Review

- Compliance problems
  - Incomplete control over treatment delivery
    - Intent to treat
    - Placebo design
  - Accidental delivery to control group

More Implementation Problems

- Randomization system doesn't work as intended.
- Treatment not delivered or received (e.g., letters not read, e-mails not opened).
- Unintended effects of treatment.

Hawthorne Effect

- Effect of change
- Named after social experiment at Hawthorne factory
  - Effects on worker productivity.
    - Brightness of lights
    - Temperature in factory
  - Every variable **seemed**to affect productivity.
    - Effects were only **temporary**.

Hawthorne Effect (contd)

- Websites
  - Changes to a site cause users to act differently.
  - User activity often normalizes after adjusting to change.
  - Example: Change in purchase flow might cause **temporary** conversion increase.

Demand Effects

- Subjects who are aware of being studied often give different answers.
    - Subjects may give answers they think researcher wants to hear.
        - Treatment can alert subjects to what researchers are looking for.
    - Example: study to determine price for new product.
        - Subjects might lie if they feel they can actually affect price of new product.
- Researchers strive to conceal connection between survey and treatment.

Blind Trials

1. Single-blind trial
    - Subject doesn't know.
2. Double-blind trial
    - Experimenter doesn't know.
3. Triple-blind trial
    - Analyst doesn't know group assignments.
    - Eliminates fishing expedition problem.
    - Biased analysis may cause study to be halted prematurely.

Unexpected Results of Study Implementation

- Abstract concepts may result in missed real-world contexts.
  - Example: restorative justice meetings
    - Attempt at reconciliation between criminals and victims.
    - Criminals didn't show up, further damaging victims.
    - Pilot study could have prevented problem.
- Field workers refuse to randomize; think they're helping.
  - Example: canvassers who knock on every door
    - Unintentionally ruins control group
  - Example: TiVo's effect on people remembering advertisements
    - Trouble getting randomized group to redeem TiVo device
    - Had to treat some in control group as well

Overview

- Pilot studies
  - Discover unanticipated problems.
- Placebo tests
  - No effect desired
- Manipulation checks
  - Was treatment successfully delivered?
  - Large effect desired.

Pilot Studies

- Always run a pilot study...no excuses!
- Unanticipated reactions.
- Example: fake emails to state legislators.
  - Intended to study responses of legislators based on geographic origin of messages
  - Spillover between legislators
- Flawed power calculations
  - Likelihood of effects of different sizes
  - Baseline
  - Response rates
  - Variance of y

Pilot Studies (contd)

- Training staff
  - Correct implementation of experiment
- Determining if systems really work
  - Subtle issues can arise.
- New ideas from implementation of treatment
  - Potential for improvement

Placebo Tests

- If experiment/experimentation system works, there should be **no difference**in a variable.
  - Difference in variable suggests problem with system.
- A/A test:
  - Tests a treatment against itself to detect any difference in outcome
- Traditional placebo test:
  - Checks other outcomes that treatment shouldn't affect

A/A Placebo Test Failing

- Website optimizer code.
- In-house system randomly assigned users to page A or page B.
  - Control: User arrived on and loaded page A.
  - Treatment: User arrived on page A and page B loaded.
  - Outcome measured: purchases on page.

A/A Placebo Test Failing (contd)

- Issues:
  - Load time for treatment group was longer.
  - Treatment group code didn't run on all web browsers.
    - Some users always in control group.
  - Unrealized manipulations occurred, e.g., delay loading web page.
  - Page A consistently bested page B, despite pages being identical.
    - Natural bias led to fake conclusions about efficacy of changes.

Placebo Test Failing

- Great tool for questioning observational studies.
  - To prove bias, look for conditions that shouldn't exist if assumption is right.
- Krueger (1993) found workers who use computers earn 15–20% more.
  - Ideal experiment: Computers increase productivity, warranting higher wages.
  - DiNardo and Pischke saw potential bias.
    - Found workers who use pencils, pens, and phones also earn more.
    - Workers who use screwdrivers earn less.
    - Wages depend on types of jobs that use these tools, not use of tools themselves.

