Digital Advertising Measurement

Florian Zettelmeyer
Kellogg School of Management
Northwestern University and NBER

Based on joint work with Brett Gordon (Northwestern), Neha Bhargava (Facebook), and Dan Chapsky (Facebook)

FTC Microeconomics Conference 2016

Northwestern | Kellogg School of Management

Advertising effectiveness measurement is an age-old problem

JOHN WANAMAKER (1838-1922)

“Half the money I spend on advertising is wasted; the trouble is, I don’t know which half.”

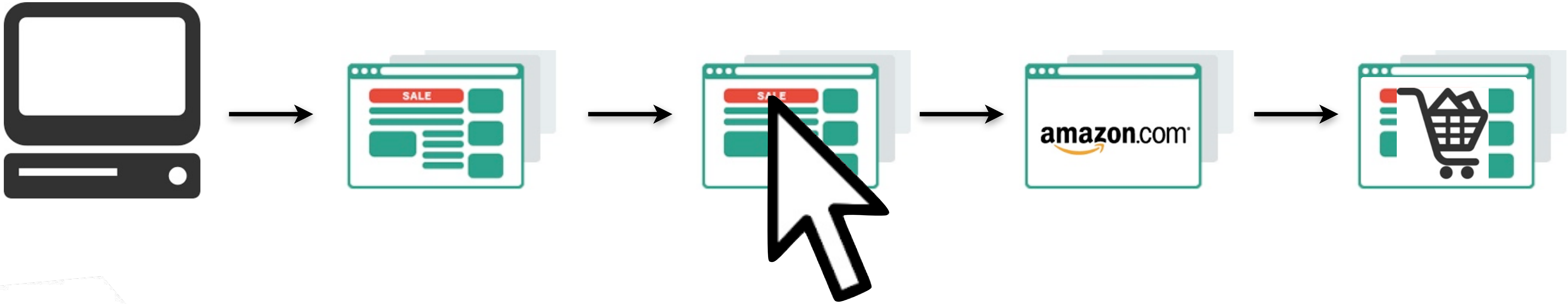
Conventional wisdom: Problem is the inability to track ad exposure and purchase outcomes at the individual level

TRADITIONAL VIEW OF AD MEASUREMENT PROBLEM

- **We did not know who saw an advertisement**
  - (At best) we knew how many consumer saw an ad
- **We did not know who purchased**
  - We know only how many products were purchased

Digital media was supposed to make measurement easier

Digital media was supposed to make measurement easier



Industry insiders have suggested that digital tracking largely solves the measurement problem

"Measuring the online sales impact of an online ad campaign... is **straightforward**: We determine who has viewed the ad, then compare online purchases made by those who have and those who have not seen it."

-Founder and Former CEO of comScore

Source: https://hbr.org/2008/04/the-off-line-impact-of-online-ads

Understanding Behavioural Impact Of Ad Exposure: comScore's Methodology

AD EXPOSED GROUP

Test and control groups matched on demographic and behavioural variables

BALANCED UNEXPOSED GROUP

LIFT METRICS

- Site Visitation
- Site Engagement
- Search Behaviour
- Buying Behavior

In practice, many firms avoid running advertising experiments

REASONS

- **Technical limitations** of advertising platforms
- Viewed as **expensive**
  - Waste of advertising opportunities
  - PSAs are used as "control ads"
- **Viewed as unnecessary** in light of observational methods

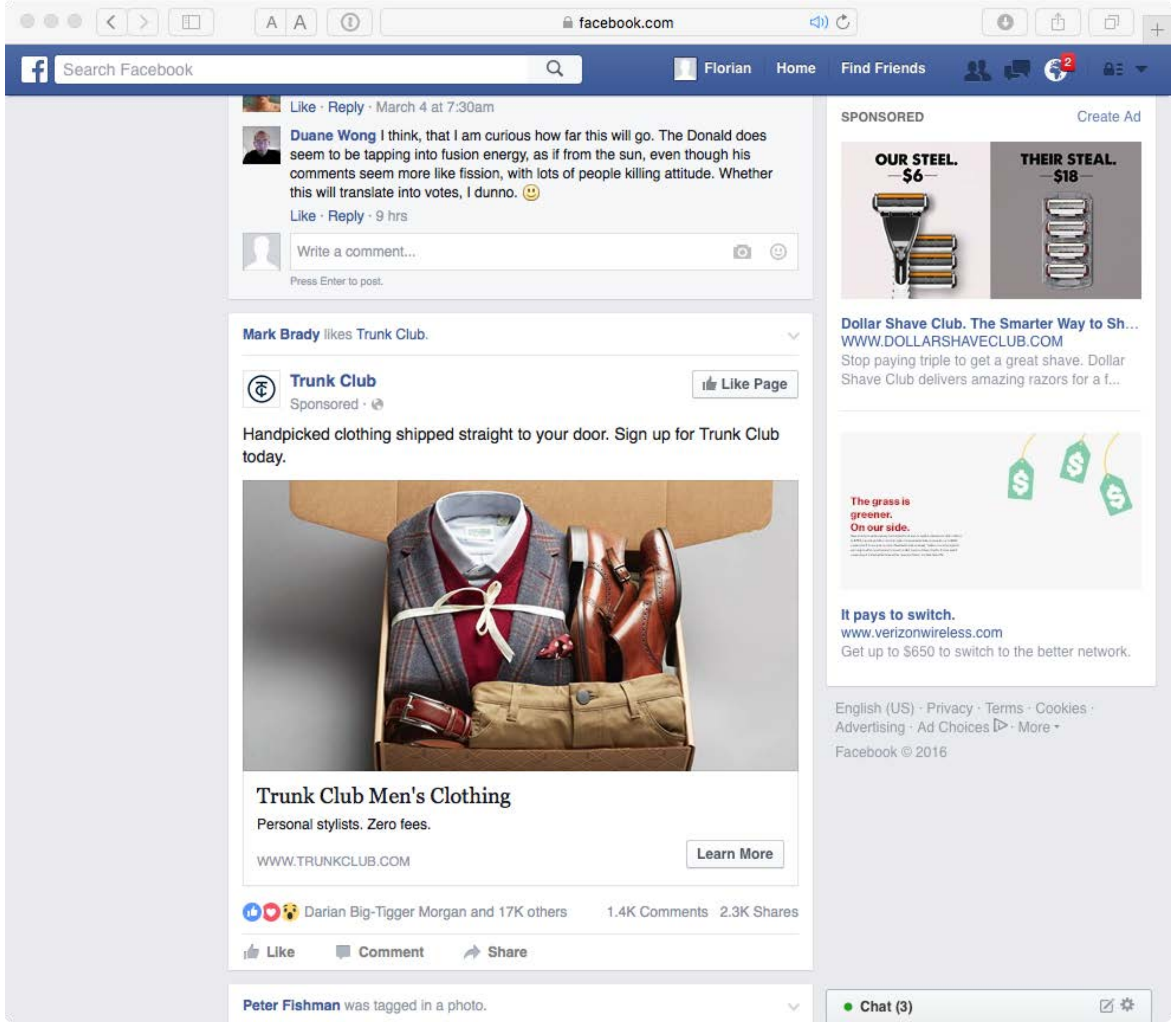
MY GOAL TODAY

*Characterize the degree to which **observational methods** can **substitute** for **randomized experiments** in online advertising measurement*

Source: Gordon, Zettelmeyer, Bhargava, Chapsky (2016): "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," Kellogg School of Management, Northwestern University
No data contained PII that could identify consumers or advertisers to maintain privacy. Based upon data from 15 US advertising lift studies. The studies were not chosen to be representative of all Facebook advertising.

Facebook advertising show up in the newsfeed or to the right of the page

TRUNK CLUB EXAMPLE



Facebook recently built an experimentation platform

FEATURES OF OUR DATA

Facebook recently built an experimentation platform

FEATURES OF OUR DATA

- **15 large-scale randomized advertising experiments** across verticals

Facebook recently built an experimentation platform

FEATURES OF OUR DATA

- **15 large-scale randomized advertising experiments** across verticals
- **Statistical power**
  - Between 2 million and 150 million users per experiment
  - 492 million user-study observations
  - 1.5 billion total ad impressions

Facebook recently built an experimentation platform

FEATURES OF OUR DATA

- **15 large-scale randomized advertising experiments** across verticals
- **Statistical power**
  - Between 2 million and 150 million users per experiment
  - 492 million user-study observations
  - 1.5 billion total ad impressions
- **Single-user login**
  - Eliminates issues with cookie-based measurement
  - Captures cross-device activity

