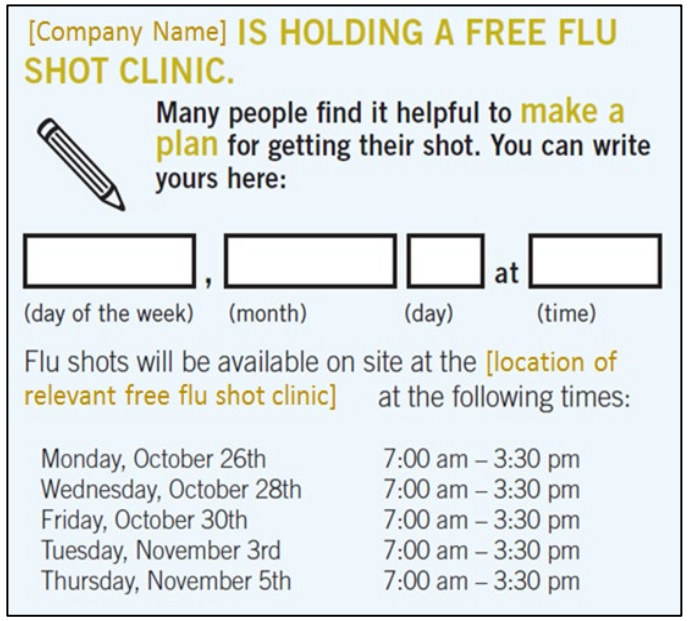
High-Throughput Experimentation and Experimental Design Innovations

Katherine L. Milkman
James G. Dinan Professor at The Wharton School
The University of Pennsylvania

Joint work with Angela L. Duckworth, Linnea Gandhi, Dena M. Gromet, Heather N. Graci, Hung Ho, Joseph S. Kay, Timothy W. Lee, Pepi Pandiloski, Yeji Park, Aneesh Rai, Jake Rothschild, Modupe N. Akinola, Max Bazerman, Ron Berman, John Beshears, Jon E. Bogard, Lauri Bonacorsi, Ilana Brody, Alison M. Buttenheim, Colin Camerer, Christopher F. Chabris, Edward Chang, Gretchen B. Chapman, James J. Choi, Robert Cialdini, Hengchen Dai, Jennifer Dannals, Lauren Eskreis-Winkler, Chalanda N. Evans, Ayelet Fishbach, Craig R. Fox, Noah Goldstein, Amir Goren, James J. Gross, Hal Hershfield, Matt D. Hilchey, Alex Hirsch, Jillian Hmurovic, Samantha Horn, Alexa Hubbard, Leslie K. John, Steven J. Jones, Dean S. Karlan, Tim Kautz, Melanie Kim, Erika Kirgios, Joowon Klusowski, Ariella Kristal, Rahul Ladhania, David I. Laibson, Catherine Lamberton, George Loewenstein, Jens Ludwig, Brigitte C. Madrian, Barbara Mellers, Michelle N. Meyer, Maria Modanu, Sendhil Mullainathan, Jimin Nam, Allison Oakes, Mitesh S. Patel, Todd Rogers, Renante Rondina, Silvia Saccardo, Maurice Schweitzer, Maheen Shermohammed, Christopher K. Snider, Dilip Soman, Jehan Sparks, Jann Spiess, Gaurav Suri, Joachim H. Talloen, Jamie Taxer, Eli Tsukayama, Yaacov Trope, Lyle Ungar, Christophe Van den Bulte, Kevin G. Volpp, Caleb Warren, Megan Weber, Ashley Whillans, Kuldeep Yadav, Jonathan Zinman, and Julian Zlatev

FIELD EXPERIMENTS: THE IDEAL METHOD FOR BEHAVIORAL SCIENTISTS TACKLING POLICY CHALLENGES

For Example: Could **prompting planning in reminder mailings** boost vaccination during the H1N1 pandemic?



[Company Name] IS HOLDING A FREE FLU SHOT CLINIC.

Many people find it helpful to make a plan for getting their shot. You can write yours here:

_____ , _____ _____ at _____

(day of the week) (month) (day) (time)

Flu shots will be available on site at the [location of relevant free flu shot clinic] at the following times:

Monday, October 26th	7:00 am – 3:30 pm
Wednesday, October 28th	7:00 am – 3:30 pm
Friday, October 30th	7:00 am – 3:30 pm
Tuesday, November 3rd	7:00 am – 3:30 pm
Thursday, November 5th	7:00 am – 3:30 pm

FIELD EXPERIMENTS: THE IDEAL METHOD FOR BEHAVIORAL SCIENTISTS TACKLING POLICY CHALLENGES

Thousands of Employees Randomized

Control

[Company Name] IS HOLDING A FREE FLU SHOT CLINIC.

Flu shots will be available on site at the [location of relevant free flu shot clinic] at the following times:

Monday, October 26th	7:00 am – 3:30 pm
Wednesday, October 28th	7:00 am – 3:30 pm
Friday, October 30th	7:00 am – 3:30 pm
Tuesday, November 3rd	7:00 am – 3:30 pm
Thursday, November 5th	7:00 am – 3:30 pm

Time Plan

[Company Name] IS HOLDING A FREE FLU SHOT CLINIC.

Many people find it helpful to make a plan for getting their shot. You can write yours here:

_____, _____ _____ at _____

(day of the week) (month) (day) (time)

Flu shots will be available on site at the [location of relevant free flu shot clinic] at the following times:

Monday, October 26th	7:00 am – 3:30 pm
Wednesday, October 28th	7:00 am – 3:30 pm
Friday, October 30th	7:00 am – 3:30 pm
Tuesday, November 3rd	7:00 am – 3:30 pm
Thursday, November 5th	7:00 am – 3:30 pm

FIELD EXPERIMENT: CONFIRMED VALUE OF PLANNING PROMPTS TO BOOST VACCINATION

[Company Name] IS HOLDING A FREE FLU SHOT CLINIC.

Flu shots will be available on site at the [location of relevant free flu shot clinic] at the following times:

Monday, October 26th	7:00 am – 3:30 pm
Wednesday, October 28th	7:00 am – 3:30 pm
Friday, October 30th	7:00 am – 3:30 pm
Tuesday, November 3rd	7:00 am – 3:30 pm
Thursday, November 5th	7:00 am – 3:30 pm

[Company Name] IS HOLDING A FREE FLU SHOT CLINIC.

Many people find it helpful to make a plan for getting their shot. You can write yours here:

_____, _____ _____ at _____

(day of the week) (month) (day) (time)

Flu shots will be available on site at the [location of relevant free flu shot clinic] at the following times:

Monday, October 26th	7:00 am – 3:30 pm
Wednesday, October 28th	7:00 am – 3:30 pm
Friday, October 30th	7:00 am – 3:30 pm
Tuesday, November 3rd	7:00 am – 3:30 pm
Thursday, November 5th	7:00 am – 3:30 pm

