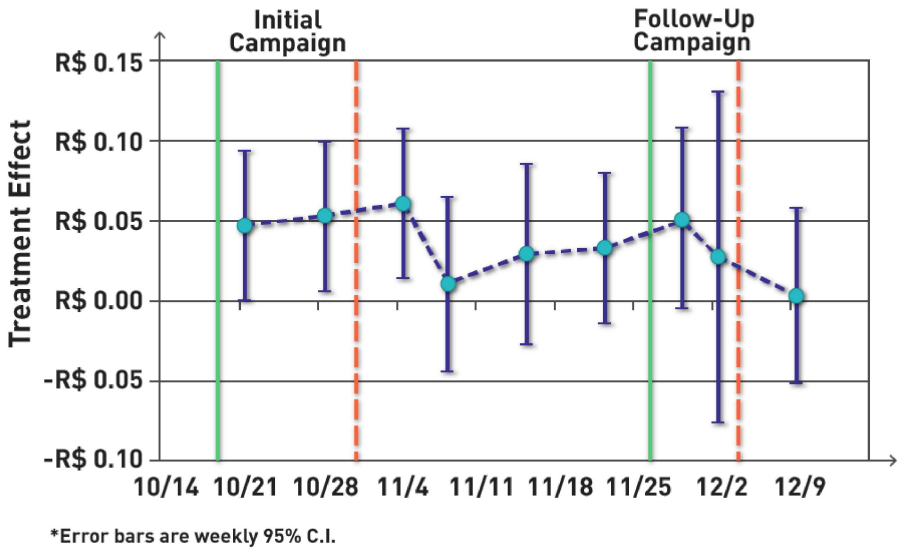
Nothing to show for this

- According to *Field Experiments*, 8.4, what assumption has to hold for within-subjects experiments to work?

Longer-Term Effects of Advertising

- Lewis and Reiley (2014) estimate persistent effects of advertising in subsequent weeks.
- Without rerandomization in second campaign, we can't distinguish its separate effects.

Figure 5. Weekly DID Estimates of the Treatment Effect



Findings

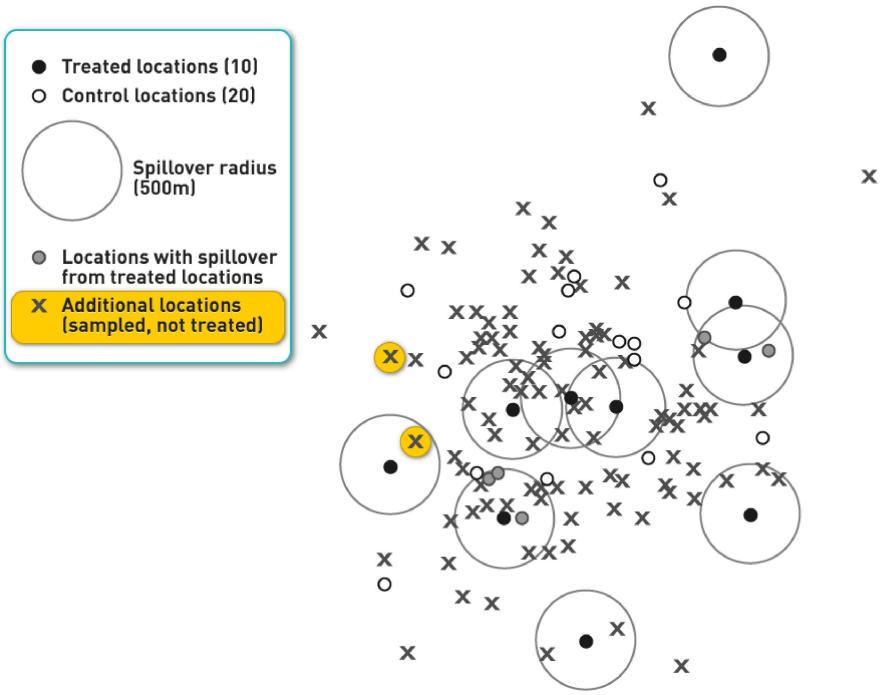
- Treatment point estimates uniformly positive, every single week.
- Difference-in-difference assumption may have caused overestimation.
    - All observations based on same preperiod.
    - Overestimate due to preperiod differences will carry over to all subsequent weeks.
- Results show evidence of long-term effects of advertising.

Reading Assignment

- Read *Field Experiments*, Section 8.4.
- Temporal spillover effects in experiments over time.
    - Spillover from present to future.
- Within-subjects experimental designs.

Spatial Spillover Example

- Skip the reading in *Field Experiments*, Section 8.3.
- See Figure 8.4 for example of spillovers in two dimensions.



just screen capture this

Making Assumptions to Deal With Spillovers

- Arbitrary assumptions about geographic extent of spillovers can radically change estimates.
  - What if range of spillover is 750 rather than 500 meters?
  - What if spillover effects are not constant out to some threshold distance?
    - Linear decline in treatment effect
    - Quadratic decline in treatment effect
- Contrast with usual virtue of experiments: Make measurements with minimal assumptions.
- Assumption problems not unique to geographic spillovers.
  - Example: spillovers via social networks (Facebook, Twitter)

Reading Assignment

- Coming up: Read *Field Experiments*, 8.4.
- First: Read a motivating example on longer-run effects of advertising.
  - Lewis and Reiley (2014), Section 4
  - Link to paper

Assumptions in 8.1

- Translated from page 257, first full paragraph, last sentence:
  - No spillovers will come from outside household.
  - Majority of spillovers will come from roommates.
  - Outside (nonroommate) spillovers complicate modeling too much to handle; assume they are small enough to ignore.