Facebook recently built an experimentation platform

FEATURES OF OUR DATA

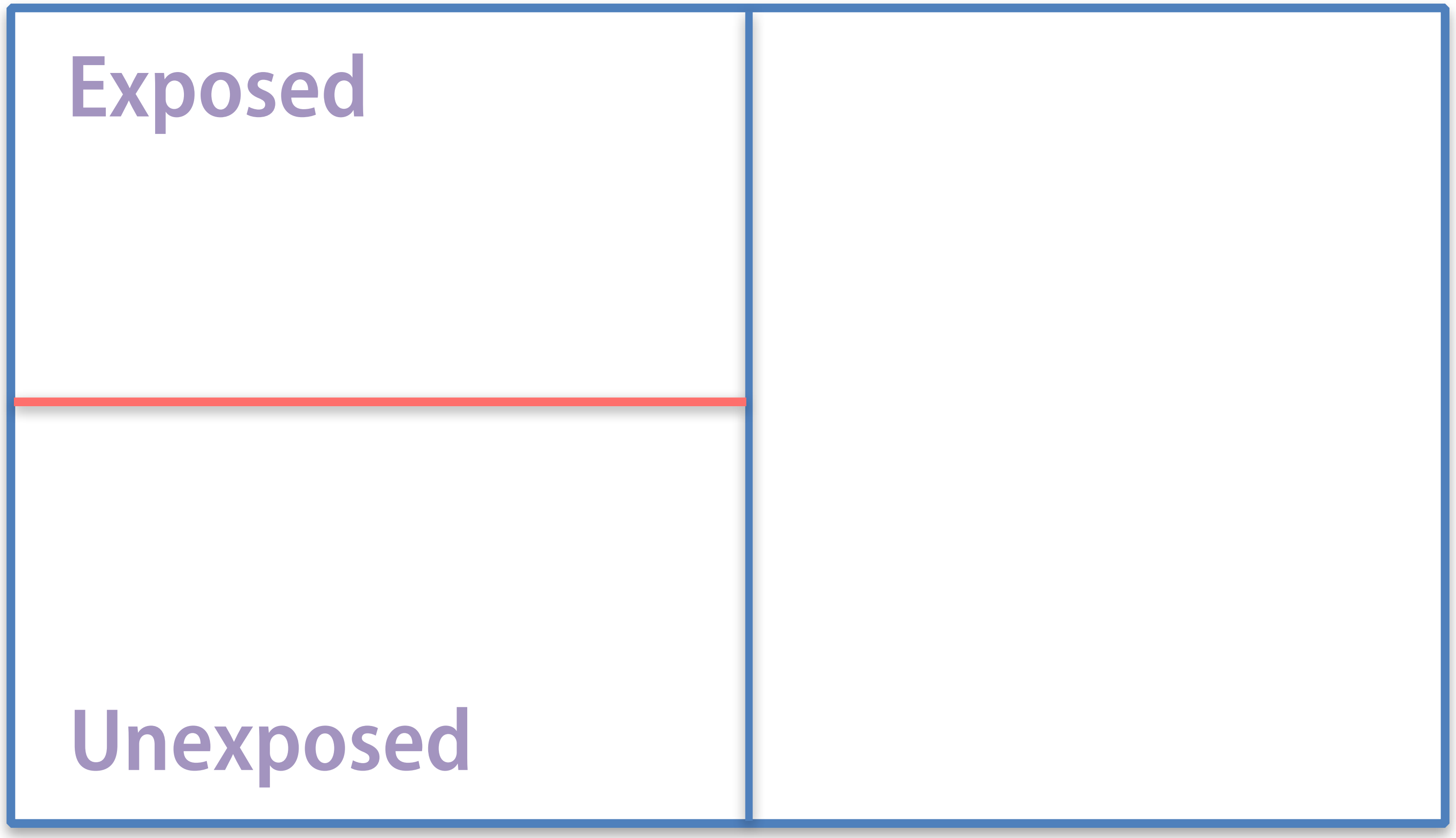
- **15 large-scale randomized advertising experiments** across verticals
- **Statistical power**
  - Between 2 million and 150 million users per experiment
  - 492 million user-study observations
  - 1.5 billion total ad impressions
- **Single-user login**
  - Eliminates issues with cookie-based measurement
  - Captures cross-device activity
- **Measure outcomes** (e.g., purchases, registrations) directly via conversion pixels on advertisers' websites—no ad clicks required

CONTENTS

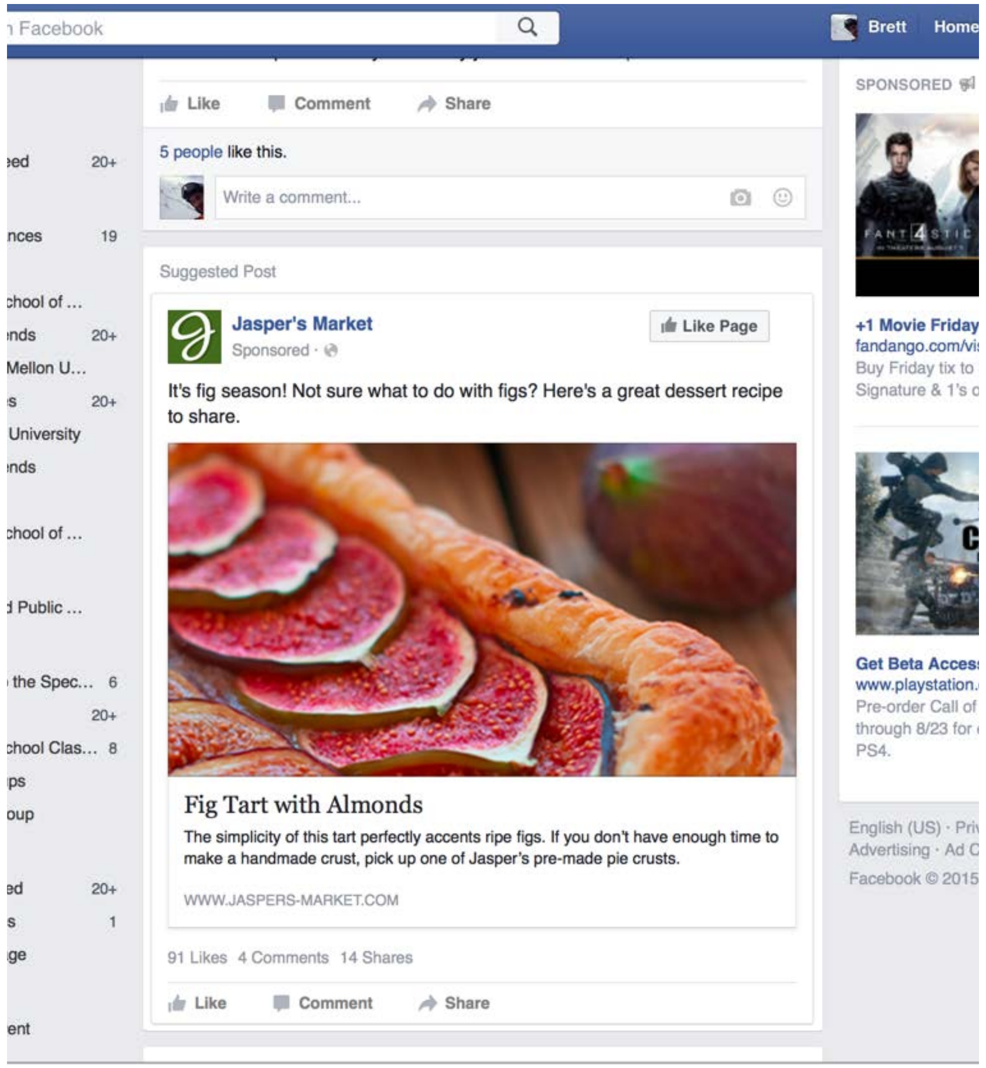
Randomized experiment with one-sided noncompliance

Test
(Eligible to be exposed)

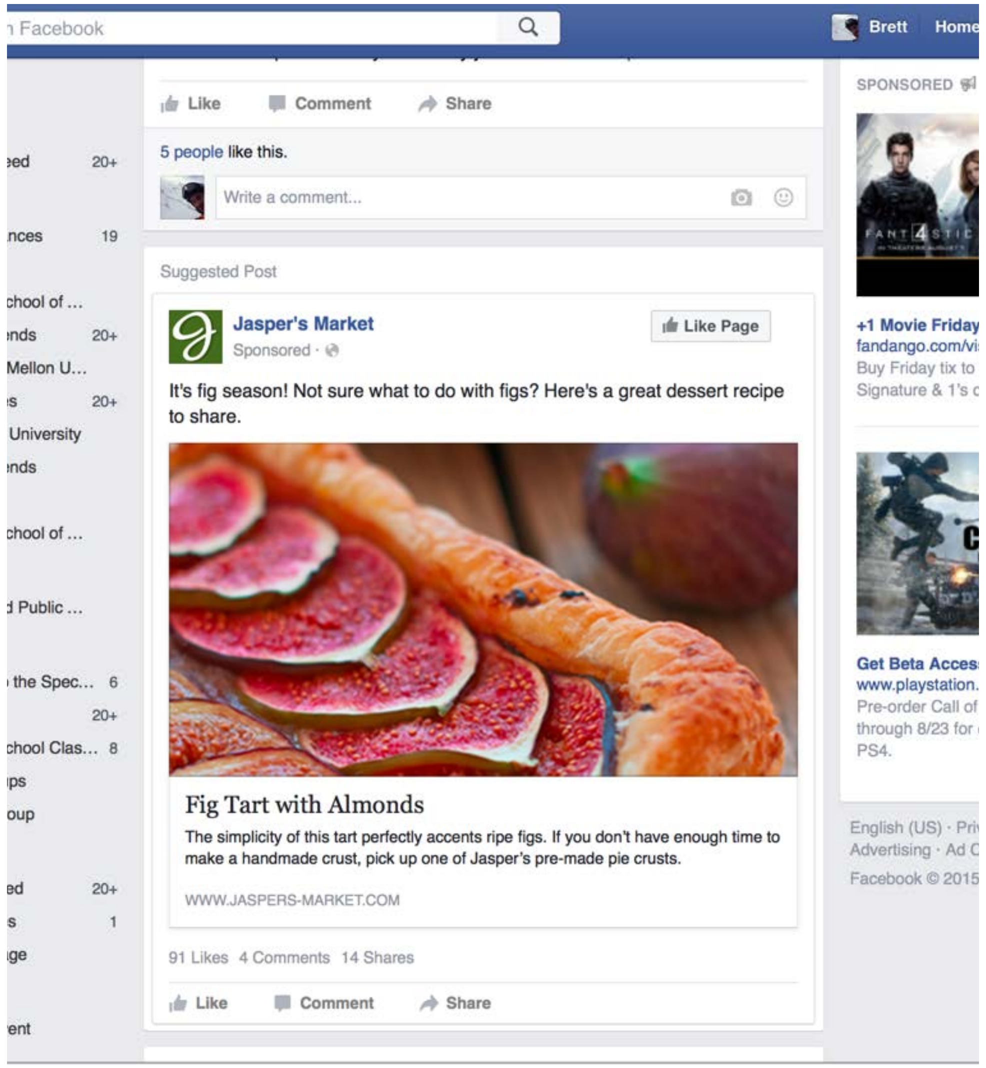
Control
(Unexposed)



Imagine two identical users are randomly assigned to test and control groups for Jasper's Market