Another Placebo Test Failing

- Can't compare compliers to everyone in control group
  - Have to compare entire treatment group to entire control group, or treatment compliers to control compliers

Social Influence Example

- Do people influence their friends?
  - Ask for person's weight and weight of friend.
  - Social network models find that being fatter causes friends to get fatter (causal claim).
    - Likely **not** causal relationship
  - What can be controlled for?
  - Same model finds being tall causes friends to get taller.
  - Social network models struggle to recover causal effects.

Manipulation Checks

- Successful experiment **should** find differences in variable.
- Example: Broockman and Butler (2014) worked with legislators to send mail.
    - Control group received no letters.
    - Did constituents receive letters and learn their legislator's positions?
    - Asked people if they recalled recently receiving a letter from legislator.
        - People in treatment group more likely to report having received letter, understand legislator's position.
        - Can't use this information in estimation.

Manipulation Checks (contd)

- Example: Sending postcards to voters causes 10 percent increase in understanding candidate's position.
    - Can't use information in same way as one-sided non-compliance
    - Way of making sure treatment had expected effect
- Important for proving null effect.
    - Was treatment actually delivered?

Covariate Balance Checks

- Experiments guarantee balance in observable and unobservable characteristics.
- Check balance on observable covariates to ensure random assignment was done correctly.
- Example of balance not holding: Users in treatment group are more active Internet users.
    - Might give impression of not conducting clean random assignment
- Especially important when randomization scheme is complex.
    - Blocking, clustering, different probabilities
    - Complex systems between researchers and subjects
- See regression week for more on this.

Example of Covariate Balance Failing

- Lewis and Reiley's first dataset
    - Strange patterns
        - Negative treatment effect
    - Large imbalance on pretreatment sales
        - Control group bought more in preperiod than treatment group.
        - Vendor had truncated data by number of sales.
        - Treatment group sales data were lost.

Preventing and Detecting Problems

- Conduct a pilot study.
- Manipulation check to measure delivery of treatment.
- Placebo test.
- Check for covariate balance.

Advocating Experimentation

1. Increase perceived benefits.
2. Decrease perceived costs.

Increase Perceived Benefits

- Stimulate curiosity.
    - Intellectual interest.
    - Get people excited.
- Vivid examples of current data leading to bad decisions.
    - Conduct placebo test.
        - Show practices failing.
    - Present potential conclusions that experiment could produce.
        - Ways those conclusions could change practices
    - Tell story of causal inferences leading to incorrect conclusions.
        - Example: Playing outside improves children's eyesight.
        - How did people receive treatment?
        - What's different about those people?
            - Reasons people differ

Increase Perceived Benefits (contd)

- "Investment in information for future decisions."
  - Example: Firm worries that holding back advertising for sake of experimenting will cost them money.
  - Short-term sacrifices pay off in future.
  - I.e., down payment.
- Build rapport.
  - Personal connections lead to willingness to run experiments.
- Do small studies first as proofs of concept.
  - Example: Broockman and Butler (2014)
    - One congressman signing on for small study led to results that garnered interest for larger study.
  - Helps secure greater cooperation

Decrease Perceived Costs

- Administration, unfairness, giving up potential gains from treatment (e.g., ads)
- Delay for some units
  - Randomize order of mailings.
  - Campaign contributions.
    - Spread out donations to allow experimentation.

Decrease Perceived Costs (contd)

- Limited resources
  - Example: Charity gives 5,000 bed nets to people in Africa.
    - Withholding bed nets seems morally objectionable.
    - Expanding population creates control group without withholding bed nets.
- Experimentation as investment in information

Building Knowledge Over Time

- Try pilot studies.
- System should produce covariate balance.
- Parameters should look as expected.
    - E.g., 80/20 treatment/control split.
    - Example: Internet Explorer skewed data.
- Understand parameters for power analysis.
- When creating a system, studies should be useful.

Pooling Results

- Precision-weighted average:

1 standard error 2

- Double the standard error gets one-fourth the weight.
- One-fourth sample size means one-fourth the information.
- If variance is the same, weight by number of subjects.
- Standard error used to represent adjustment of overall view of treatment.
- Meta-analysis summarizes data.