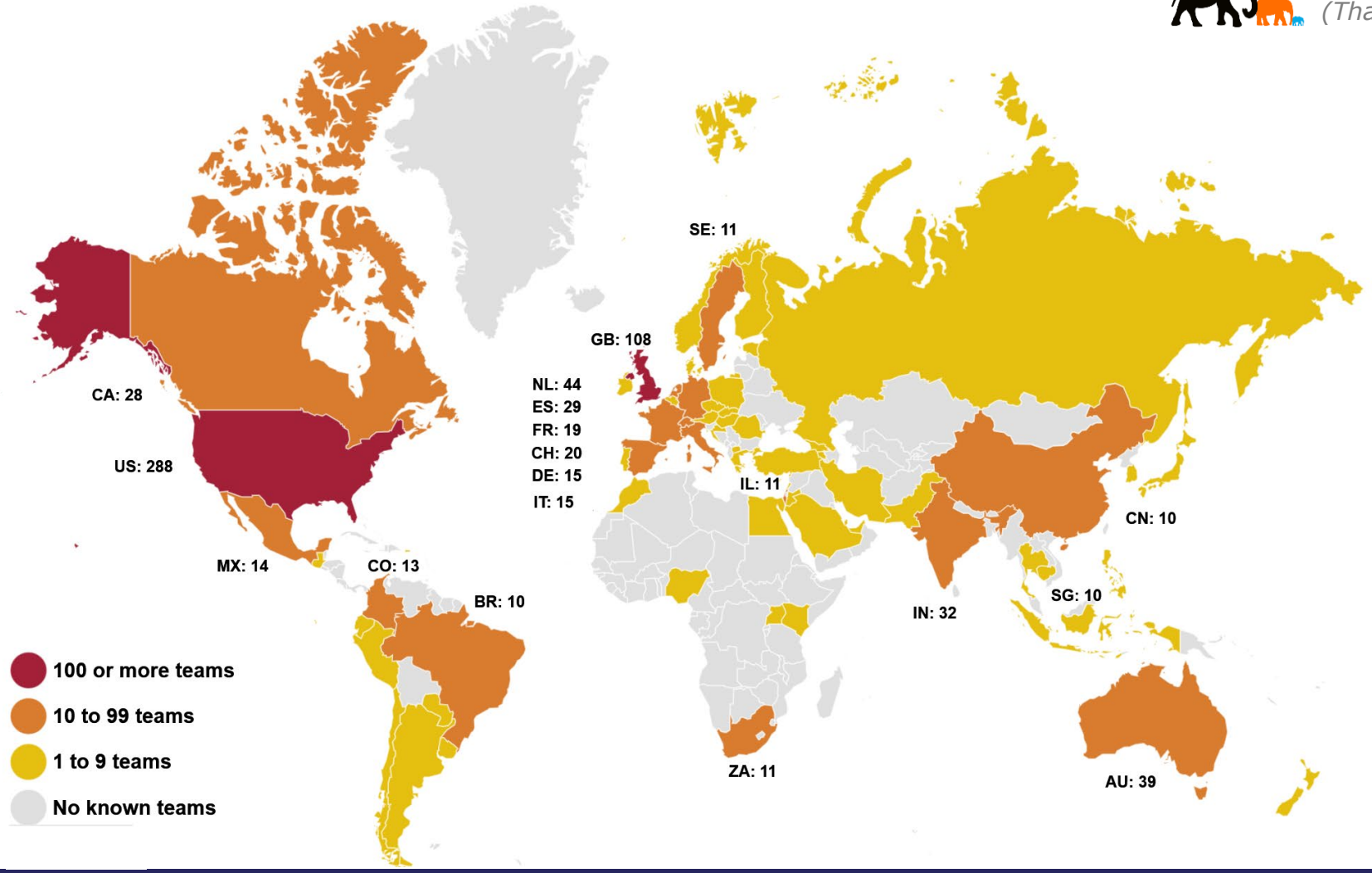
12%

increase in vaccinations

Milkman, Beshears, Choi and Madrian (2011)

HUNDREDS OF NUDGE UNITS OPERATING GLOBALLY, IN GOVERNMENTS, RUNNING FIELD EXPERIMENTS

(Thaler and Sunstein, 2008)



THE AVERAGE EFFECTS OF GOVT NUDGE UNITS' INTERVENTIONS AREN'T HUGE (1.39 pp, or 8%)

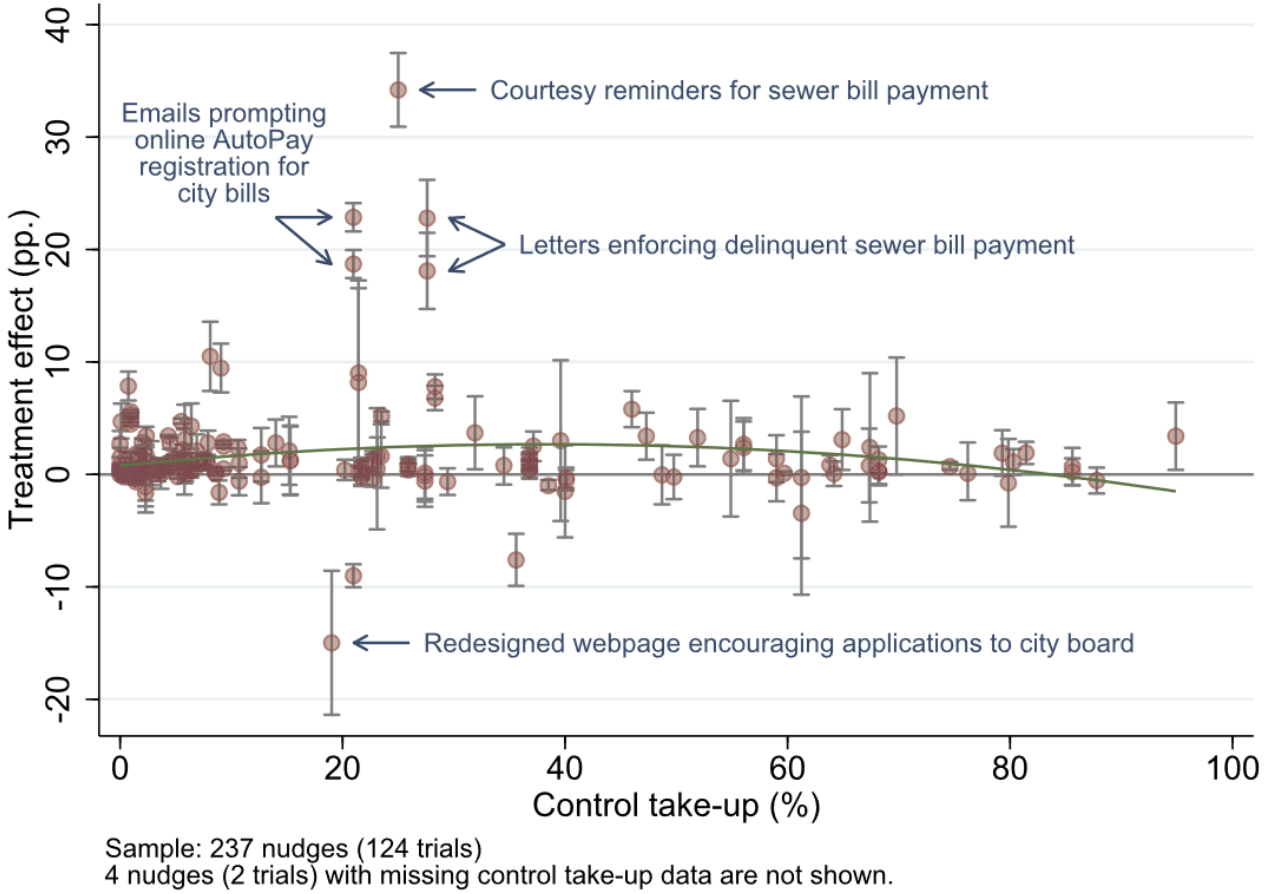
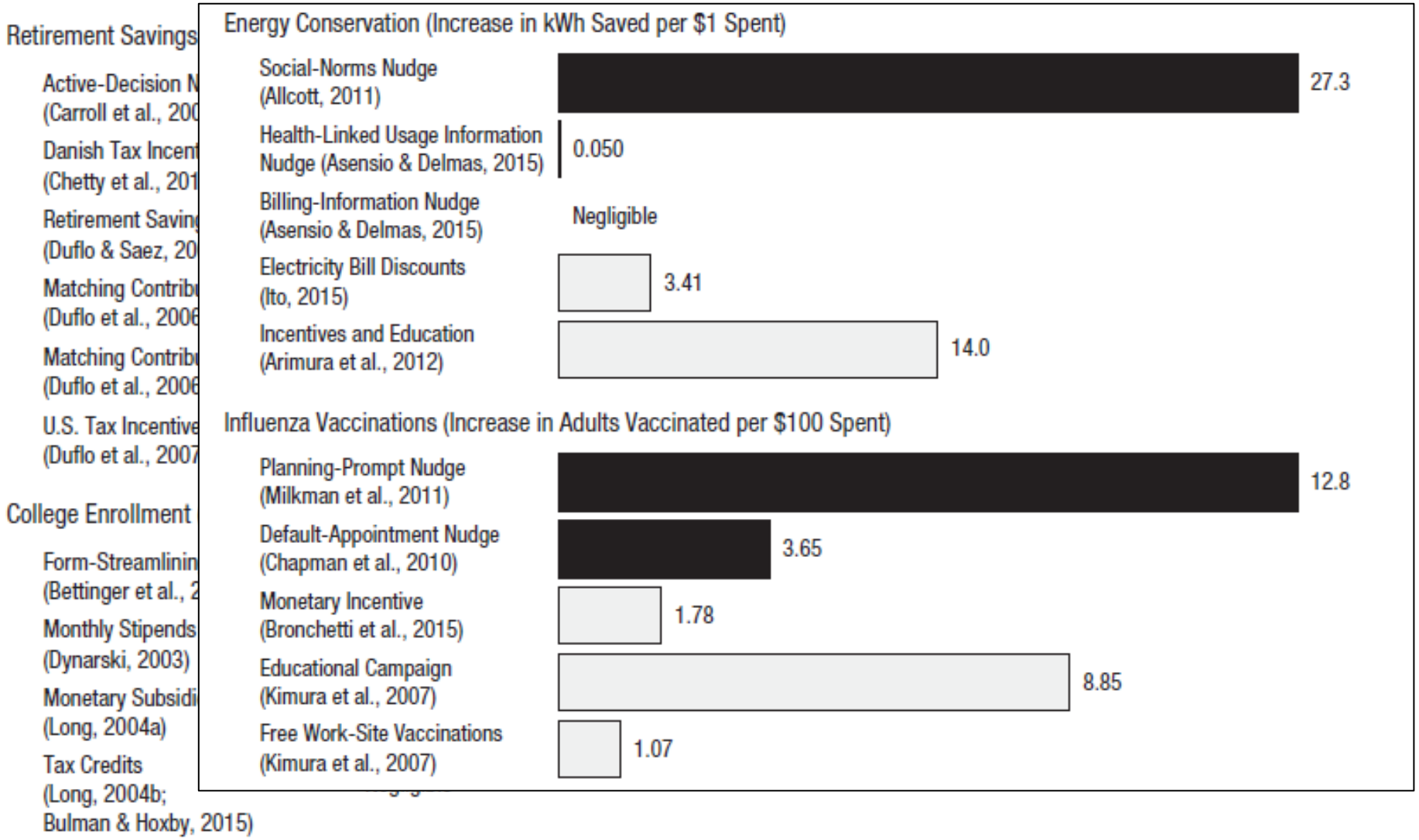