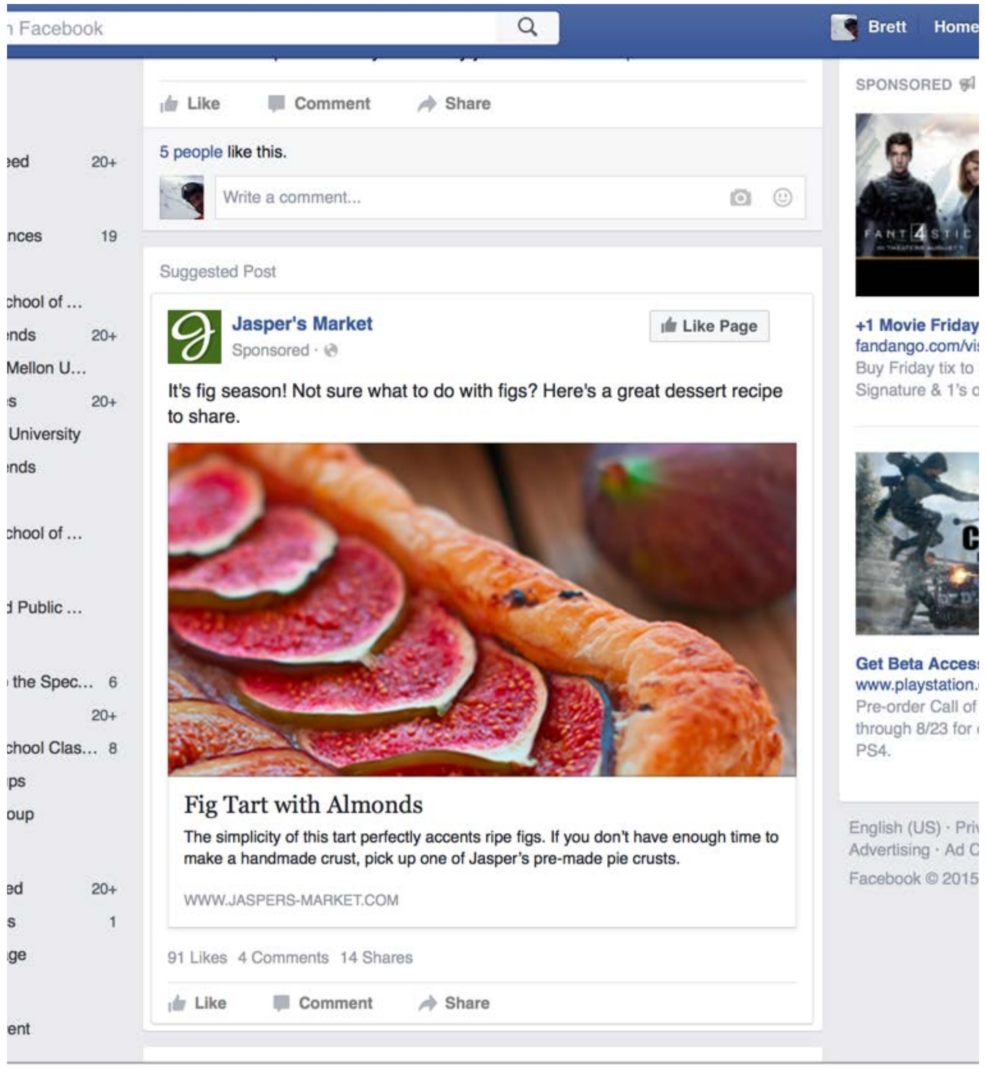


Test

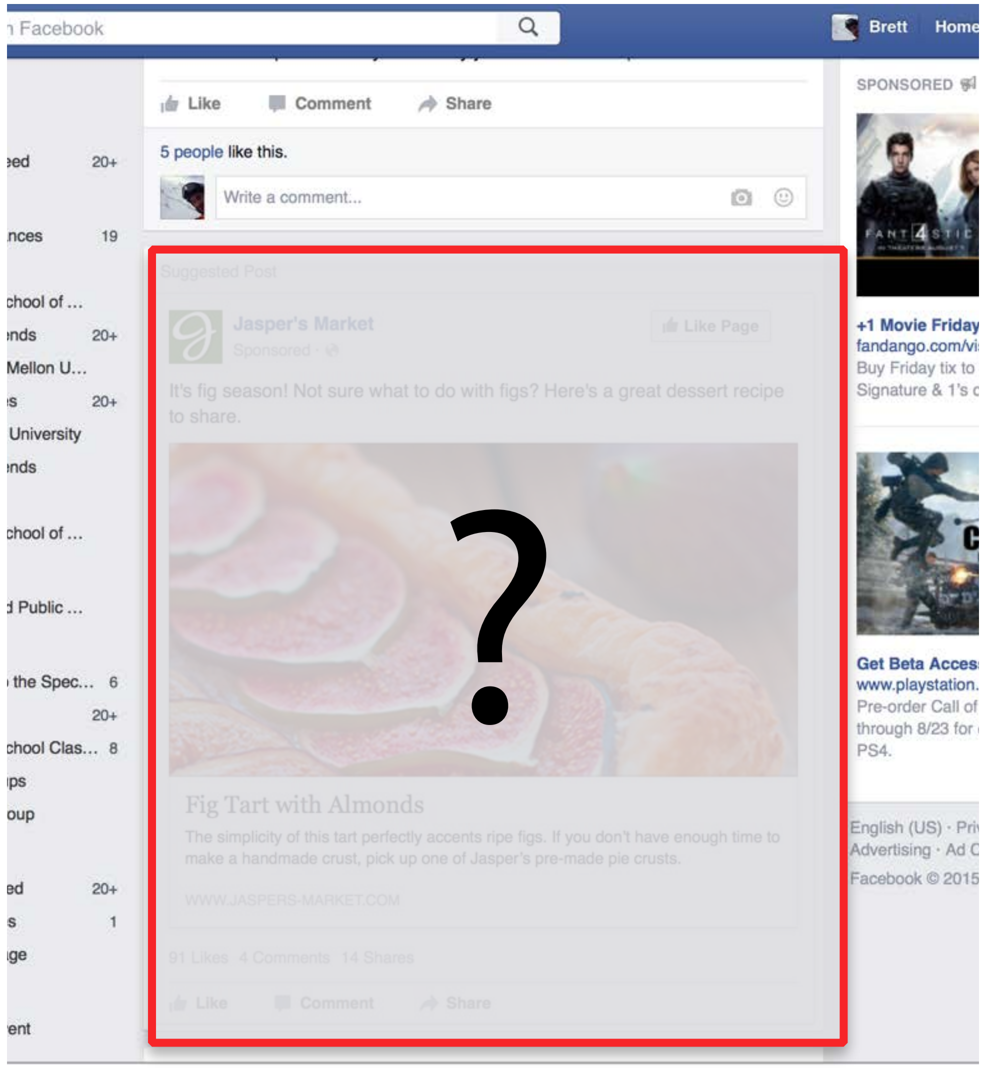


Control

What ad should the control user see?



