w241: Experimental Design

Week 09: Spillovers in Experiments

David Reiley, David Broockman, D. Alex Hughes
UC Berkeley, School of Information
Updated: 2021-06-24

9.1 Introduction

9.2 Spillovers

What are spillovers?

Effects of one person's treatment on the outcome of another person

- Regardless of whether the second person was treated
- Eg. Person A in ad campain treatment discusses the ad with person B in the control group
- When this interaction causes B to make a purchase she would not have made otherwise, this is **spillover**

Reading Assignment

Read *Field Experiments*, Chapter 8

- Note for discussion any jargon you don't understand
- Make sure you understand the bulleted list of examples on pages 255 and 266

9.3 Example: Job-Performance Monitoring

Example: Nagin et al (2002)

Research Question: Are employees padding their results to qualify for bonuses?

- Measuring effects of increased monitoring on quality of job performance
  - Telemarketing firm solicits donations for charity
  - Employee earn bonus for each successful pledge
  - Some reported pledges never materialize

Approach

- Spot-check a fraction of each employee's reported donors
- "Bad calls" are removed from the employee's bonus
- Some of these may be employee bad luck (eg. donor changed mind), but some may be deliberate cheating
- Minimize expensive spot-checks (callbacks)

Example: Nagin et al (2002) cont'd

Experiment

- How does the audit probability effect employee performance?
  - Pre-experiment callback rate: 10%
  - Experimental rates reported back to employees: 0%, 2%, 5%, and 10%
  - Actual experimental rate: 25%
  - An employee in the 10% group would have 25/100 calls audited, but only 10 of those used to report back to employee

Results

- When zero audited calls are reported back to an employee, the fraction of bad calls inreases by 3%
  - Baseline mean: 2%
  - Minimum rate to affect performance: 2%
  - No statistically significant difference between 2%, 5%, and 10%

The Spillover Effect

Employees may compare notes

- People in 0% treatment group may realized peers audited at 5%
- May expect increased monitoring in the future -professor's correction: audit probability is about to 5%, not above 5%

Without spillover effect:

- Treatment effect should show what would happen if company switched policy from 0% to 5% on every employee

With spillover effect:

- 5% treatment should create spillover effect
- 0% group should expect more auditing and make fewer bad calls

9.5 Example Continued: Spillovers and Experimental Design

Quiz Answer

Question: With spillovers, would we expect a simple randomized experiment to underestimate or overestimate the effect of changing the auditing policy?

Answer: Underestimate

- Word-of-mounth effect makes 0% and 5% groups more similar and reduces estimated treatment effects

What can we do to improve design?

Randomize at work-site level (eg. different cities) instead of individual level

- Distance between groups of employees minimizes chance of spillover
- Represents clustered design (see *Field Experiments* 3.6.2)

Clustered Design

Used in telemarketing example (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 gropu having correlated behavior for reasons other than the treatment

Advantages and Disadvantages

Clustered Design Advantages:

- Reduced word-of-mounth spillovers
- Easier administration

Clustered Design Disadvantages:

- Less randomization
- Less precision in estimated treatment effect
  - Due to possible correlation of outcomes within treatment
  - Eg. 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, specifically:

- How ignoring spillover (externality) affects bias
- New subscript notation on potential outcomes
  - First subcript is roommate
  - Second subscript is treated individual (self)

9.6 Identifying Causal Effects in the Presense of Spillovers

Example in 8.1

Assumptions

From pg 257, first paragraph, last sentence

- No spillovers will come from outside household
- Majority of spillovers will come from roommates
- Outside (non-roommate) 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 direction of bias if we ignore spillover and study individual randomization
- Underestimate true effect on voter turnout
- Word-of-mouth constitutes indirect treatment
  - Treatment group member happens to talk about political content with someone in the control group
  - Reduces treatment effect

Reading Assignment

Fields Experiments Section 8.2

Goal is to understand:

- How spillover can bias main treatment effect
- How to use clustered designs to correct bias

Focus primarily on *pgs 260-261*

- Kinds of proximity that might cause spillovers
- Last sentence of page 261: **We are ignoring spillovers from a distance of two or more units**
- Compute 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)