FIGURE 3.—This figure plots the treatment effect relative to control group take-up for each nudge with the quadratic fit. Some of the outliers are labeled for context. Error bars show 95% confidence intervals.

DellaVigna and Linos (2022)

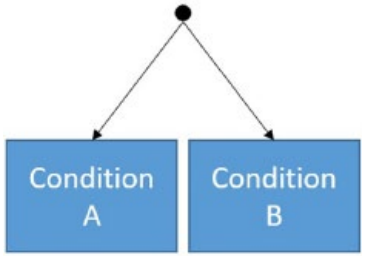
BUT LEVERAGING NUDGES TO PROMOTE POLICY GOALS IS *VERY* COST EFFECTIVE…



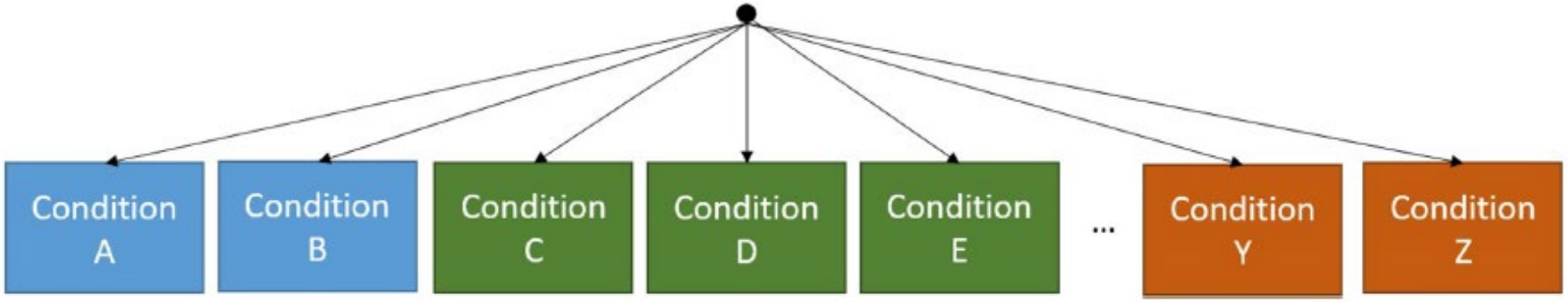
Benartzi et al. (2017)

HOW CAN WE EFFICIENTLY ASSESS WHICH BEHAVIORAL INSIGHTS ARE MOST USEFUL?

Traditional Field Experiment



Megastudy



A megastudy is a ***very large*** field experiment in which many smaller, sub-experiments are run synchronously with the same dependent variable

Milkman et al. (2021). Megastudies Improve the Impact of Applied Behavioural Science. *Nature*.
Duckworth and Milkman (2022). A Guide to Megastudies. *PNAS Nexus*.

BENEFITS OF THE MEGASTUDY APPROACH

1. Ensures **comparability** of results across studies
2. **Fixed costs** of executing the study can be borne by central organizer (low marginal costs)
3. Can run as a **tournament** with interdisciplinary teams
4. **Reduces risk** of learning nothing useful
5. Vastly **accelerates pace of scientific discovery**
6. Can use ML to ask **what works for whom**
7. Ensures **all findings are published** (even nulls)

A SIMILAR IDEA REVOLUTIONIZED AI

Common Task Framework

- Researchers compete to solve the same problem (e.g., image recognition)
- Subject to the same constraints (e.g., the same validation method)
- Using the same dataset—with complete transparency in terms of hypotheses tested and results

(Liberman and Jelinik, 2010; Donoho, 2015)

Scientific Tournaments

- Similar in flavor but don’t involve random assignment

(Mellers et al., 2014)

BEHAVIOR CHANGE FOR GOOD

OUR SCIENTIFIC TEAM

A TOUR OF FOUR MEGASTUDIES

1. **Exercise at 24 Hour Fitness**
2. Flu Vaccinations at Healthy Checkups
3. Flu Vaccinations at the Pharmacy
4. COVID Boosters at the Pharmacy