Test

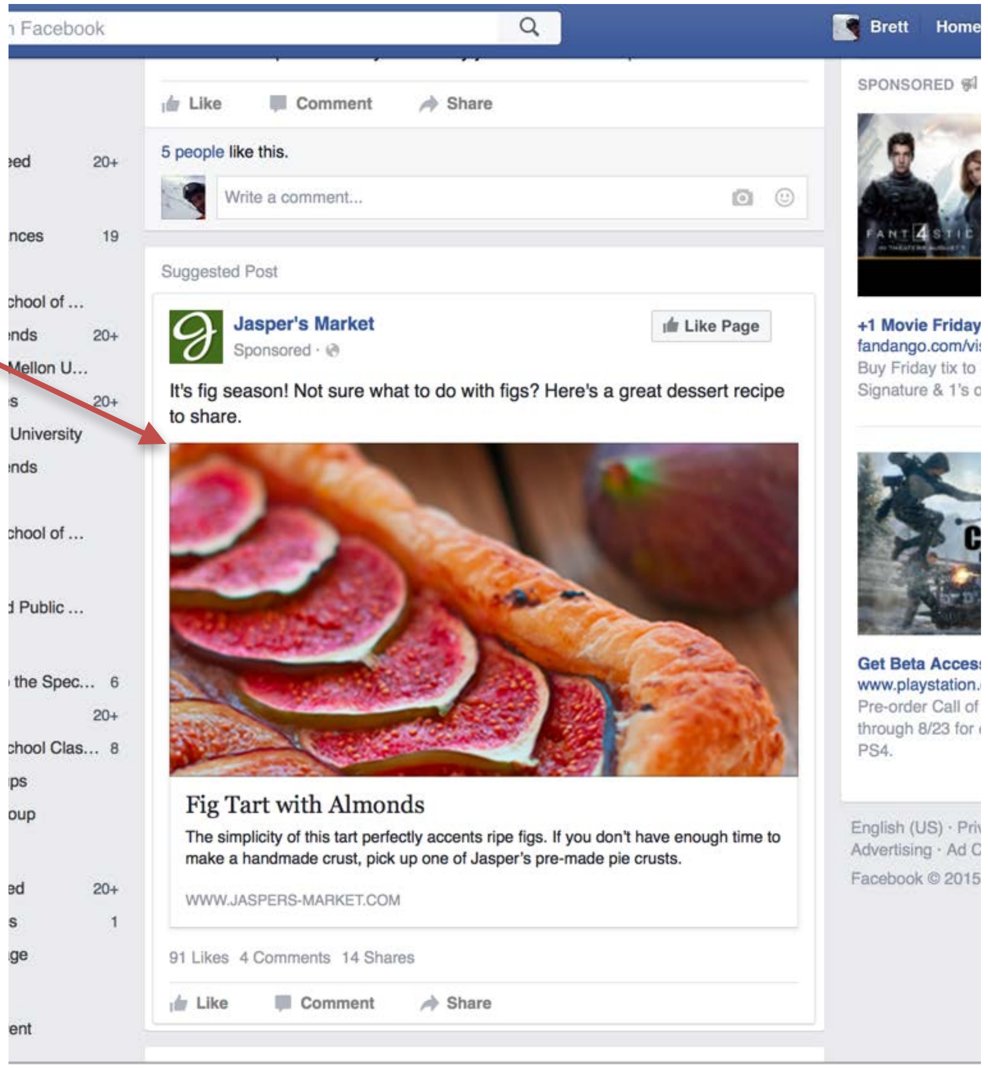


Control

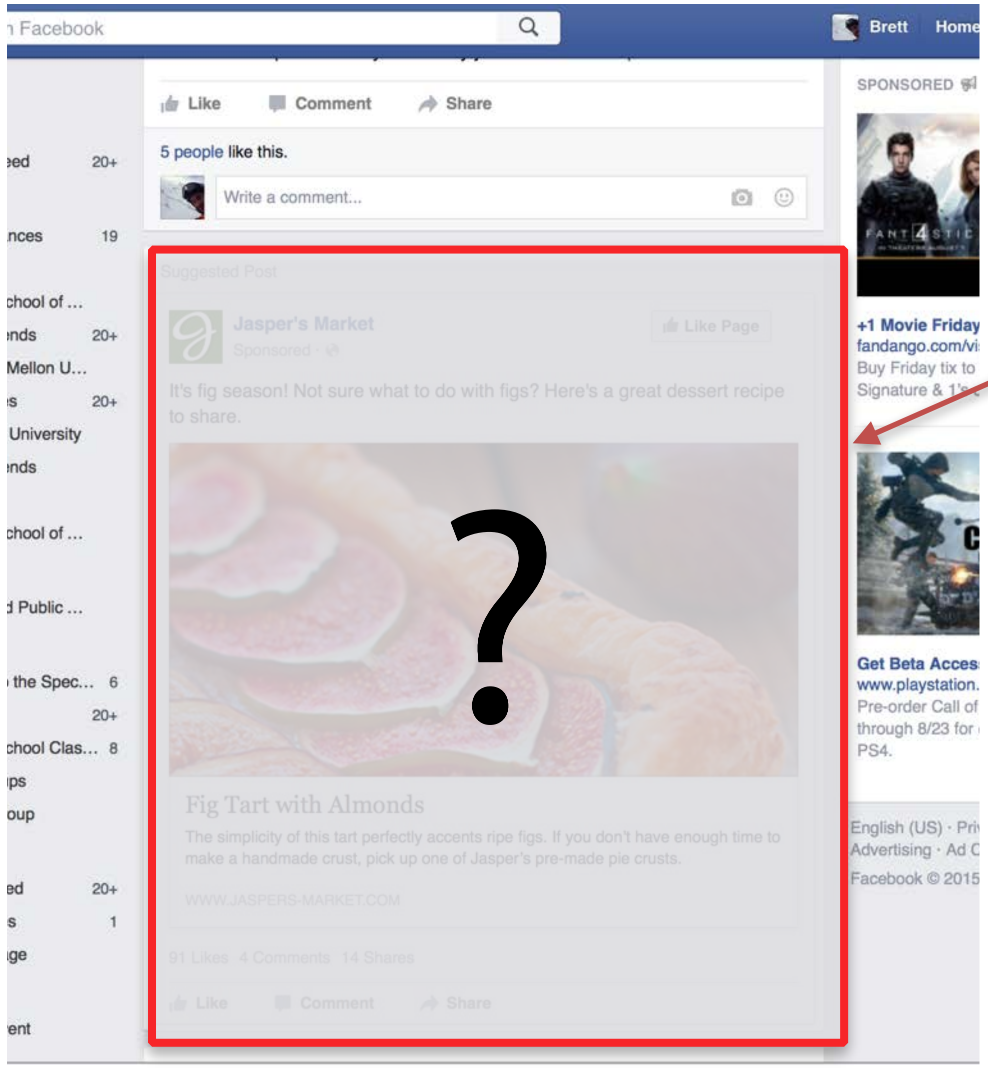
Serve the ad that would have been shown in the absence of the Jasper's Market ad campaign

Ad Auction

1.
2.
3.
4.



Test



Control

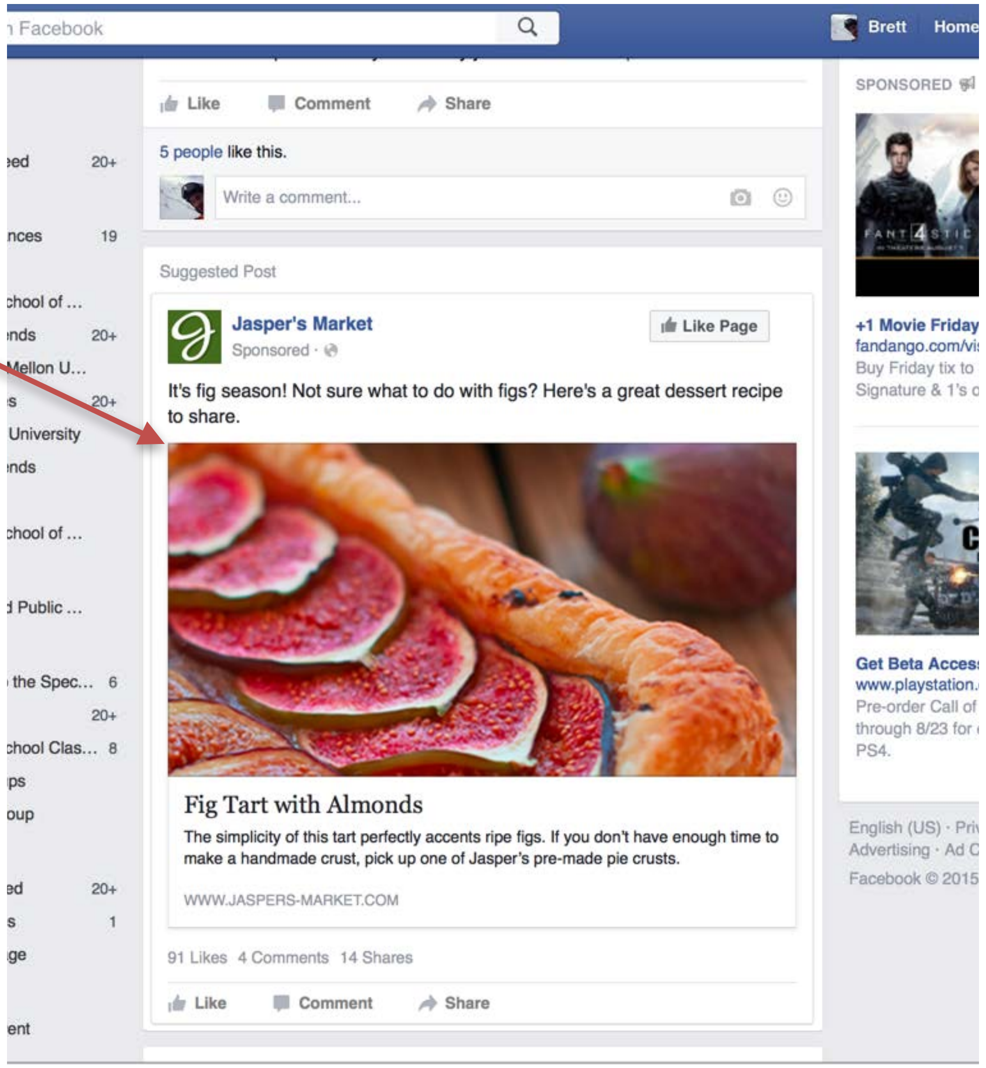
Ad Auction

1.
2.
3.
4.

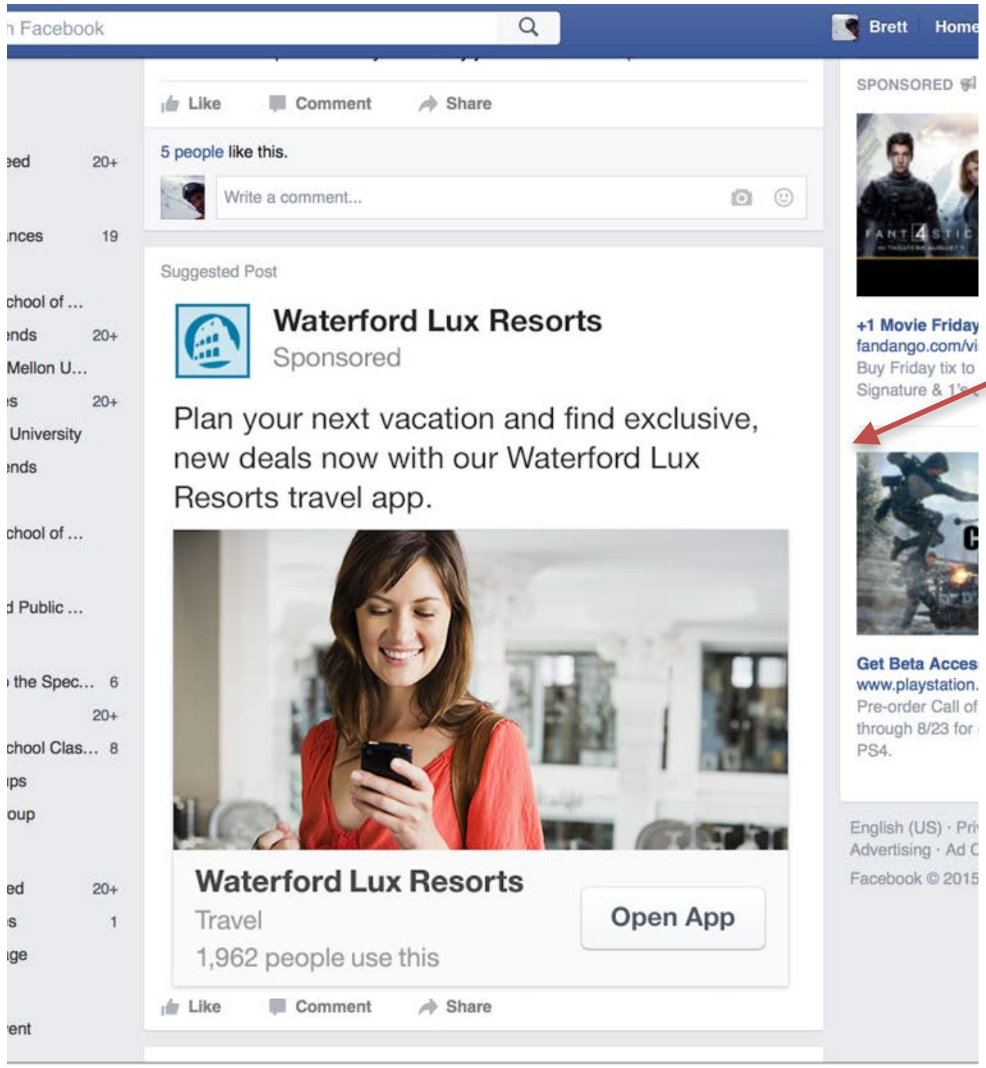
Serve the ad that would have been shown in the absence of the Jasper's Market ad campaign

<u>Ad Auction</u>

1.
2.
3.
4.



Test



Control

<u>Ad Auction</u>

1.
2.
3.
4.