9.7 Quiz 2

9.8 Spatial Spillovers

Spatial Spillover Example

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

Figure 8.4: Spillover in Two Dimensions

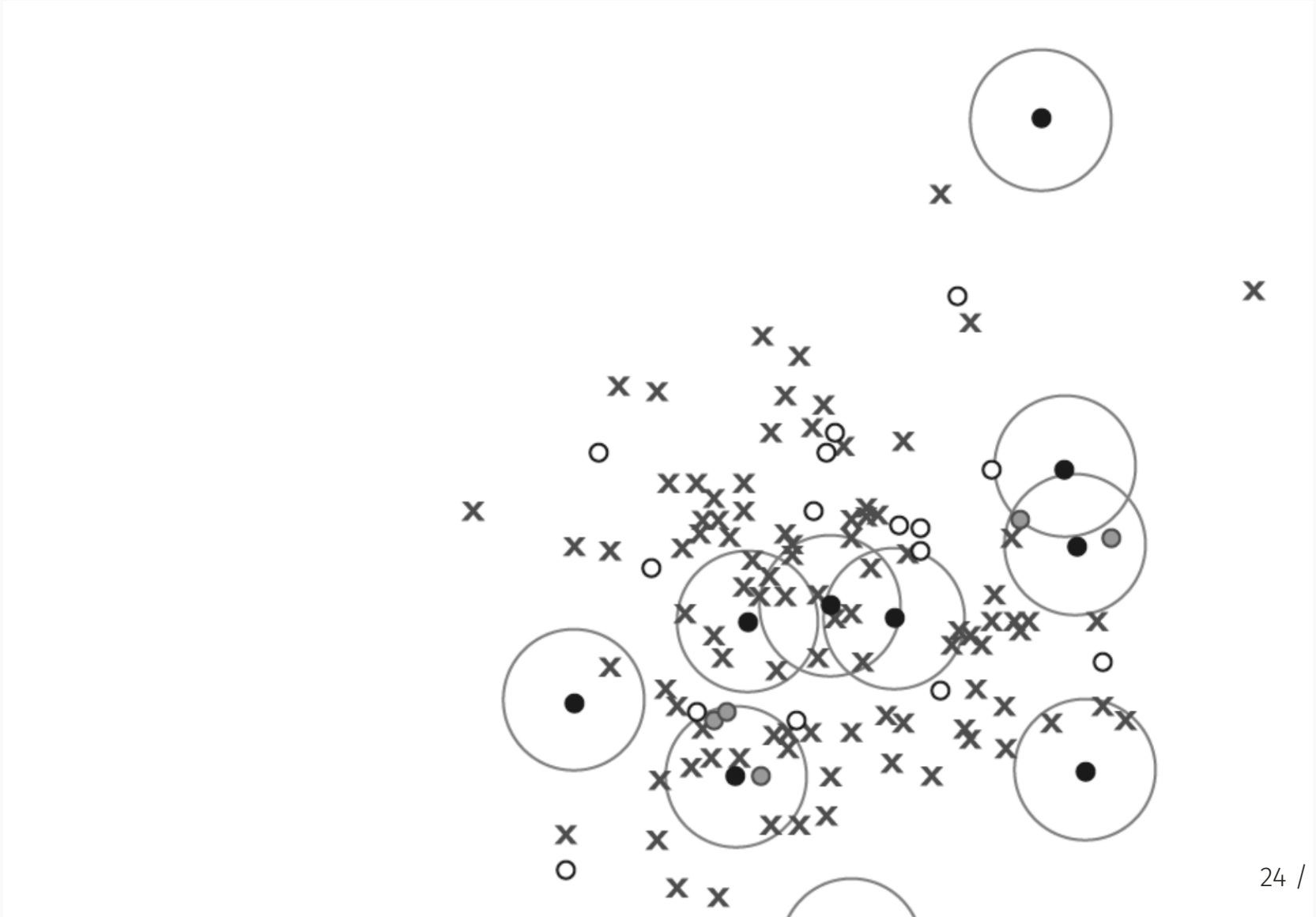


Figure 8.4: Spillover in Two Dimensions

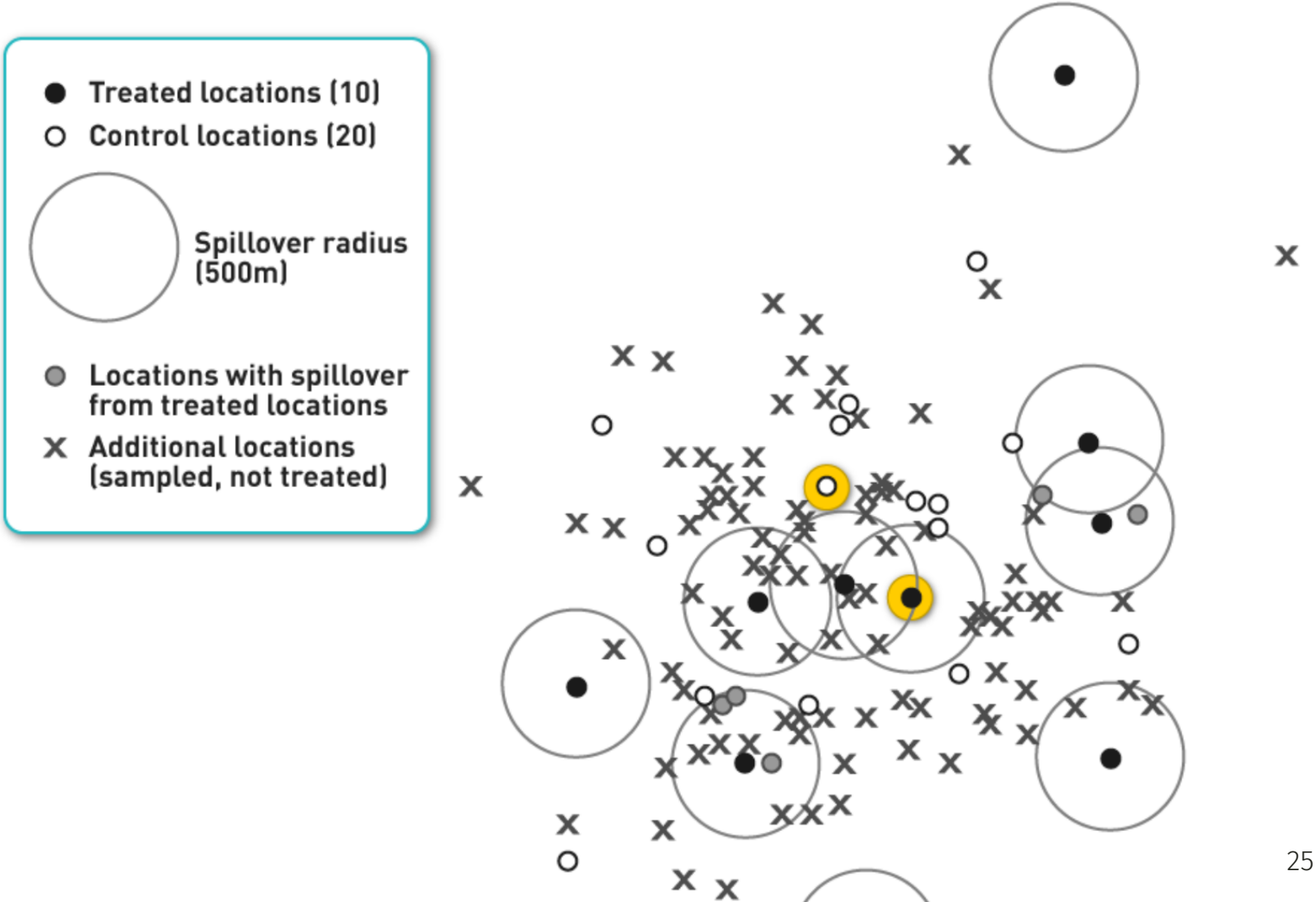


Figure 8.4: Spillover in Two Dimensions

Figure 8.4: Spillover in Two Dimensions

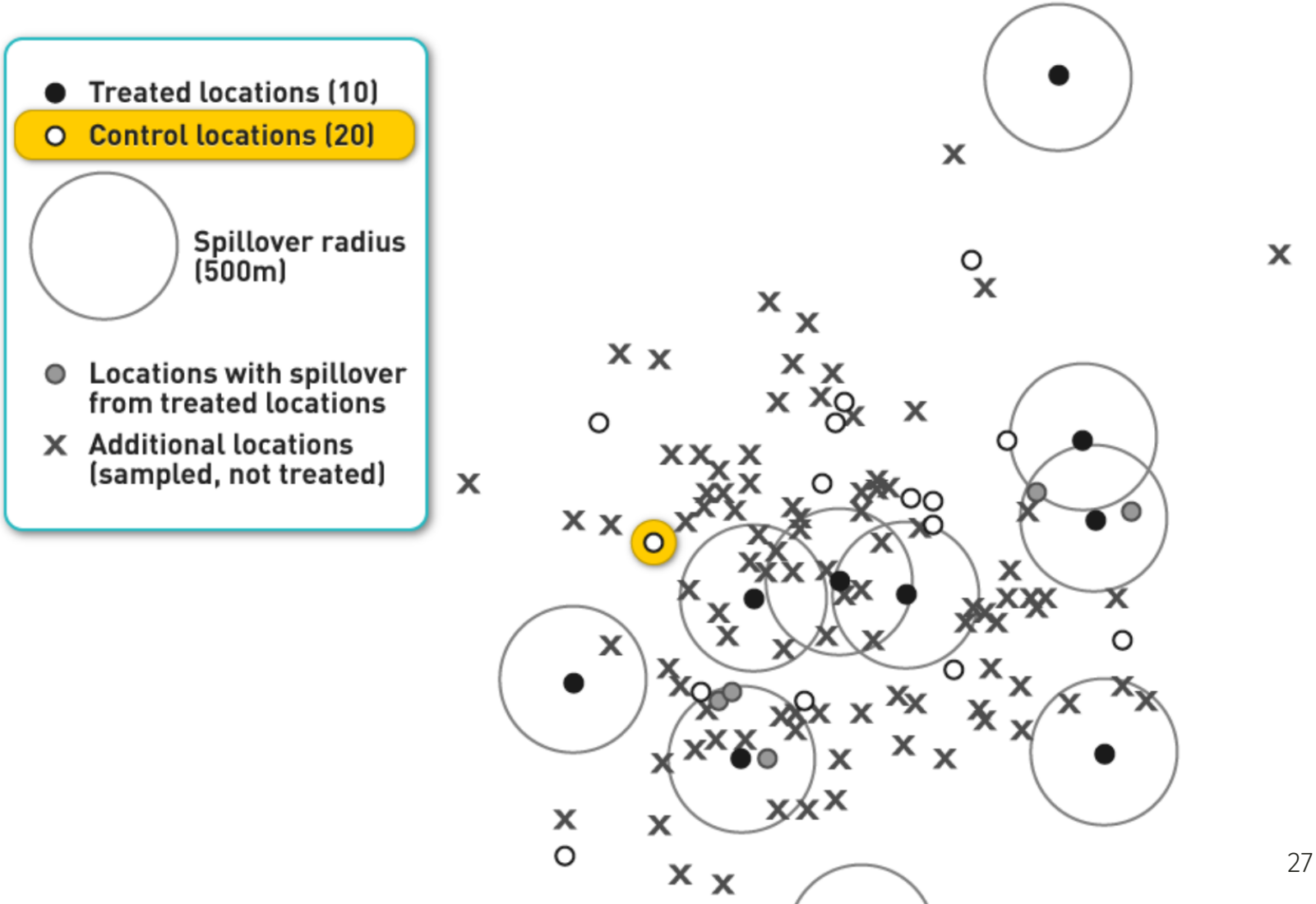


Figure 8.4: Spillover in Two Dimensions