Example in 8.1

- Political mailings and word of mouth:
  - Consider the direction of bias we get if we ignore spillover and study individual randomization.
  - We underestimate the true effect of the political mailing on voter turnout.
  - Word of mouth (treatment-group member happens to talk about the political content with someone in the control group) constitutes indirect treatment and reduces treatment effect.

Reading Assignment

- Read *Field Experiments*, 8.2.
- Overall goal of this unit is to understand:
  - How spillover can bias main treatment effect
  - How to use clustered designs to correct bias
- Focus primarily on pages 260–261.
- Think of examples of the different kinds of proximity that might cause spillovers in different applications.
- Understand last sentence of last paragraph on page 261: We are ignoring spillovers from a distance of two or more units.
- Compute the probabilities in Table 8.3 by understanding the relationship to Figure 8.1.
- Don't worry about pages 262–263 (location-adjacent version of probability-weighting problem in Section 4.5).

Quiz Answer

- **Q:** With spillovers, would we expect a simple randomized experiment to underestimate or overestimate the effect of changing the auditing policy?
- **A:**Underestimate
  - Word-of-mouth effect makes 0% and 5% groups more similar and reduces estimated treatment effects.
  - If the question is how much job performance would improve under increased auditing frequency companywide, we will underestimate.

no need to include this quiz

What Can We Do?

- Randomize at work-site level (different cities) instead of individual level.
  - Uses geographic distance between groups of employees to minimize chances of cross-group spillover
  - Represents clustered design (see *Field Experiments* 3.6.2)

Clustered Design

- Used by Nagin et al. (2002) to avoid spillover problems:
  - 16 sites.
  - 12 in control group (5% audit rate).
  - Four treatment sites: 0%, 2%, 5%, and 10% rates reported (data collected at 25%).
  - Cross-site variation in treatment provides basis for results.
  - Within-site variation obtained by reshuffling treatments after six weeks.
    - Addresses concerns about everyone in a given treatment group having correlated behavior for reasons other than the treatment

Advantages and Disadvantages

- Clustered design advantages:
  - Reduced word-of-mouth spillovers
  - Easier administration
- Clustered design disadvantages:
  - Less randomization
  - Less precision in estimated treatment effect
    - Due to possible correlation of outcomes within treatment.
    - Perhaps rain is correlated with certain employee or donor behaviors.
    - Clustered standard errors correctly estimate this uncertainty; regular standard errors are underestimated.

Reading Assignment

- Read *Field Experiments*, Section 8.1.
- Make sure you understand the roommate example, especially:
    - How ignoring spillover (externality) affects bias
    - New subscript notation on potential outcomes
        - First subscript is roommate.
        - Second subscript is treated individual (self).

The Scenario

- Measuring effects of increased monitoring on quality of job performance (Nagin et al. 2002):
  - Your telemarketing firm is employed by charities to solicit donations.
  - Employees earn a bonus for each successful pledge they obtain.
  - Some reported pledges never materialize.
  - You want to know if employees are padding their results to qualify for bonuses.

The Approach

- You randomly spot-check a fraction of each employee's reported donors.
- "Bad calls" are debited from the employee's pay and noted in the paycheck itemization.
- Some of these may be employee bad luck (e.g., donor changed mind), but some may be deliberate cheating.
- You hope to minimize expensive spot-checks (callbacks).

The Experiment

- How does employee performance vary with the audit probability?
  - Pre-experiment callback rate: 10%
  - Experimental feedback to employees: 0%, 2%, 5%, and 10%
  - Actual experimental rate: 25%
- Results:
  - When zero audited calls are reported to an employee, the employee's fraction of bad calls increases by 3%.
    - Baseline mean: 2%
    - Minimum rate to affect performance: 2%
    - No statistically significant difference between 2%, 5%, and 10%

The Spillover Effect

- If employees compare notes:
  - People treated at 0% may realize that peers are treated at 5%.
  - 0% group may expect increased monitoring in the future.
    - *Professor's correction: I might reasonably assume that my audit probability is closer to 5%, not above 5%.*
- Without spillovers: Treatment effect should show what would happen if the company switched its policy from 0% to 5% on every employee.
- With word-of-mouth effects:
  - 5% treatment should create spillover effect.
  - 0% group should expect more auditing and make fewer bad calls.

What Are Spillovers?

- Effects of one person's treatment on the outcome of another person, regardless of whether the second person was treated.
- E.g., person A in ad campaign treatment group discusses the ad with person B in the control group.
- When the interaction with A causes B to make a purchase she would not have made otherwise, this is **spillover**.

Assignment

- Read introduction to *Field Experiments* Chapter 8.
- Note for later discussion any jargon you don't understand.
- Make sure you understand the bulleted list of examples on pages 255 and 256.