This mechanism produces a *distribution* of control ads

KEY IMPLICATION

- The focal ad might be “replaced” by a different control ad for each exposure
  - Sometimes Gap wins
  - Sometimes Audi wins
  - etc…

This is the distribution of control ads a user would have seen, had the focal advertiser’s campaign *never existed*

CONTENTS

We illustrate the RCT estimates using one of the studies

STUDY #4: Omni-channel retailer

- **Sample size:** 25.5 million users over two weeks in 2015
  - 30% Control / 70% Test

- **Treatment:** exposed vs. unexposed (binary)

- **Outcome**: purchase at the digital retailer via "conversion pixel," which the advertiser placed after the checkout page (binary)

Results: ATT Lift

Average Treatment Effect on the Treated (ATT)

- Intent-to-Treat (ITT) effect = 0.012%
- 25% of users exposed in the test group
- **ATT = 0.012%/0.25 = 0.045%**

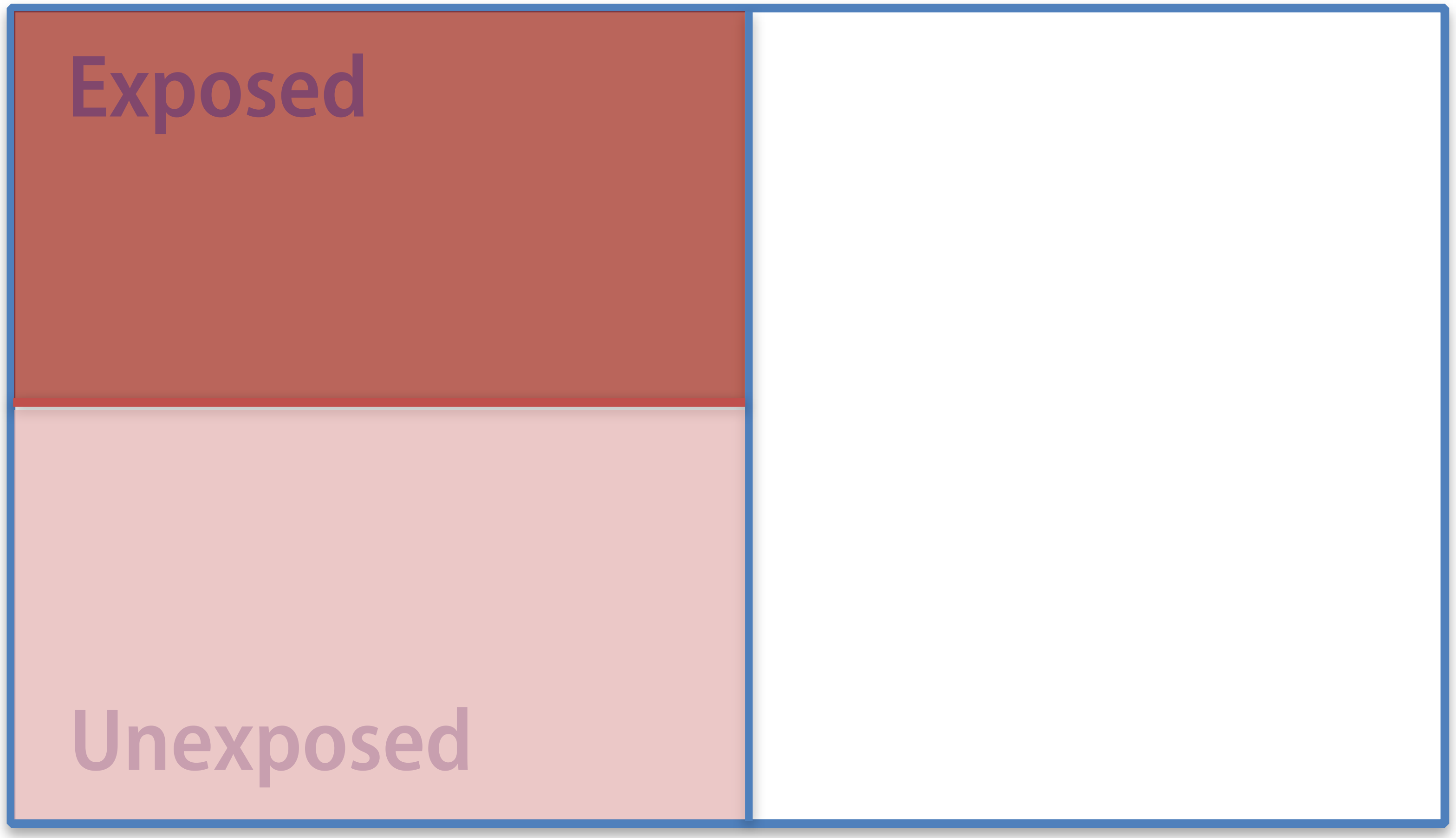
ATT Lift

- Conversion rate of treated (exposed) users: 0.107%
- Conversion rate if treated had not been treated: 0.107% - 0.045% = 0.062%
- **Lift = 0.045%/0.062% = 73%   95% CI = [33, 113]**

In practice, many firms don't have a control group

Test
(Eligible to be exposed)

Control
(Unexposed)



Exposed vs. unexposed yields very different estimates

EXPOSED-UNEXPOSED COMPARISON

- Unexposed (in test) : 0.020% conversion rate
- Exposed (in test): 0.107% conversion rate

Lift = 416%
CI = [308, 524]

Significantly overstates RCT lift of 73%

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

The problem is that, within the test group, unexposed and exposed users differ

	Control	Test	
		Unexposed	Exposed
age	31.67	32.07	30.45
gender	1.17	1.22	1.05
facebookage	2288	2295	2264
married	0.20	0.19	0.21
single	0.14	0.14	0.14
friend_count	486	462	554
web_l7	1.64	1.81	1.15
mobile_l7	5.99	5.77	6.63
primary_phone_os_2	0.47	0.47	0.45
primary_phone_os_1	0.43	0.40	0.51
primary_phone_os_0	0.08	0.10	0.03

Core question: How well can we do without an experiment?

Since our goal is to mimic an observational data set, we ***only use data from the test group***

Observational Methods

- Exact Matching (EM)
  - Age and gender
- Propensity Score Matching (PSM)
  - Logit propensity, 4 nearest neighbors
- Regression Adjustment (RA)
  - Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRAT)

Unconfoundedness Assumption

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp W_i \;\mid\; X_i$$

Observational Methods

- Exact Matching (EM)
  - Age and gender
- Propensity Score Matching (PSM)
  - Logit propensity, 4 nearest neighbors
- Regression Adjustment (RA)
  - Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRAT)

Group exposed/unexposed users into age-gender strata

Remove observations without overlap across exposure status

Reweigh unexposed observations to equalize age-gender distribution

Observational Methods

- Exact Matching (EM)
  - Age and gender
- Propensity Score Matching (PSM)
  - Logit propensity, 4 nearest neighbors
- Regression Adjustment (RA)
  - Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRAT)

Rosenbaum & Rubin (1983), Abadie & Imbens (2006)

Estimate propensity scores Pr(W | X)

Match each exposed user to the four unexposed users with the closest propensity scores

Observational Methods

- Exact Matching (EM)
  - Age and gender
- Propensity Score Matching (PSM)
  - Logit propensity, 4 nearest neighbors
- Regression Adjustment (RA)
  - Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRAT)

Robins & Rotnitzky (1995), Wooldridge (2007)

Regress outcomes on covariates separately for exposed/ unexposed

Weigh observations by the inverse propensity scores to achieve double robustness

Observational Methods

- Exact Matching (EM)
  - Age and gender
- Propensity Score Matching (PSM)
  - Logit propensity, 4 nearest neighbors
- Regression Adjustment (RA)
  - Inverse Probability-Weighed Regression Adjustment (IPWRA)
- Stratification & Regression (STRAT)

Rosenbaum & Rubin (1983), Imbens & Rubin (2015)

Partition the sample into strata by discretizing the propensity score (larger N —> more strata)

Regress outcome on exposure and covariates separately within each strata

Sequence of variables for the observational methods

EM: Age and gender

PSM, IPWRA, STRAT:

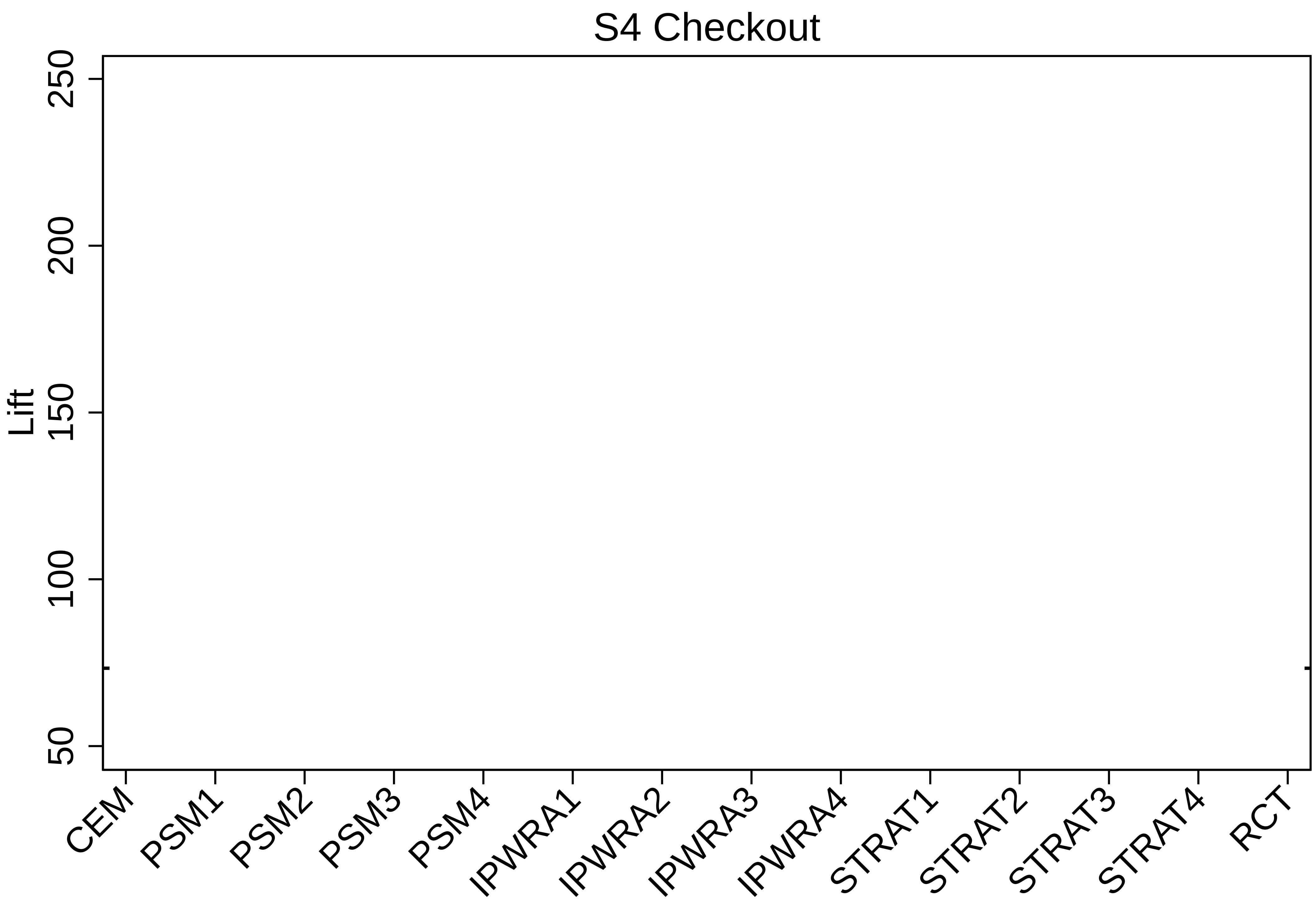
1. Age, gender, # days on FB, FB age, friends, initiated friends, relationship status, mobile OS, tablet OS, market fixed effects, day fixed effects, etc.