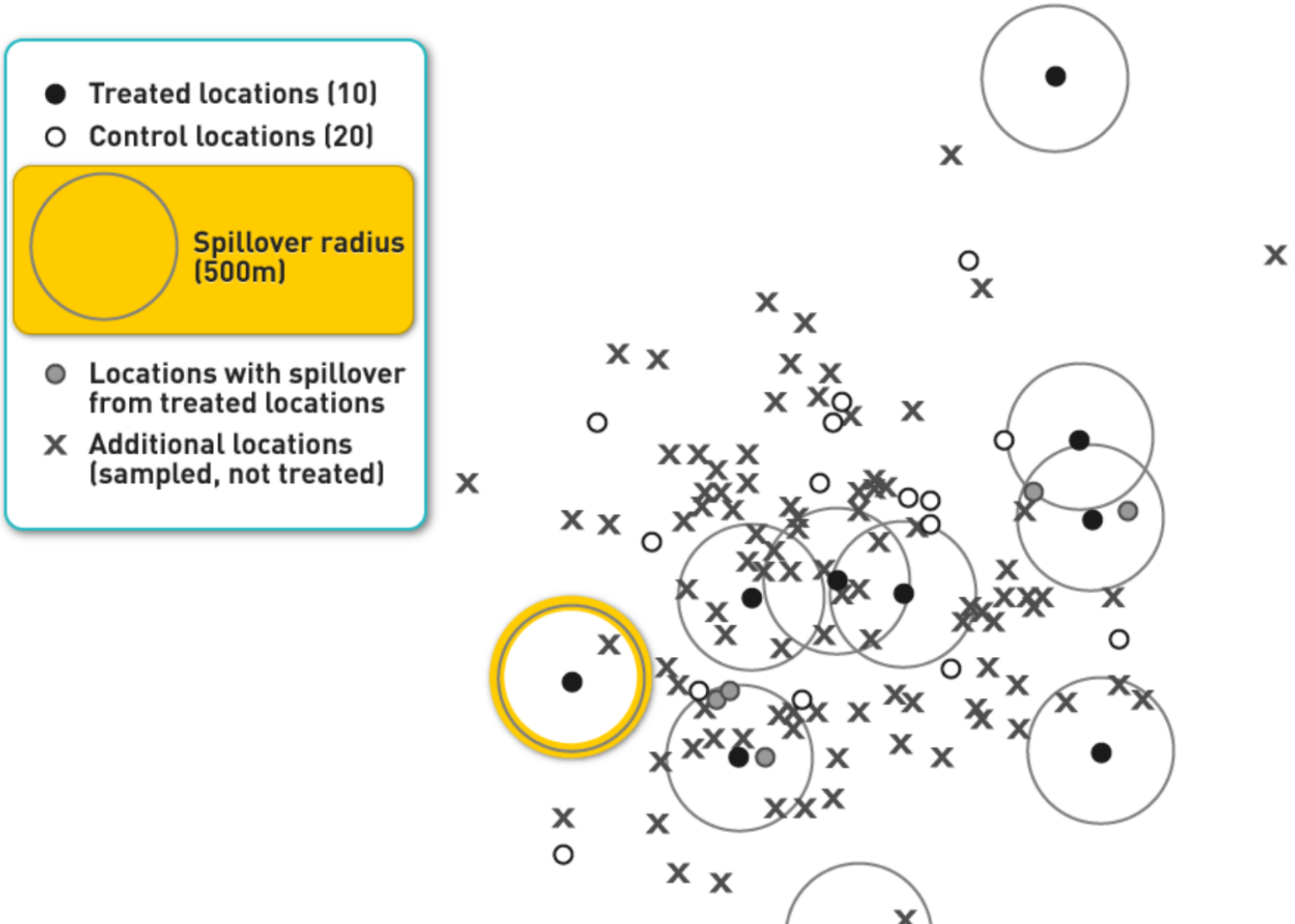


Figure 8.4: Spillover in Two Dimensions

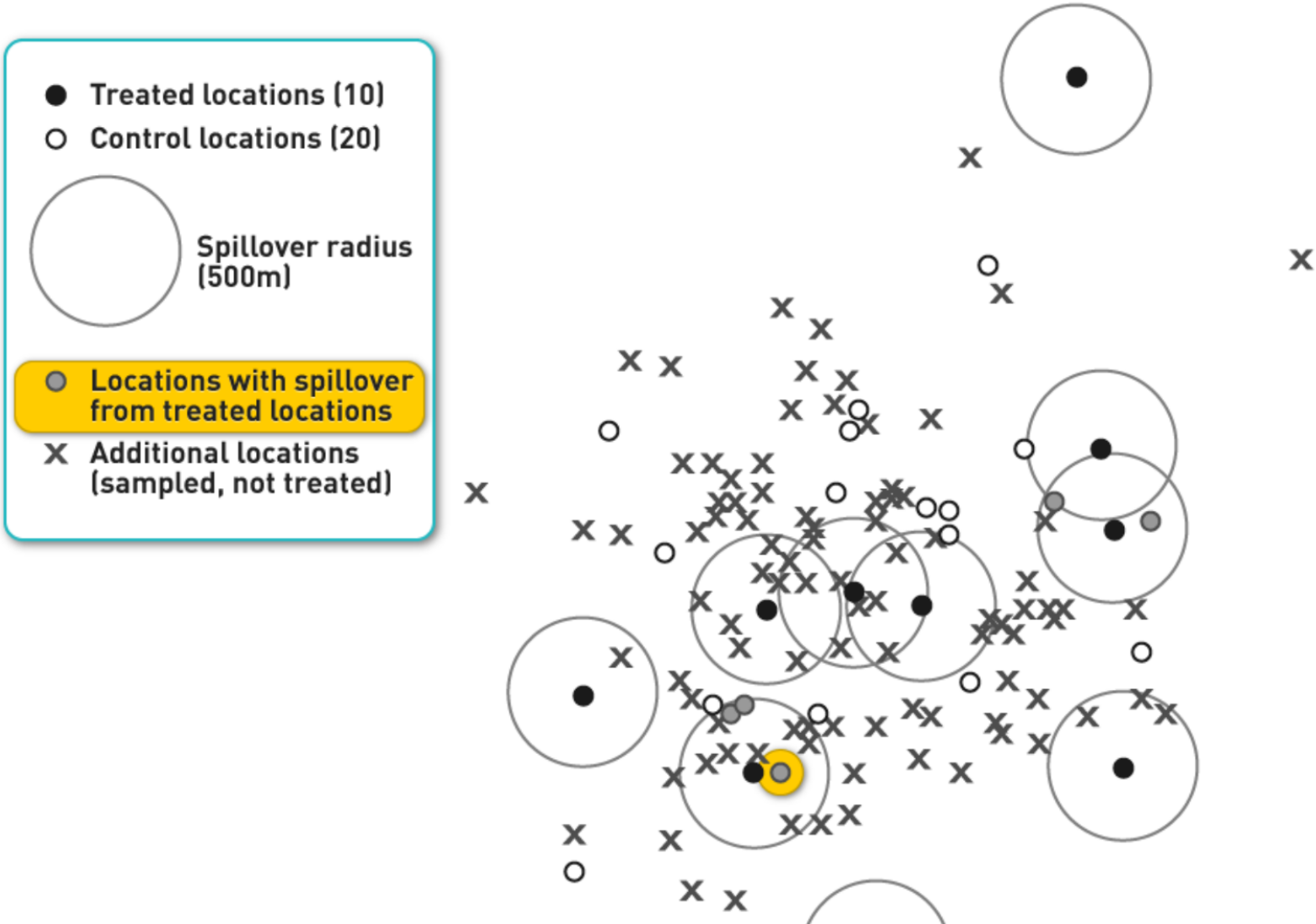


Figure 8.4: Spillover in Two Dimensions

Figure 8.4: Spillover in Two Dimensions

Making Assumptions

Arbitrary assumptions about geographic extent of spillover can radically change estimates

- Eg. what if spillover range is 750 m vs 500 m?
- What if not constant?
  - Linear decline
  - Quadradic decline