A MEGASTUDY WITH *24 Hour Fitness*

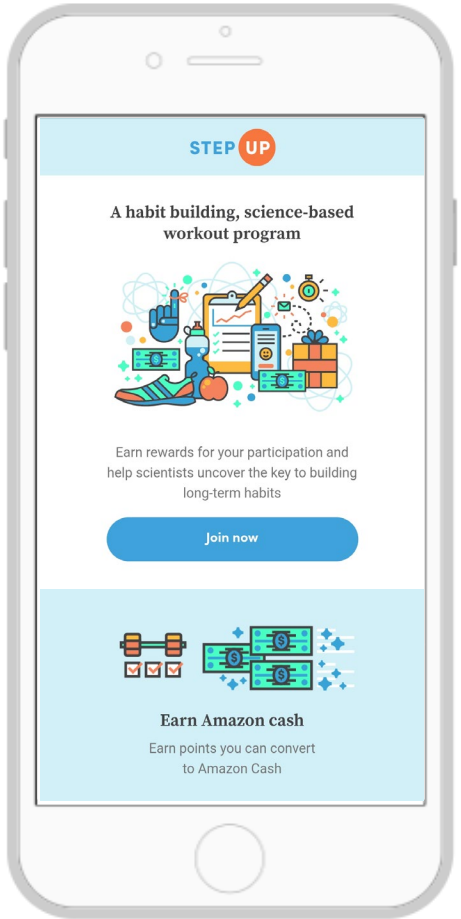
(4 million member gym chain, 430 U.S. locations, 62,746 participants)

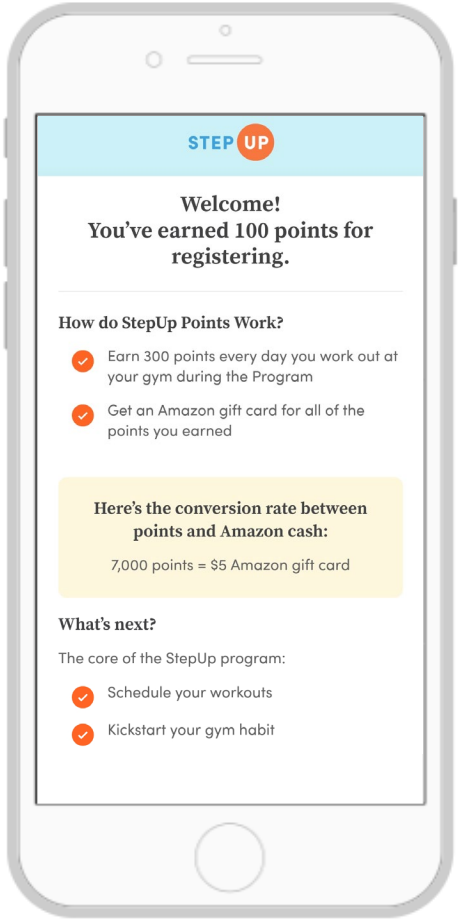
Milkman et al. (2021). Megastudies Improve the Impact of Applied Behavioral Science. *Nature*.

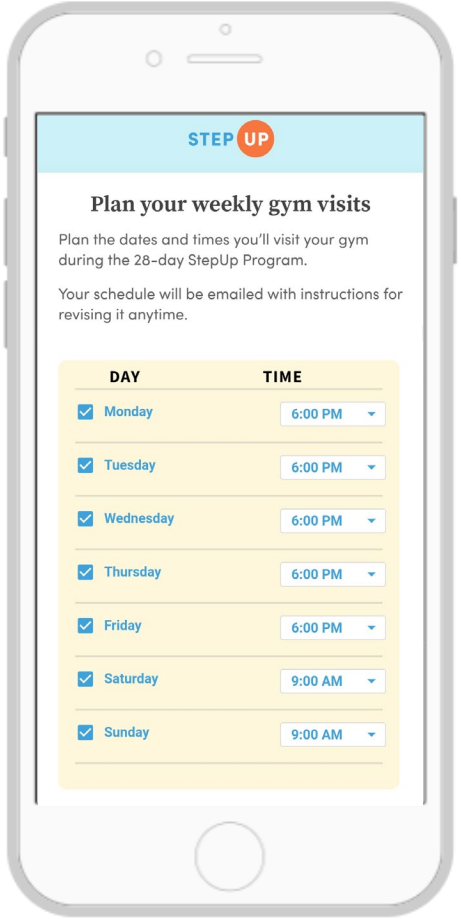
WHAT DID WE TEST?

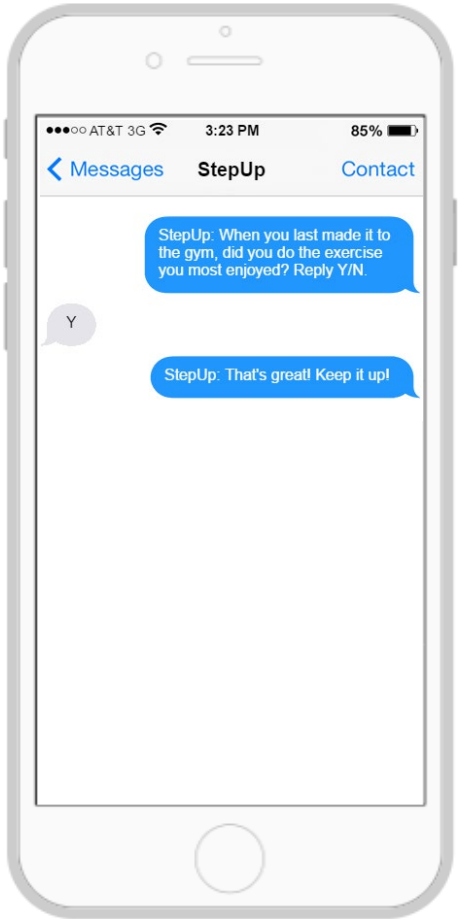
30 scientists designed **20** different pre-registered research studies with **53** different interventions