2. Same as 1 + Census/ACS data matched by zip code

3. Same as 2 + Facebook User Activity (binned)

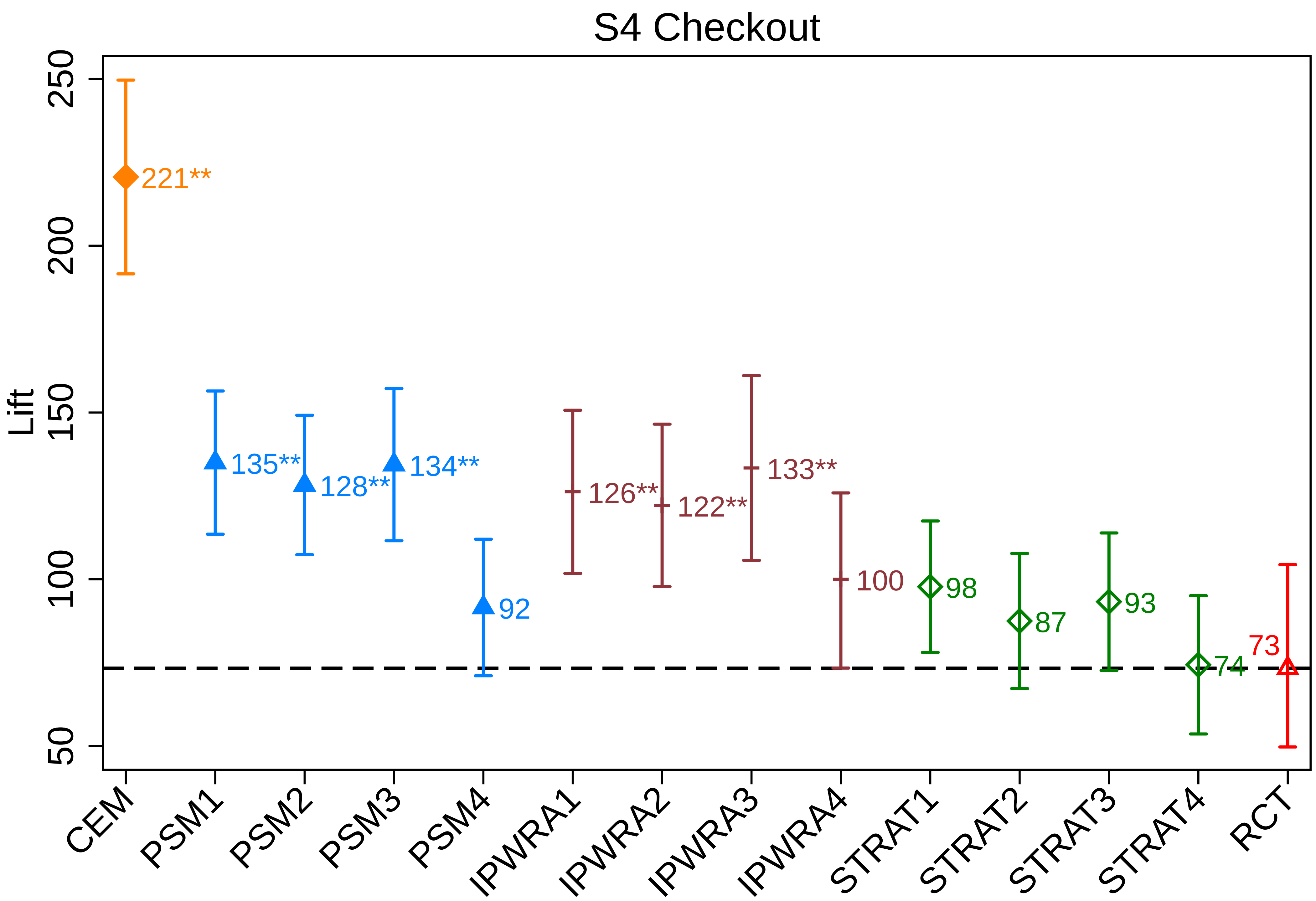
4. Same as 3 + Facebook Match Score

Exposed-unexposed
Lift = 416%



Benchmark (RCT)
Lift = 73%

Exposed-unexposed
Lift = 416%



Benchmark (RCT)
Lift = 73%

CONTENTS

We analyzed a total of 15 studies

STUDY SELECTION PROCEDURE

- **Brett and Florian** selected these studies using the following criteria:
  - Experiment conducted recently (Jan 2015 or later)
  - Minimal sample size (>1 million users)
  - Business-relevant conversion tracking in place
  - No retargeting by advertiser during experiment
- Our samples are **not representative** of all Facebook advertising

Note: Some numbers have been scaled to preserve confidentiality.

We observe a variety of studies

Study	Vertical	Observations	Test	Control	Impressions	Clicks	Conversions	Outcomes*
1	Retail	2,427,494	50%	50%	39,167,679	45,401	8,767	C, R
2	Finan. serv.	86,183,523	85%	15%	577,005,340	247,122	95,305	C, P
3	E-commerce	4,672,112	50%	50%	7,655,089	48,005	61,273	C
4	Retail	25,553,093	70%	30%	14,261,207	474,341	4,935	C
5	E-commerce	18,486,000	50%	50%	7,334,636	89,649	226,817	C, R, P
6	Telecom	141,254,650	75%	25%	590,377,329	5,914,424	867,033	P
7	Retail	67,398,350	17%	83%	61,248,021	139,471	127,976	C
8	E-commerce	8,333,319	50%	50%	2,250,984	204,688	4,102	C, R
9	E-commerce	71,068,955	75%	25%	35,197,874	222,050	113,531	C
10	Tech	1,955,375	60%	40%	2,943,890	22,390	7,625	C, R
11	E-commerce	13,339,044	50%	50%	11,633,187	106,534	225,241	C
12	Retail	5,566,367	50%	50%	10,070,742	54,423	215,227	C
13	E-commerce	3,716,015	77%	23%	2,121,967	22,305	7,518	C, R
14	E-commerce	86,766,019	80%	20%	36,814,315	471,501	15,722	C
15	Retail	9,753,847	50%	50%	8,750,270	19,365	76,177	C

* C = checkout, R = registration, P = page view

We observe a variety of studies

Study	Vertical	Observations	Test	Control	Impressions	Clicks	Conversions	Outcomes*
1	Retail	2,427,494	50%	50%	39,167,679	45,401	8,767	C, R
2	Finan. serv.	86,183,523	85%	15%	577,005,340	247,122	95,305	C, P
3	E-commerce	4,672,112	50%	50%	7,655,089	48,005	61,273	C
4	Retail	25,553,093	70%	30%	14,261,207	474,341	4,935	C
5	E-commerce	18,486,000	50%	50%	7,334,636	89,649	226,817	C, R, P
6	Telecom	141,254,650	75%	25%	590,377,329	5,914,424	867,033	P
7	Retail	67,398,350	17%	83%	61,248,021	139,471	127,976	C
8	E-commerce	8,333,319	50%	50%	2,250,984	204,688	4,102	C, R
9	E-commerce	71,068,955	75%	25%	35,197,874	222,050	113,531	C
10	Tech	1,955,375	60%	40%	2,943,890	22,390	7,625	C, R
11	E-commerce	13,339,044	50%	50%	11,633,187	106,534	225,241	C
12	Retail	5,566,367	50%	50%	10,070,742	54,423	215,227	C
13	E-commerce	3,716,015	77%	23%	2,121,967	22,305	7,518	C, R
14	E-commerce	86,766,019	80%	20%	36,814,315	471,501	15,722	C
15	Retail	9,753,847	50%	50%	8,750,270	19,365	76,177	C

* C = checkout, R = registration, P = page view

We observe a variety of studies

Study	Vertical	Observations	Test	Control	Impressions	Clicks	Conversions	Outcomes*
1	Retail	2,427,494	50%	50%	39,167,679	45,401	8,767	C, R
2	Finan. serv.	86,183,523	85%	15%	577,005,340	247,122	95,305	C, P
3	E-commerce	4,672,112	50%	50%	7,655,089	48,005	61,273	C
4	Retail	25,553,093	70%	30%	14,261,207	474,341	4,935	C
5	E-commerce	18,486,000	50%	50%	7,334,636	89,649	226,817	C, R, P
6	Telecom	141,254,650	75%	25%	590,377,329	5,914,424	867,033	P
7	Retail	67,398,350	17%	83%	61,248,021	139,471	127,976	C
8	E-commerce	8,333,319	50%	50%	2,250,984	204,688	4,102	C, R
9	E-commerce	71,068,955	75%	25%	35,197,874	222,050	113,531	C
10	Tech	1,955,375	60%	40%	2,943,890	22,390	7,625	C, R
11	E-commerce	13,339,044	50%	50%	11,633,187	106,534	225,241	C
12	Retail	5,566,367	50%	50%	10,070,742	54,423	215,227	C
13	E-commerce	3,716,015	77%	23%	2,121,967	22,305	7,518	C, R
14	E-commerce	86,766,019	80%	20%	36,814,315	471,501	15,722	C
15	Retail	9,753,847	50%	50%	8,750,270	19,365	76,177	C