Contrast with usual virtue of experiments: make measurements with minimal assumptions

Assumption problems not unique to geography

- Eg. social networks

Reading Assignment

Read *Field Experiments*, Section 8.4

First: A motivating example on longer-run effects of advertising

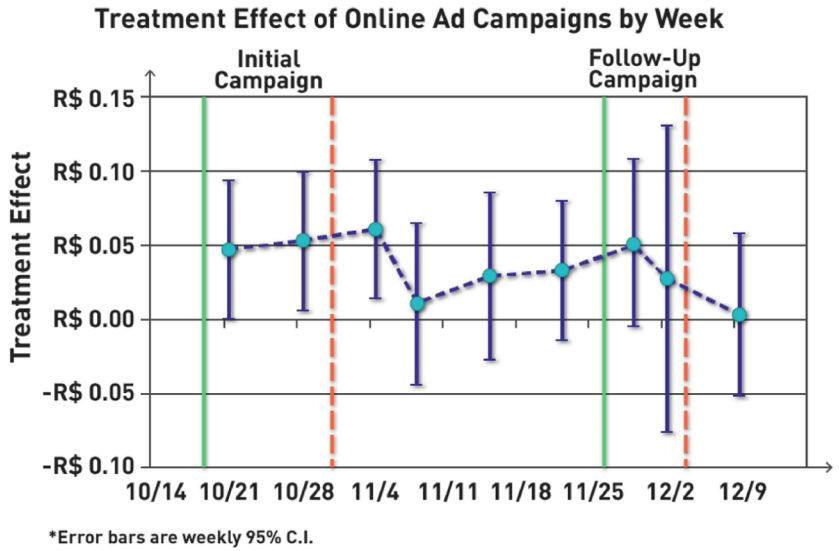
- Lewis and Reiley (2014), Section 4
- Access this paper here

9.9 Spilling Across Time: Example on the Effects of Advertising

Longer-Term Effects of Advertising

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

Figure 5. Weekly DID Estimates of the Treatment Effect



*Error bars are weekly 95% C.I.

Findings

- Treatment point estimates uniformly positive every week
- Difference-in-difference assumption may have caused overestimation
    - All obsvations based on same pre-period
    - Overestimation due to pre-period differences will carry over to all subseguent weeks
- Results show evidence of long-term effects of advertising

Reading Assignment

Field Experiments, Section 8.4

- Temporal spillover effects over time, from present to future
- Within-subjects experimental designs

9.10 Yoga: Estimating in the Presense of Spillover

THE SLIDES FOR THIS SECTION ARE NOT HERE

9.11 Quiz 3

9.12 Within-Subjects Design and Time-Series Experiments

Within-Subjects Design: Why?