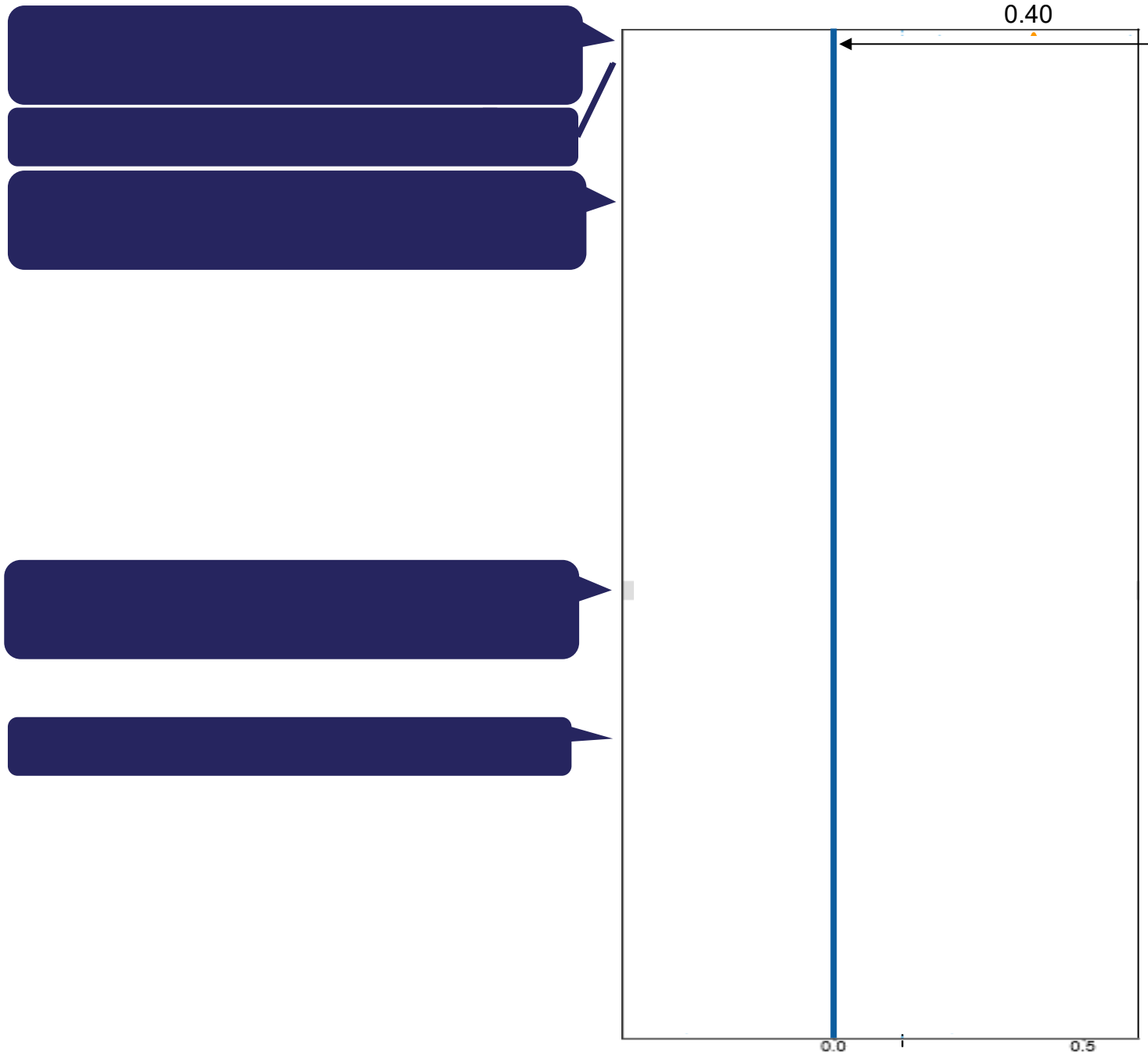
Varied sign-up surveys, incentives, reminders, interactive texts and weekly emails for 28 days





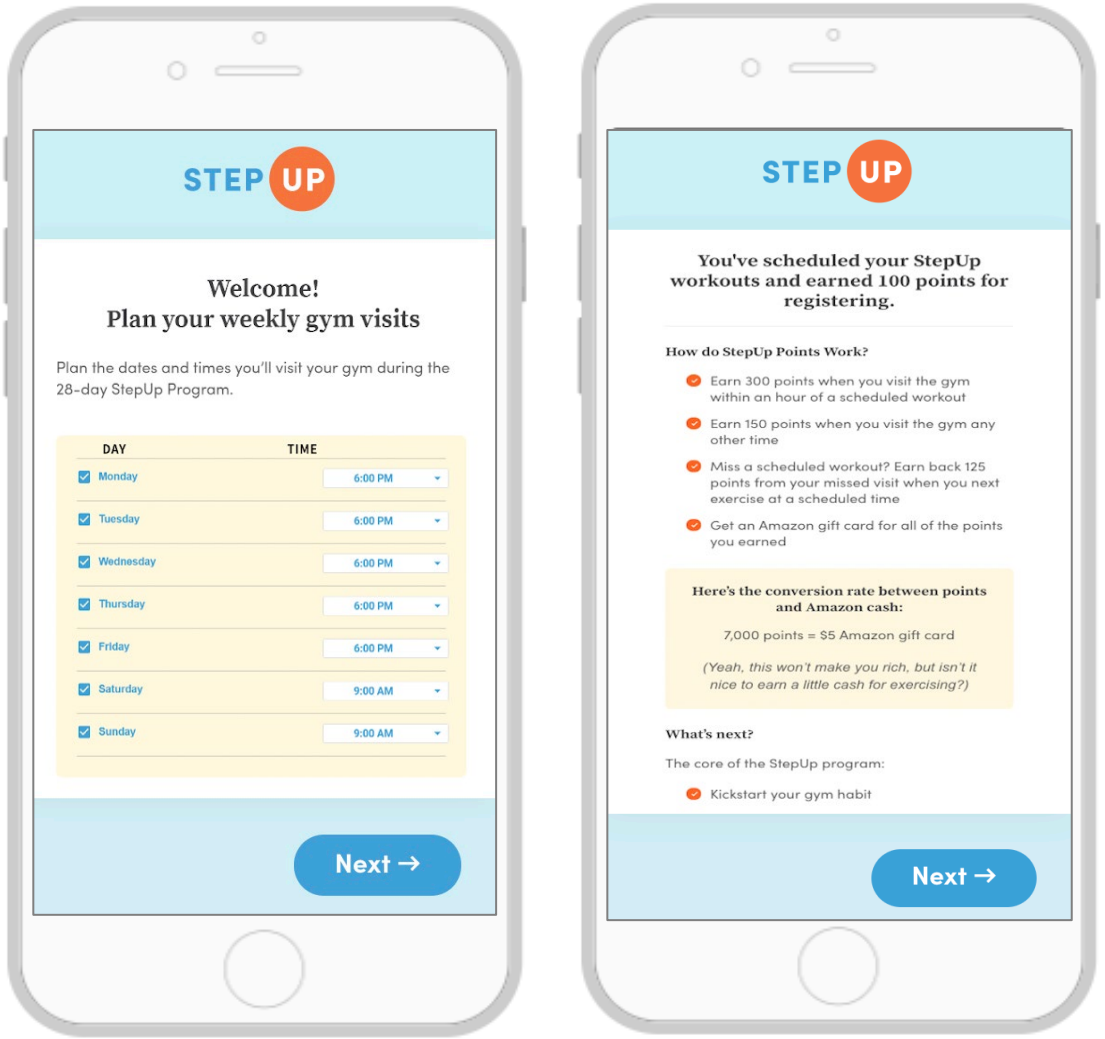





0.40
placebo control
(36% of treatments outperform placebo control at BH adjusted p < 0.05)
Significance
BH adj. p <.05
p < .05
0.0
0.5
whiskers depict +/-95% CIs
Increase in avg. weekly gym visits (regression-estimated)

Intervention That Worked Best:

OFFERING A BONUS TO RETURN



Participants earned an extra $0.09 if they came back to the gym the day after missing a workout