* C = checkout, R = registration, P = page view

We observe a variety of studies

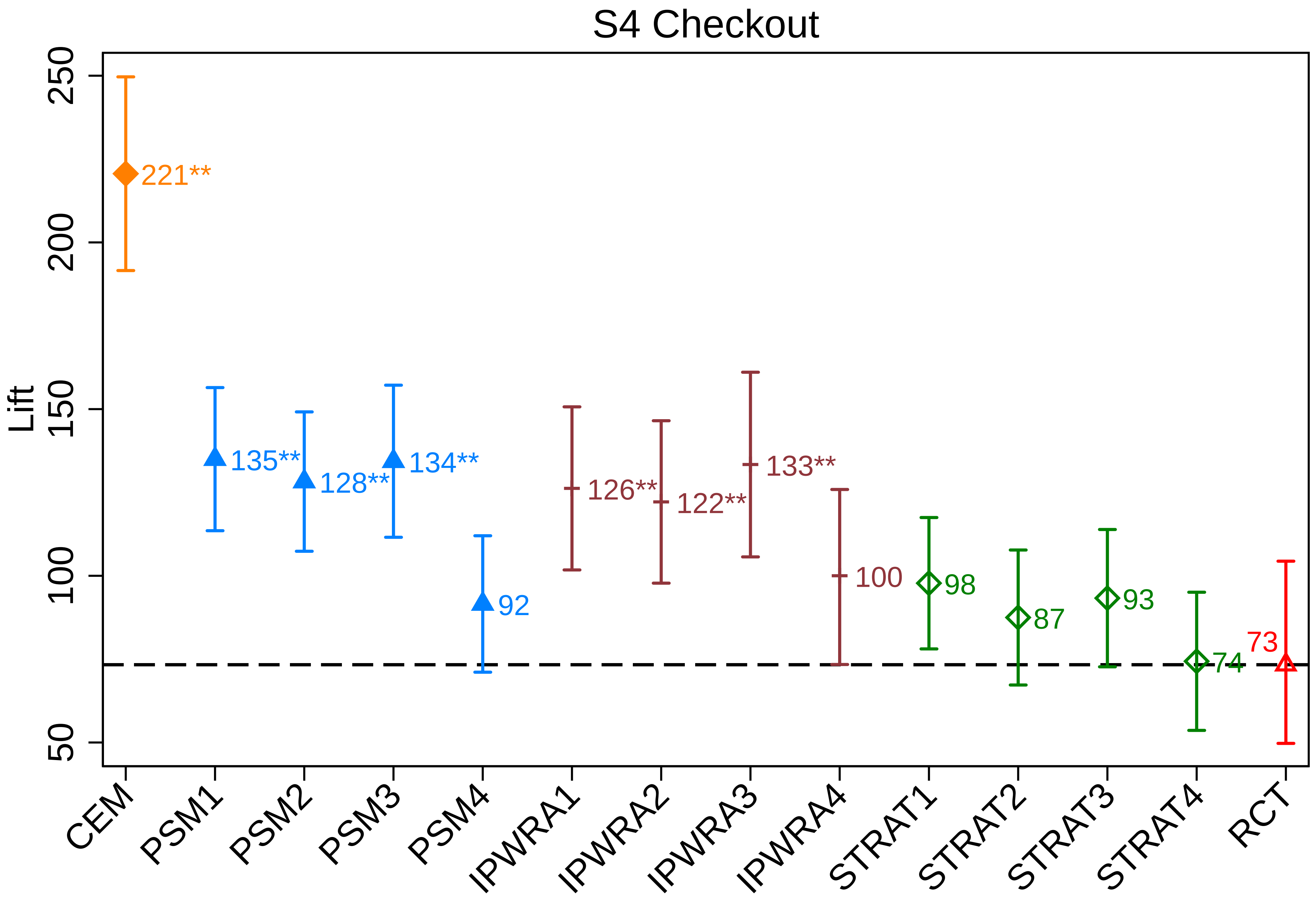
Study	Vertical	Observations	Test	Control	Impressions	Clicks	Conversions	Outcomes*
1	Retail	2,427,494	50%	50%	39,167,679	45,401	8,767	C, R
2	Finan. serv.	86,183,523	85%	15%	577,005,340	247,122	95,305	C, P
3	E-commerce	4,672,112	50%	50%	7,655,089	48,005	61,273	C
4	Retail	25,553,093	70%	30%	14,261,207	474,341	4,935	C
5	E-commerce	18,486,000	50%	50%	7,334,636	89,649	226,817	C, R, P
6	Telecom	141,254,650	75%	25%	590,377,329	5,914,424	867,033	P
7	Retail	67,398,350	17%	83%	61,248,021	139,471	127,976	C
8	E-commerce	8,333,319	50%	50%	2,250,984	204,688	4,102	C, R
9	E-commerce	71,068,955	75%	25%	35,197,874	222,050	113,531	C
10	Tech	1,955,375	60%	40%	2,943,890	22,390	7,625	C, R
11	E-commerce	13,339,044	50%	50%	11,633,187	106,534	225,241	C
12	Retail	5,566,367	50%	50%	10,070,742	54,423	215,227	C
13	E-commerce	3,716,015	77%	23%	2,121,967	22,305	7,518	C, R
14	E-commerce	86,766,019	80%	20%	36,814,315	471,501	15,722	C
15	Retail	9,753,847	50%	50%	8,750,270	19,365	76,177	C

* C = checkout, R = registration, P = page view

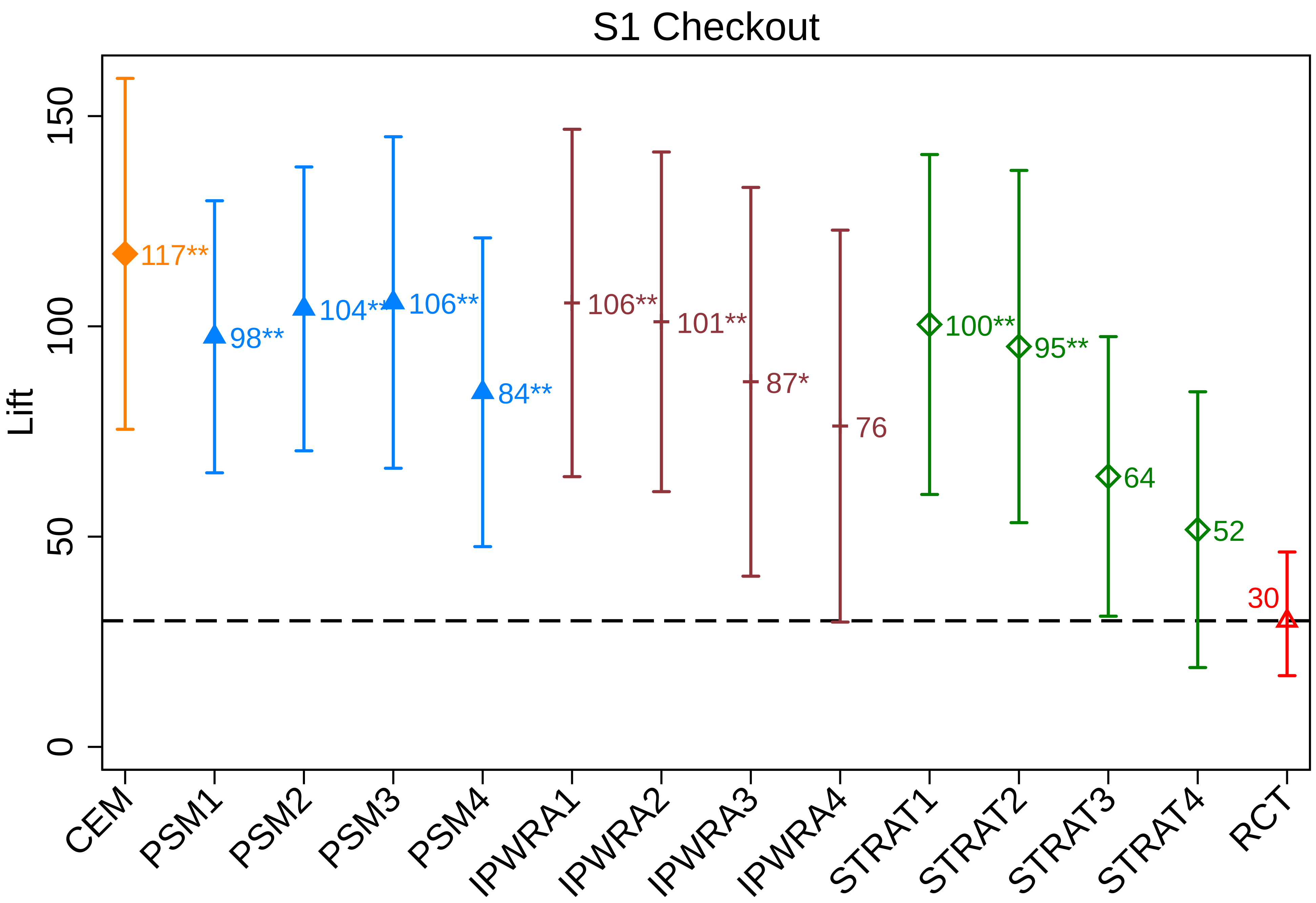
Study	Conversion	Control Conv	Test Conv	Expos %	ATT Lift	p-val	Exp-Unexp Lift
1	checkout	0.14%	0.17%	76%	30.0%	0.000	118%
2	checkout	0.04%	0.04%	47%	0.7%	0.407	278%
3	checkout	0.26%	0.27%	65%	8.6%	0.012	105%
4	checkout	0.04%	0.06%	37%	73.3%	0.000	213%
5	checkout	0.01%	0.03%	29%	410.4%	0.000	571%
7	checkout	0.32%	0.32%	50%	2.6%	0.048	33%
8	checkout	0.06%	0.06%	26%	-2.7%	0.404	81%
9	checkout	0.24%	0.24%	7%	2.4%	0.021	3836%
10	checkout	0.15%	0.15%	65%	1.6%	0.422	37%
11	checkout	0.33%	0.36%	42%	9.2%	0.000	294%
12	checkout	7.17%	7.25%	77%	1.3%	0.010	133%
13	checkout	0.37%	0.29%	43%	-56.7%	0.000	-66%
14	checkout	0.03%	0.05%	34%	63.4%	0.000	263%
15	checkout	1.81%	1.85%	81%	2.5%	0.006	26%

Study	Conversion	Control Conv	Test Conv	Expos %	ATT Lift	p-val	Exp-Unexp Lift
1	Registration	0.10%	0.74%	76%	786%	0.000	1018%
5	Registration	0.10%	0.45%	29%	899%	0.000	1343%
8	Registration	0.01%	0.02%	26%	68%	0.073	232%
10	Registration	0.47%	0.50%	65%	9%	0.035	35%
14	Registration	0.21%	0.39%	34%	165%	0.000	450%
2	Page View	0.01%	0.16%	47%	1532%	0.000	3332%
5	Page View	0.11%	0.36%	29%	605%	0.000	902%
6	Page View	0.46%	0.51%	60%	14%	0.000	271%

