

Field Experiments I

POLSCI 4SS3

Winter 2023

Last time

- Hypothesis testing as a standard to evaluate if experiments suggest that policies works
- We don't know how things look when the policy work, but we can contrast with the hypothetical world in which it doesn't work
- We can only do this if we know how the experiment was conducted!
- **Lab:** Statistical power as a diagnosand to determine if an experiment is well equipped to detect “true” effects
- **Today:** Start applying these concepts to **field experiments**

Field Experiments

- **Field:** Interventions in real world settings (*vs. surveys, laboratory*)
- **Experiment:** Randomization determines assignment of units to conditions
- *AKA randomized controlled trials in policy or A/B testing in industry*
- **Core idea:** Randomization allows us to produce credible evidence on whether something works
- **In practice:** A lot of implementation details and research design choices to navigate

Examples

Banerjee et al (2021): TUP program

- **Poverty trap:** Most programs to help the poor improve living conditions in the short term, but revert afterwards
- **Solution:** Multidimensional “big push” to overcome poverty traps
- Evaluate long-term effect on poorest villages in West Bengal, India

Household eligibility criteria

1. Able bodied female member (why?)
2. No credit access

AND at least three out of

1. Below 0.2 acres of land (about 2 basketball courts)
2. No productive assets
3. No able-bodied male member
4. Kids who work instead of going to school
5. No formal source of income

Data strategy

- **Sample:** 978 eligible households
- **514 assigned** to treatment
- **266 accepted** treatment

What was the treatment?

Program

- Choose a productive asset (82% chose livestock)
- Weekly consumption support for 30–40 weeks
- Access to savings
- Weekly visits from program staff over a span of 18 months

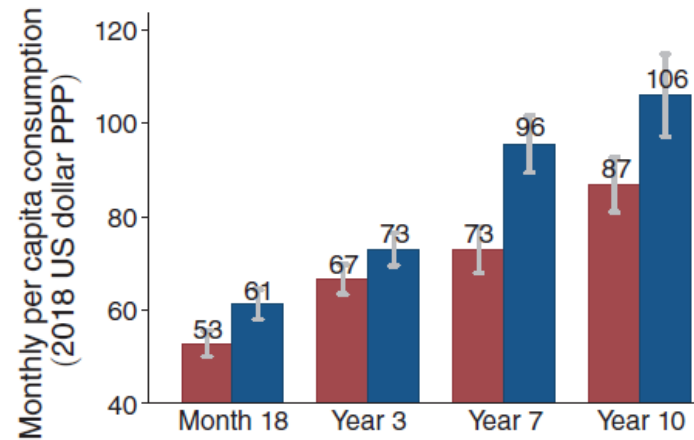
Why would someone reject this?

Answer strategy

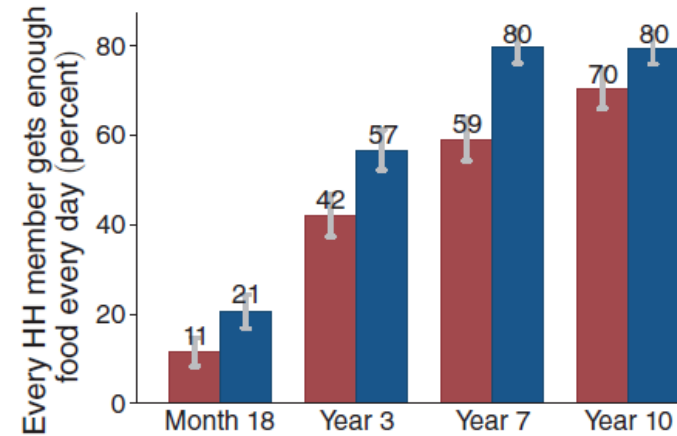
- Track economic and health outcomes after 18 months, 3, 7, and 10 years
- Of all household members
- Focus on average treatment effect **among the treated**
(more in the lab)

Findings

Panel A. Per capita consumption
(2018 US dollar PPP)



Panel B. Every household member gets enough
to eat every day (percent)



Panel C. Income (2018 US dollar PPP)