- Variance in *y* always 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 do it for 30 days?
- How do I know whether avoiding gluten will help me?
- 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 lost

- Independence from other factors that might cause 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
- Impossible to predict in advance what other events might influence outcome

Ample permutation space for statistical inference

- Best attainable p-value is 0.5

Including many time periods can solve this problem

- Room for randomization
- Enough data for valid standard-error estimates

9.13 Quiz 4

9.14 Yoga: Estimating with Clustered Standard Errors

THERE ARE NO SLIDES HERE

9.15 Review: Estimating with Clustered Standard Errors

9.16 Guide to Reading: Waitlist Designs

Reading Assignment

Field Experiments, Section 8.5

- Waitlist designs
- First three paragraphs

The example in the rest of this section is instructive if you wish to read it, but the details are beyond our scope

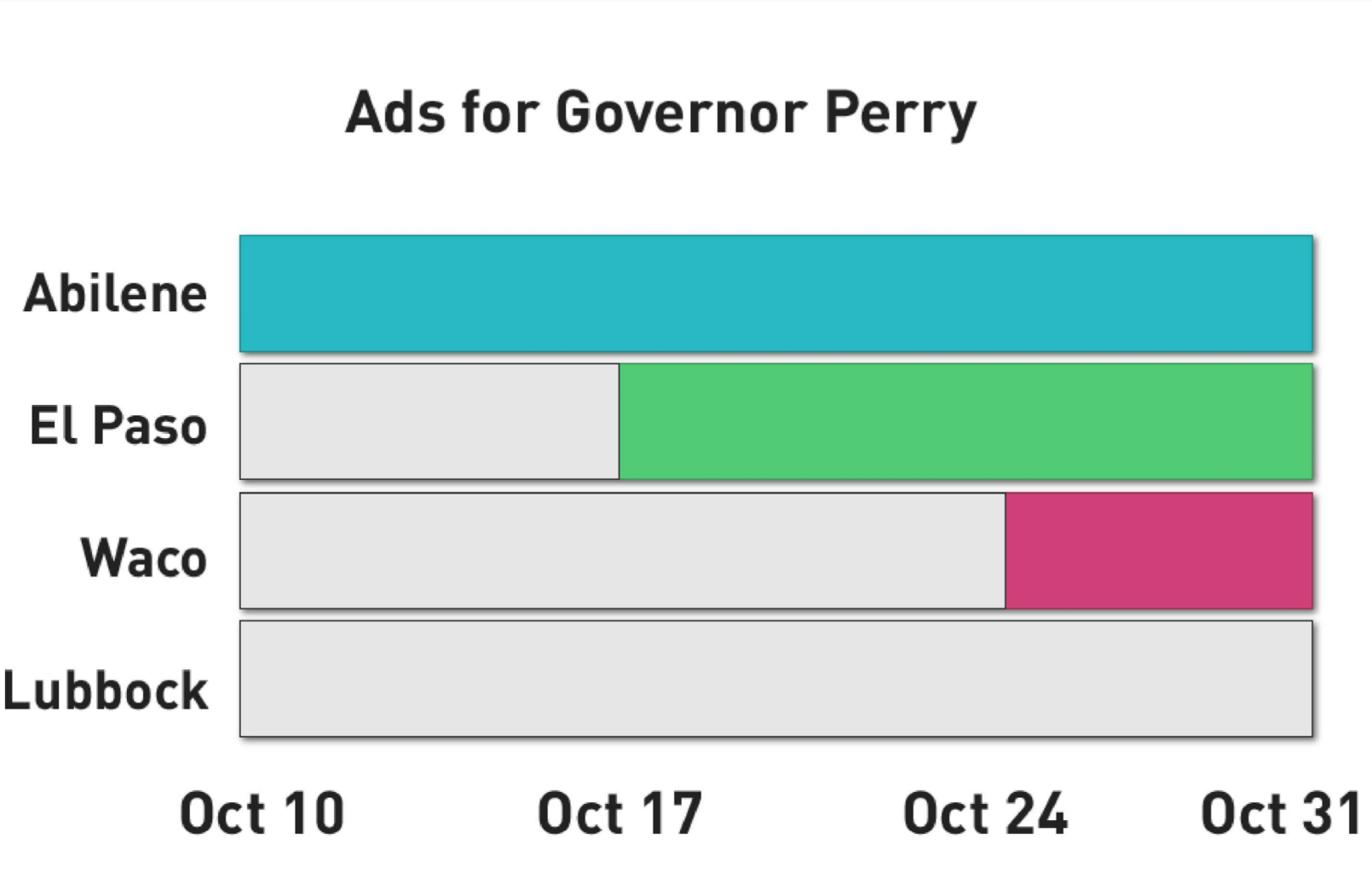
9.17 Quiz 5

9.18 Waitlist Designs

Waitlist Design

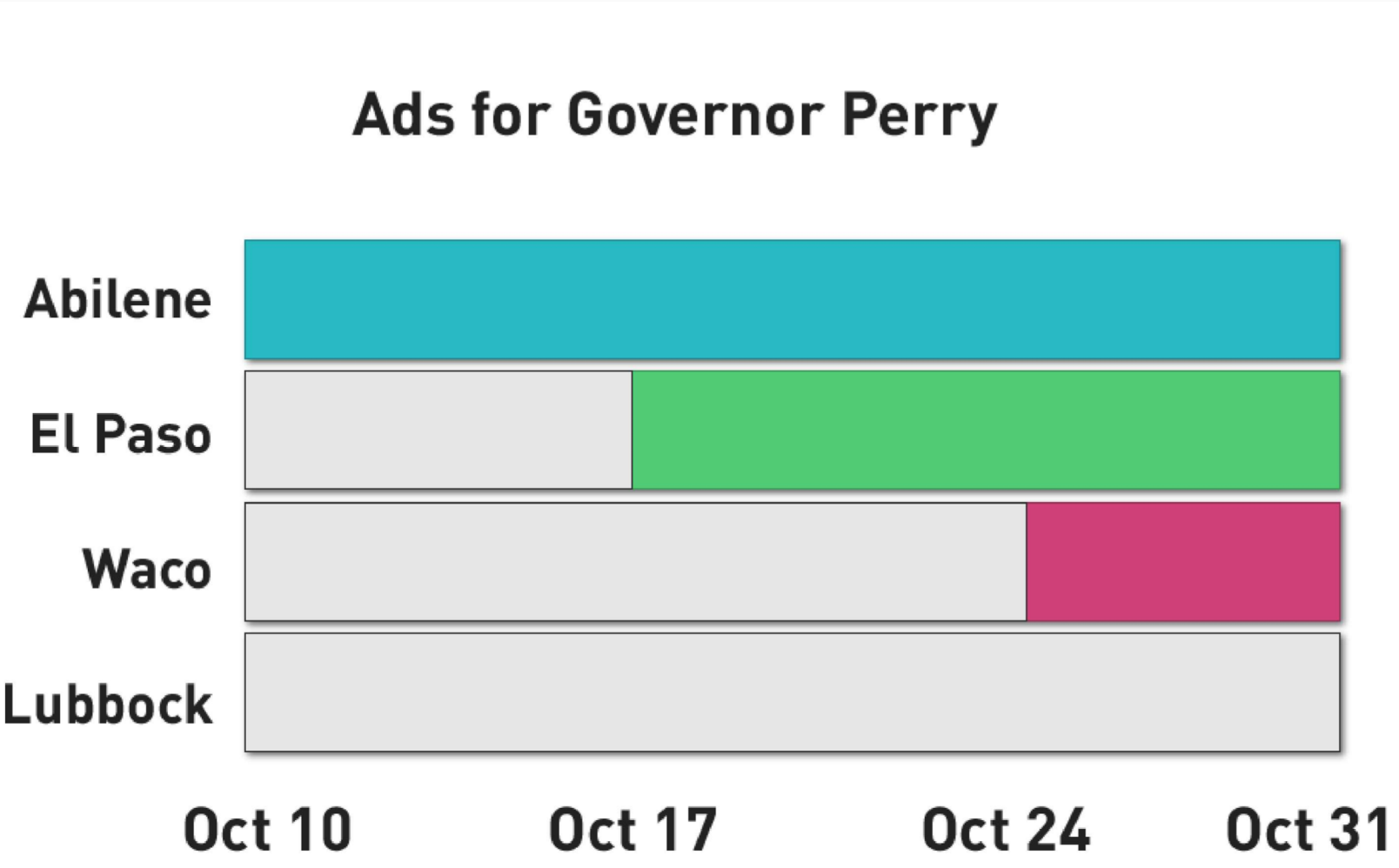
Administer randomized treatment to all subjects over time

- Roll out treatment to control group in stages
- Also known as *Stepped-wedge* designs


Ads for Governor Perry
Abilene
El Paso
Waco
Lubbock
Oct 10
Oct 17
Oct 24
Oct 31

Benefits of Waitlist Design

- Allows experimentation
- Allays concerns about witholding treatment
- Can delay widespread treatment long enough to see effects in early treatment groups before administering to all groups
- Long-run *and* simultaneous treatment effects


Ads for Governor Perry
Abilene
El Paso
Waco
Lubbock
Oct 10
Oct 17
Oct 24
Oct 31

9.19 Yoga Example: Unsccessully Estimating in the Presence of Spillover

What to Remember From This Week

All experiments aim for *Ceteris paribus* "all other things being equal" except treatment effect

Spillovers between subjects can bias results:

- Positive spillovers: underestimation (eg. I see ads and take my friend to the store)
- Negative spillovers: overestimation (eg. extra police divert crime to a difference area)

Clustered treatment can eliminate this bias

Within-subjects design can be valuable if treatment effects are limited in time

Stepped-wedge designs get variation both within and between subjects