Increases gym visits by

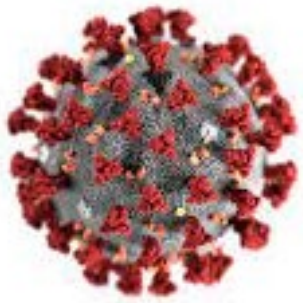
27%

during the 28-day StepUp Program

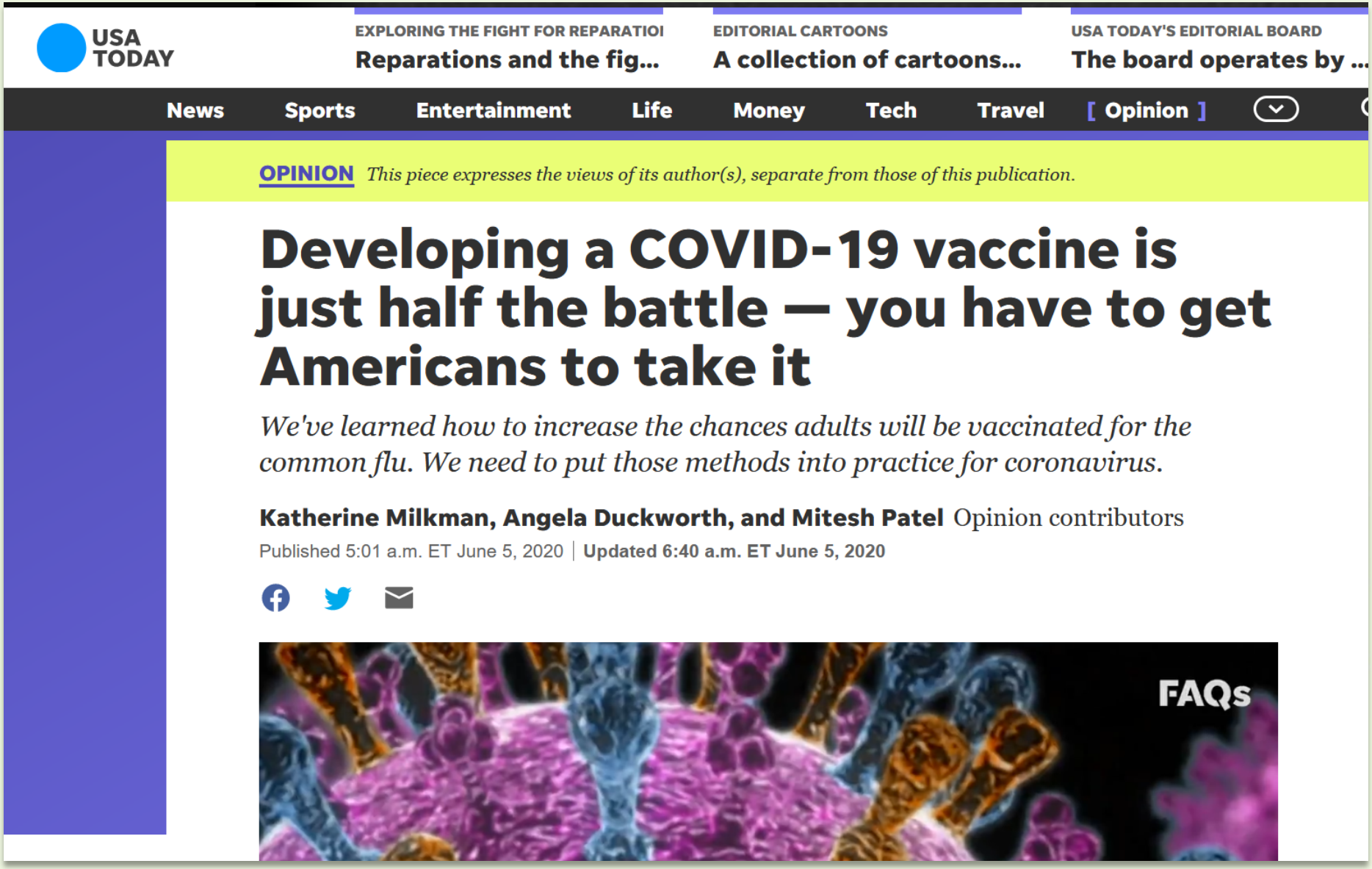
TAKEAWAYS FROM OUR FIRST MEGATUDY

- ✓ Identified behavioral insights that **could reliably increase gym attendance by up to 25% at low cost**
- ✓ **Forecasters aren't well-calibrated** at predicting what will work (no correlation between predictions of public health experts and results) so **testing is important**
- ✓ **Megastudies are a useful tool** to deploy when there is an urgent need for robust behavioral insights

WHERE DID THIS NEW TOOL FOR TESTING MANY IDEAS AT ONCE PROVE PARTICULARLY USEFUL?



INTEREST IN COVID-19 VACCINES BEFORE THEY WERE AVAILABLE IN THE U.S.



A TOUR OF FOUR MEGASTUDIES

1. Exercise at 24 Hour Fitness
2. **Flu Vaccinations at Healthy Checkups**
3. **Flu Vaccinations at the Pharmacy**
4. COVID Boosters at the Pharmacy

A MEGASTUDY WITH *Penn Medicine & Geisinger*

(Including 47,306 patients with healthy check ups)

Experimental Design

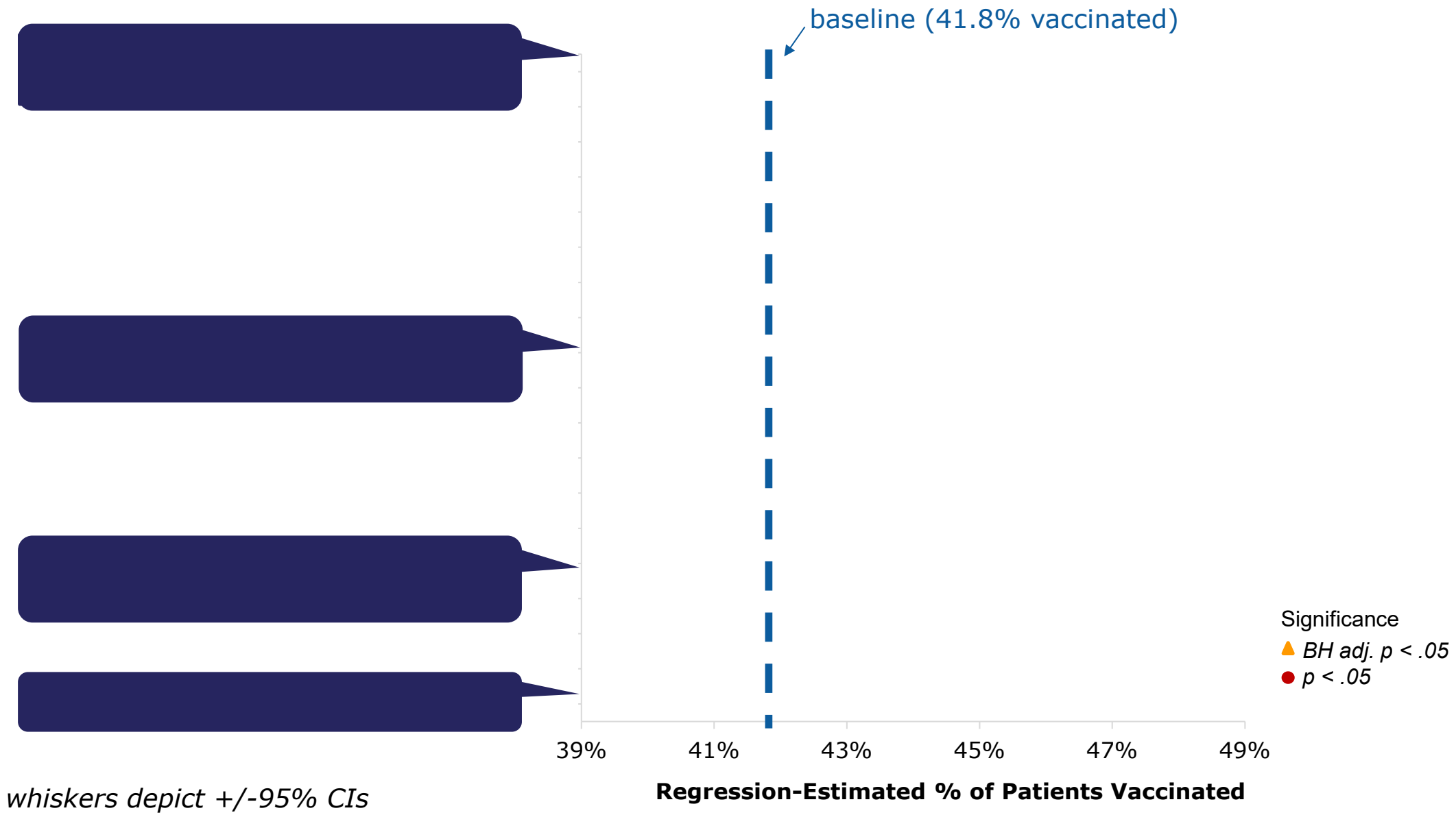
We tested 19 different text-messaging strategies

A Sample of What We Tested

- *"Dedicate your shot to a loved one"*
- *"Here's a joke about the flu"*
- *"Get a shot to protect other people"*
- *"A shot has been reserved for you"*

Milkman et al. (2021). A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *PNAS.*

DID WE INCREASE VACCINATION RATES?