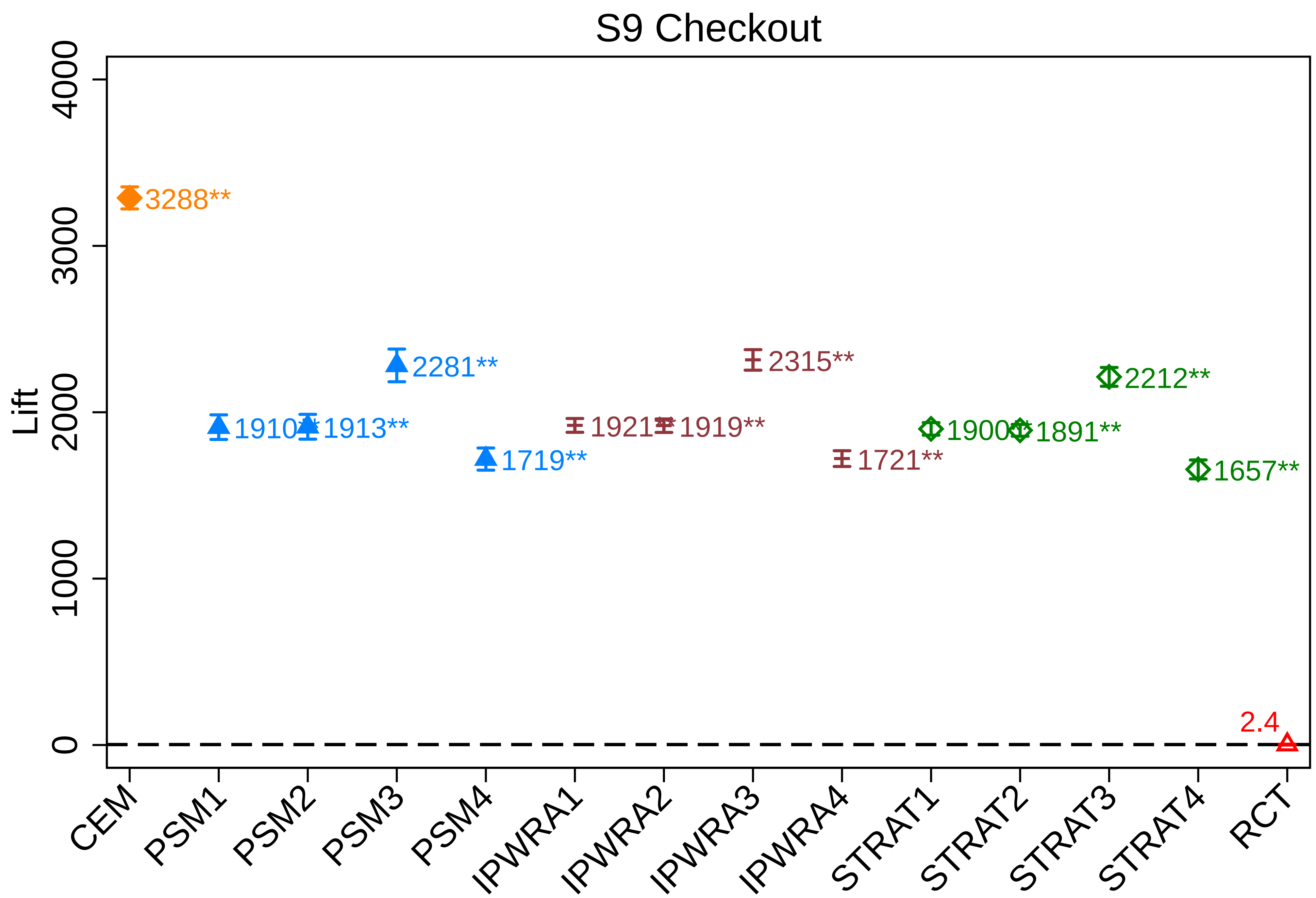
In some studies observational methods come close…



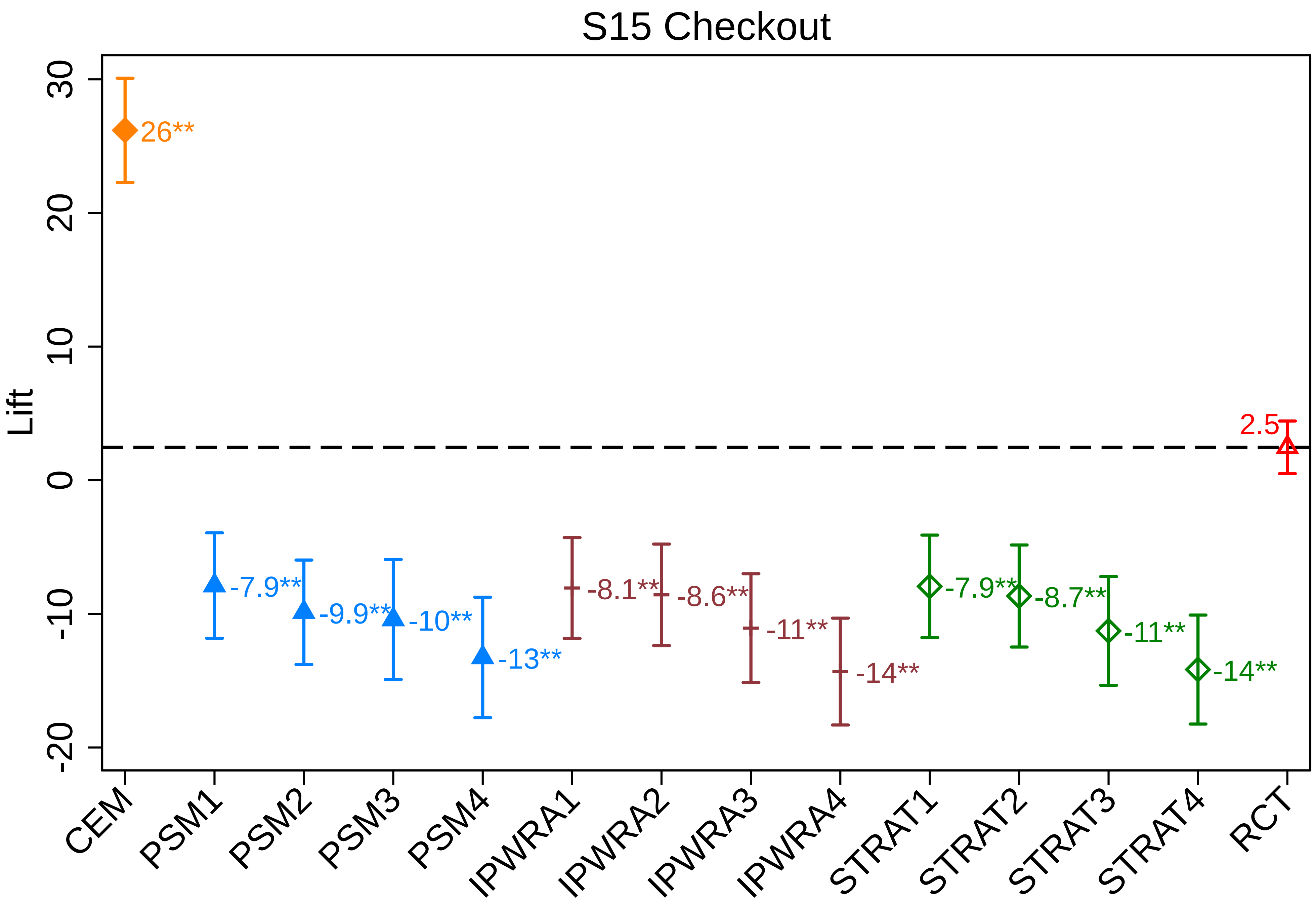
…and there might be a consistent pattern across methods



In other studies, lift estimates from observational methods widely *overstate* the RCT lift…



…and sometimes the observational methods *underestimate* the lift



Northwestern | Kellogg

		(A)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)
			CEM	Propensity Score Matching				Regression Adjustment				Stratified Regression			
ampaign	Outcome	RCT Lift*	Age, Gender	Age, Gender + FB Vars	Age, Gender + FB Vars + Census Vars	Age, Gender + FB Vars + Census Vars	Age, Gender + FB Vars + Census Vars + Activity	Age, Gender + FB Vars	Age, Gender + FB Vars + Census Vars	Age, Gender + FB Vars + Census Vars	Age, Gender + FB Vars + Census Vars + Activity	Age, Gender + FB Vars	Age, Gender + FB Vars + Census Vars	Age, Gender + FB Vars + Census Vars + Activity	Age, Gender + FB Vars + Census Vars + Activity
1	Checkout	30%	117%	98%	104%	106%	84%	106%	101%	87%	76%	100%	95%	64%	52%
2	Checkout	0.7%	428%					149%	141%	44%	35%	98%	99%	55%	40%
3	Checkout	8.6%	73%	20%	27%	55%	16%	21%	23%	41%	5%	18%	20%	33%	1%
4	Checkout	73%	221%	135%	128%	134%	92%	126%	122%	133%	100%	98%	87%	93%	74%
5	Checkout	410%	505%	407%	441%	429%	309%		429%	439%	305%	436%	429%	436%	300%
7	Checkout	2.6%	38%	19%	21%	-34%	-35%	19%	20%	-33%	-35%	19%	20%	-31%	-33%
8	Checkout	-2.7%	49%	28%	52%	47%	36%	36%	42%	55%	29%	33%	38%	54%	28%
9	Checkout	2.4%	3288%	1910%	1913%	2281%	1719%	1921%	1919%	2315%	1721%	1900%	1891%	2212%	1657%
10	Checkout	1.6%	37%	18%	17%	33%	-4%	21%	20%	35%	-13%	21%	21%	35%	-11%
11	Checkout	9%	276%	29%	31%	40%	7%	30%	31%	34%	3%	30%	31%	34%	2%
12	Checkout	1%	129%	111%	111%	82%	81%	112%	111%	82%	81%	112%	111%	83%	82%
13	Checkout	-57%	-66%	-46%	-46%	-29%	-29%	-47%	-47%	-30%	-30%	-46%	-46%	-31%	-30%
14	Checkout	63%	118%	81%	85%	103%	99%	80%	83%	91%	91%	74%	76%	84%	84%
15	Checkout	2%	26%	-8%	-10%	-10%	-13%	-8%	-9%	-11%	-14%	-8%	-9%	-11%	-14%
1	Registration	786%	1010%	1060%	979%	1042%	1002%	956%	958%	1079%	988%	823%	810%	429%	350%
5	Registration	899%	1259%	1052%	1086%	1041%	780%	1056%	1060%	1058%	728%	1099%	1098%	1081%	769%
8	Registration	68%	178%	157%	121%	121%	179%	148%	150%	155%	113%	153%	157%	159%	123%
10	Registration	9%	34%	17%	20%	27%	-2%	18%	18%	30%	0%	18%	18%	30%	2%
14	Registration	165.2%	289%	230%	227%	250%	241%	227%	227%	245%	234%	229%	227%	251%	239%
2	Page View	1532%	4311%					2471%	2479%	1182%	1190%	1225%	1243%	1777%	1258%
5	Page View	605%	839%	752%	741%	709%	491%	744%	744%	704%	476%	767%	767%	712%	497%
6	Page View	14%	235%									114%	118%	260%	289%

Conclusion

- There is a significant discrepancy between the commonly-used approaches and our true experiments in our studies
- While observations approaches sometimes come close to recovering the measurement from true experiments, it is difficult to predict a priori when this might occur
- Measurements are unreliable for checkout conversion outcomes
- Measurements are more reliable for registration or page view outcomes
- *Many industry participants seem unaware that this is a problem*