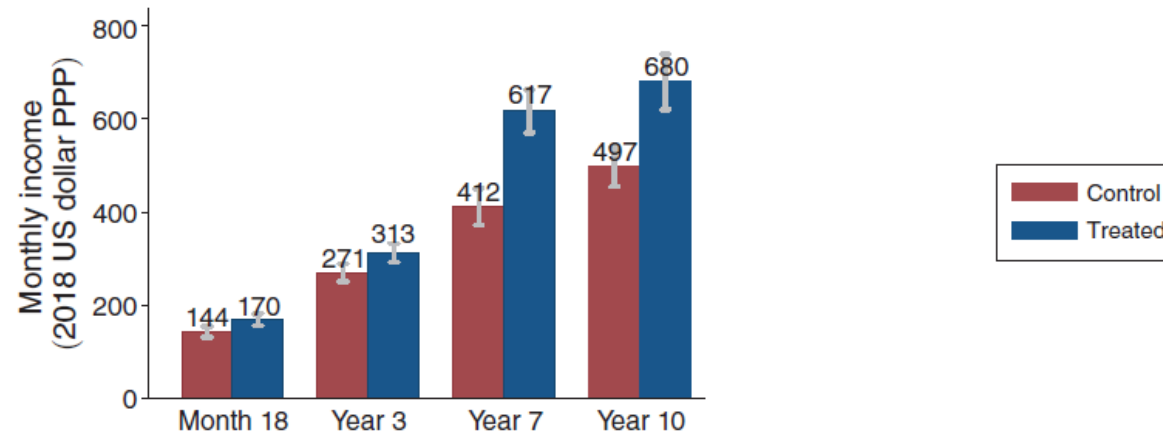


FIGURE 1. CONSUMPTION AND INCOME OVER TIME

Why does this work?

Time	ATE on income from...			
	Livestock	Micro-enterprise	Self-employment	Remittances
18 months	10.26	7.93	18.67	0.00
3 years	7.68	25.12	31.06	3.70
7 years	27.26	67.59	108.36	8.87
10 years	16.71	36.82	93.87	19.06

- **Takeaway:** Big push works because it helps people diversify their income sources over time

Pennycook et al (2021): Shifting attention to accuracy can reduce misinformation online

- Why do people share fake news in social media?

Three explanations:

1. Confusion about accuracy
2. Partisanship > accuracy
3. Inattention to accuracy

Study 7: Application to Twitter

- Studies 1-7 were all survey experiments
- Study 7 deploys intervention on Twitter to see if priming accuracy works
- N = 5,739 users who previously shared news from untrustworthy sources
- **Treatment:** Send a DM asking to evaluate accuracy of news article

Challenge to data strategy

- Can only send DM to someone who follows you
- Need to create bot accounts and hope for follow-backs
- Identify those who retweet fake news
- Limit 20 DMs per account per day
- 3 waves with many 24-hour blocks in each

HOW ACCURATE IS THIS HEADLINE?



Woman who had ovary frozen in childhood give...

She is believed to be the first woman in the world to have a baby after having ovarian tissue frozen befo...

[surveycamel.com](https://www.surveycamel.com)

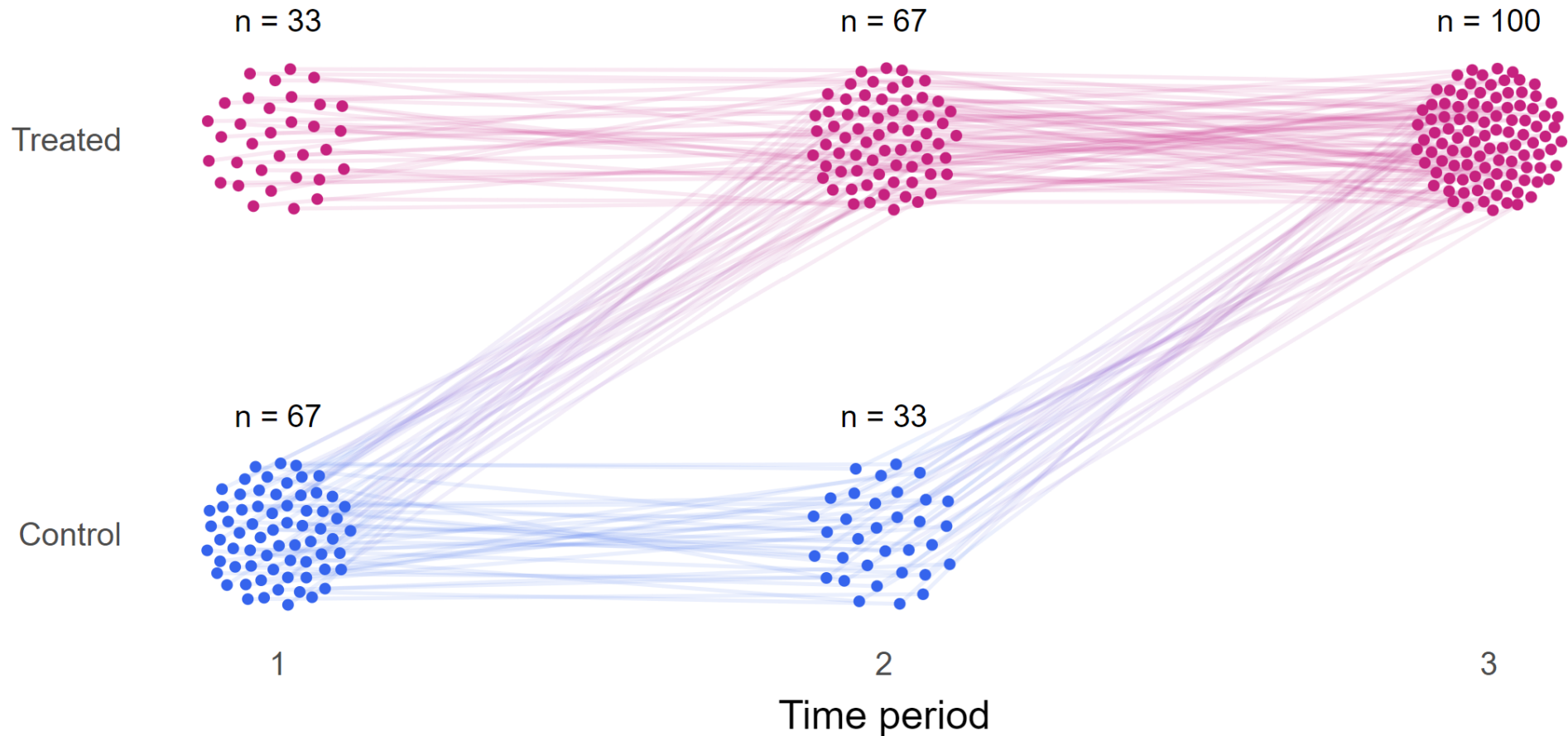
Thanks for following me! Can I ask you a favor? I'm wondering how accurate the above headline is, and I'm doing a survey to find out.