A TOUR OF FOUR MEGASTUDIES

1. Exercise at 24 Hour Fitness
2. Flu Vaccinations at Healthy Checkups
3. **Flu Vaccinations at the Pharmacy**
4. COVID Boosters at the Pharmacy

A MEGASTUDY WITH *Walmart Pharmacies*

(Including 689,693 Walmart Pharmacy patients)

Experimental Design

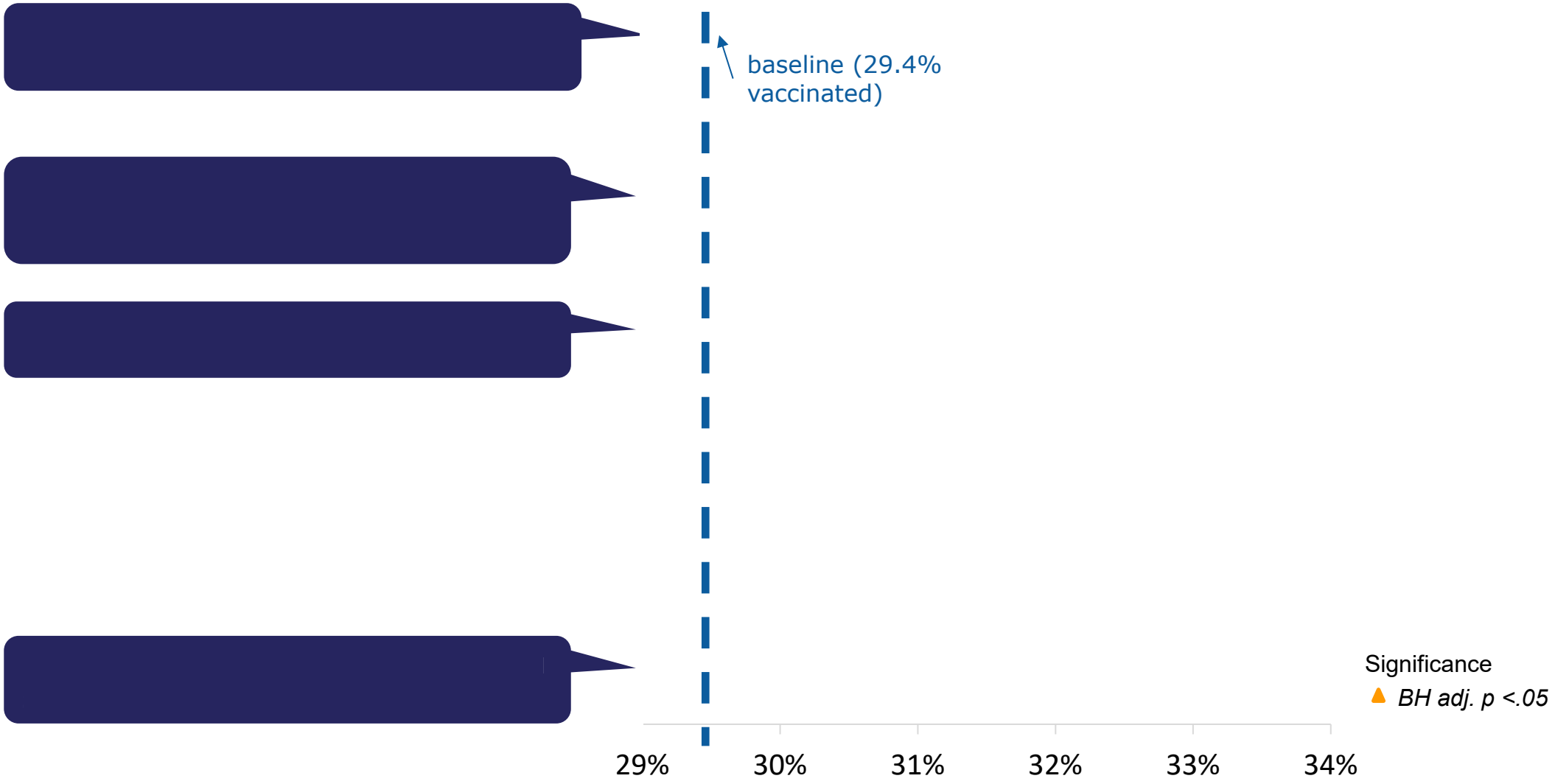
We tested 22 different text messaging strategies

A Sample of What We Tested

- *"Commit to getting a flu shot"*
- *"Get a shot to protect family & friends"*
- *"A shot is waiting for you"*

Milkman et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *PNAS.*

DID WE INCREASE VACCINATION RATES?



whiskers depict +/-95% CIs

TAKEAWAYS FROM FLU VACCINE MEGATUDIES

- ✓ **Ownership language** reliably improved the performance of vaccine reminders
- ✓ Multiple reminders outperformed single reminders **so nagging works**
- ✓ Informal, interactive texts were less effective
- ✓ **Experts could not predict ex ante what would work best** suggesting need for tests
- ✓ If effectively **scaled**, ~4-8% increases in vaccination could mean **~6-12 million extra Americans vaccinated**

A TOUR OF FOUR MEGASTUDIES

1. Exercise at 24 Hour Fitness
2. Flu Vaccinations at Healthy Checkups
3. Flu Vaccinations at the Pharmacy
4. **COVID Boosters at the Pharmacy** 

24 HOUR FITNESS

Penn Medicine

Geisinger

Walmart Pharmacy

CVS pharmacy

A MEGASTUDY TESTING HOW FREE RIDES & TEXT REMINDERS AFFECT COVID-19 *Vaccination Decisions*

(Including 3,662,548 pharmacy patients)

Experimental Design

We tested 8 different text messaging strategies

A Sample of What We Tested

- *"A free ride to and from the pharmacy"*
- *"Significant current transmission in your county"*
- *"Would Tuesday 2 pm at 1 Main St work for you?"*

DID WE INCREASE BOOSTER RATES?

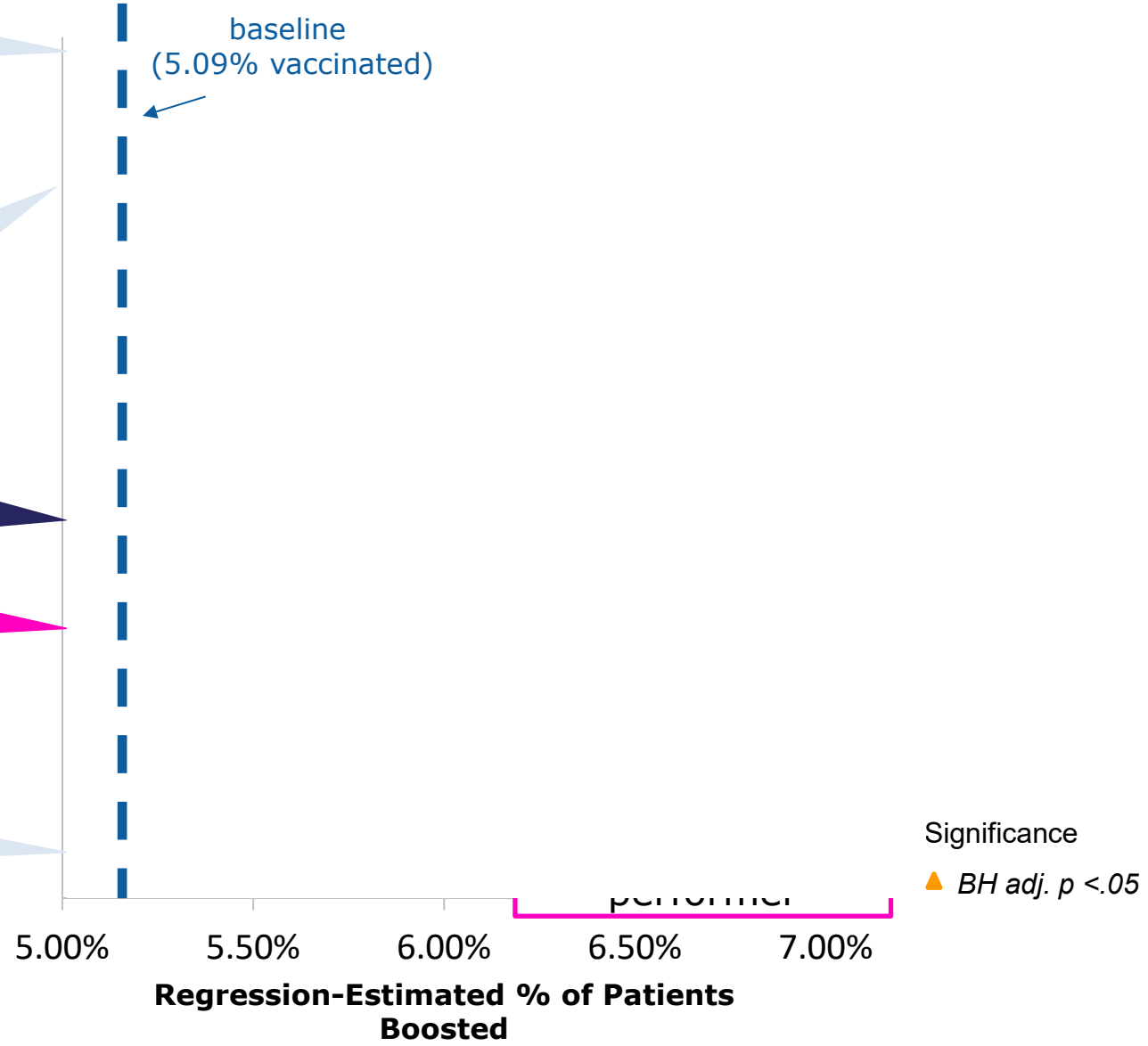
Many find it helps to make a plan. Would Tuesday at 2:00 PM at 1 Main Street work?

CDC data show significant current COVID transmission in Philadelphia County. Infection rates are in the top 30% in the US.

Updated COVID boosters are recommended to help prevent infection & severe illness. Your booster is waiting for you.

A free ride to and from the pharmacy has been reserved for your booster appointment.

Here are some important facts about why boosters are recommended: www.CDC.gov



whiskers depict +/-95% CIs

TAKEAWAYS FROM COVID BOOSTER MEGATUDY

- ✓ A behaviorally-informed reminder is as useful as a free ride, **contrary to expectations**
- ✓ Reminders had bigger effects for **men, Medicare beneficiaries,** those with **prior boosters,** those with **unknown insurance,** and **older adults**
- ✓ Reminders **spilled over** to boost **flu vaccinations**
- ✓ We estimate that this test produced **an extra 32,848 COVID-19 boosters** and **9,723 flu vaccinations** in fall 2022

IN THE WAKE OF THESE SUCCESSES, MEGASTUDIES ARE EXPLODING

1. Nudging elementary school teachers to support their students' math progress (Duckworth et al., 2024)
2. Reducing discrimination (Roy et al., 2024)
3. Increasing work for environmental protection (Vlasceanu et al., 2024)
4. Reducing support for partisan violence (Voelkel et al., 2024)
5. Increasing support for addressing climate change (Voelkel et al., 2024)
6. Nudging banking customers to save more (Milkman et al., 2024)

Also in Progress: happiness (Szaszi et al.), dishonesty (Uhlmann et al.), student success (Milkman et al.), plastic waste (Huang et al.)

DIFFERENT RESEARCHERS ARE MAKING DIFFERENT CHOICES ABOUT MEGASTUDY PARAMETERS

1. Is there an open call for intervention proposals or a call to a fixed group of collaborators?
2. Are interventions selected by peer review or are all ideas fielded?
3. How are the interventions delivered at scale? (e.g., email, text, survey)
4. Is the megastudy a collection of sub-studies or of one-off interventions?
5. Are the outcomes measured subjective or objective? One-off or longitudinal?
6. How many interventions are included? (10? 54?)

CONCLUSIONS

- ✓ **Megastudies can:**
  - ✓ Accelerate scientific discovery
  - ✓ Compare **apples to apples**
  - ✓ Reduce risk and fixed costs
  - ✓ Facilitate **interdisciplinary** collaboration
  - ✓ Lead to policy recommendations that are **better than what we'd generate based on scientific intuition**

- ✓ **Megastudies have limitations:**
  - ✓ Extra difficult to implement **complex, costly** designs
  - ✓ Require tremendously **large sample sizes**
  - ✓ Require **centralized coordination**
  - ✓ Require correction of p-values for multiple hypothesis testing (e.g., BH-correction)
  - ✓ Effects of top-performing interventions are overestimated in expectation (the winner's curse) so require adjustment (e.g., James Stein shrinkage)

THANK YOU

BCFG TEAM SCIENTISTS

BCFG FUNDERS

Abby and Jeremy Schiffman
AKO Foundation
Bill and Melinda Gates Foundation
Center for Health Incentives & Behavioral Economics
Chan Zuckerberg Initiative
Character Lab
Flu Lab
John Alexander, Alexander Production Company
Marc J. Leder, Sun Capital Partners
Michael & Susan Dell Foundation
National Bureau of Economic Research
National Institute on Aging, National Institutes of Health
Penn Center for Precision Medicine Accelerator Fund
Robert Wood Johnson Foundation
Schmidt Futures
Social Science Research Council
Warren G. Lichtenstein, Steel Partners

BCFG STAFF (Past and Present)

Dena Gromet
Lauri Bonacorsi
Katherine Chang
Sean Ellis
Melanie Ernest
Heather Graci
Aden Halpern
Hung Ho
Daniel Jaro
Jay Jung
Joseph Kay
Alec Lamon
Annie Laughlin
Tim Lee
Alex Luscher
Rayyan Mobarak
Predrag Pandiloski
Yeji Park
Madeline Paxson
Aneesh Rai
Jake Rothschild
Caitlin Satterfield
Ankur Saxena
Saurin Shah
Donna St. Louis
Robert Wisner
Ramon Silvera Zumaran

BCFG PARTNERS

24 Hour Fitness
Walmart
Geisinger
Penn Medicine
Lyft
CVS Pharmacy

BEHAVIOR CHANGE FOR GOOD

Learn more at www.bcfg.upenn.edu