[surveycamel.com/ze/news/story5...](https://www.surveycamel.com/ze/news/story5...)

Based on the headline, do you think it is accurately describing something that actually happened?

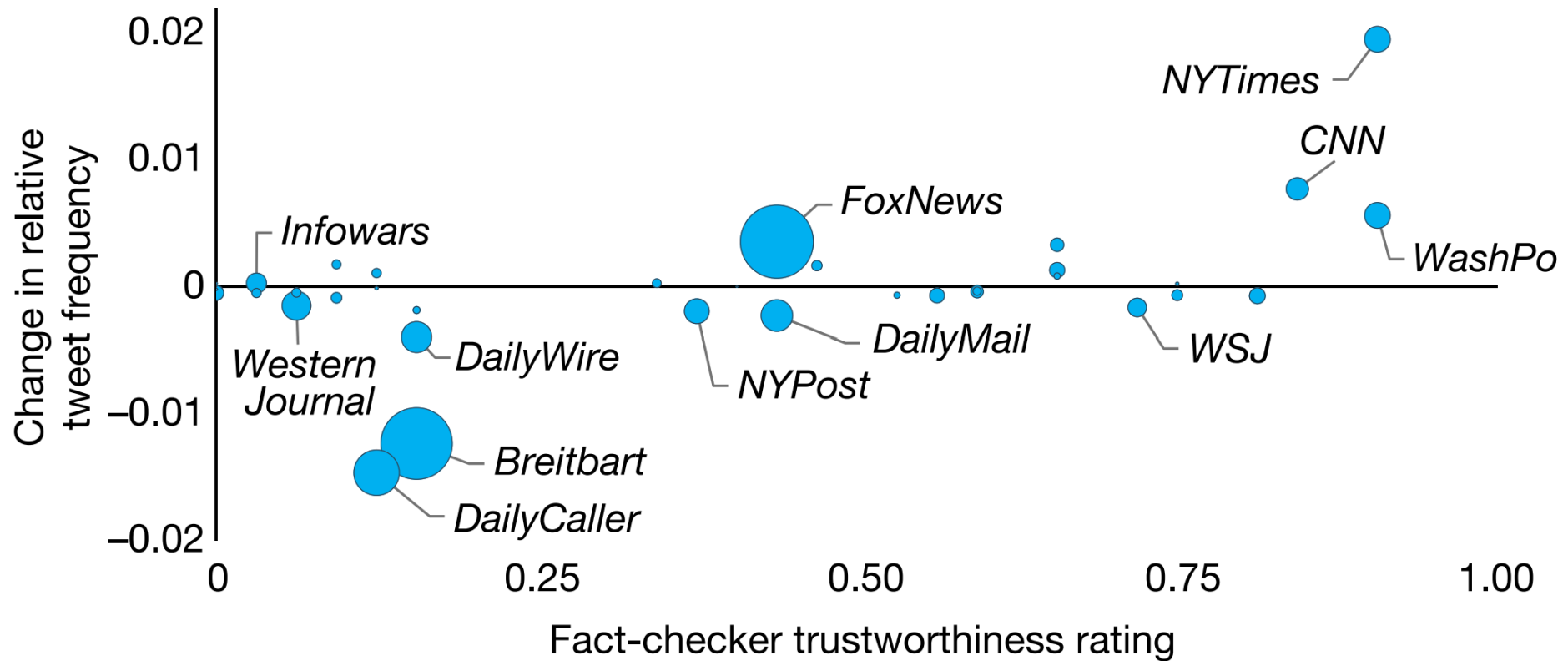
Please rate as: 1=Not at all accurate, 2 = Not very accurate, 3= Somewhat accurate, 4 = Very accurate

Stepped-wedge design



AKA staggered adoption design

Findings



The size of each dot represents the proportion of pre-treatment posts from that outlet

Next Week

More field experiments

Focus on: Sections 1-3 of Diaz and Rossitter (2023)

Break time!





