

Online Advertising Incrementality Testing And Experimentation

Industry Practical Lessons

Joel Barajas, Narayan Bhamidipati, James G. Shanahan

August 14th, 2021



Verizon confidential and proprietary. Unauthorized disclosure, reproduction or other use prohibited.



Who we are ...



Joel Barajas
Sr Research Scientist
Yahoo! Research,
Verizon Media
Sunnyvale, CA, USA

[LinkedIn](#)



Narayan Bhamidipati
Sr. Director, Research
Yahoo! Research,
Verizon Media
Sunnyvale, CA, USA

[LinkedIn](#)



James G. Shanahan
Church and Duncan
Group Inc
UC Berkeley, CA, USA

[LinkedIn](#)



Tutorial Parts

- 1. The basics: context and challenges**
- 2. Incrementality Testing: concepts, solutions and literature**
- 3. From concept to production: platform building, challenges, case studies**
- 4. Deployment at Scale: test cycle and case studies**
- 5. Emerging trends: identity challenges, industry trends and solutions**

Part 4

Deployment at Scale: test cycle and case studies

Experimentation: Incrementality Testing Cycle



Experiment Design and Planning

Minimum **Detectable Lift**, Total **Audience Available**, **Hold-out** group size, **Exposed Audience Needed**, Experiment **Duration**, etc



Intervention Execution

Start **holding out users randomly** from showing Ads. **Log “Ghost” Ad impressions** for these users



User Tracking and Metrics

Track **user cohorts**. Compute **aggregate metrics** to guide **high-level decisions** and campaign tuning



Incremental Performance and Recommendations

Document incrementality metrics, **CR lift**, **Cost per Incremental Action**. Provide **recommendations**, side effects and learnings

Experimentation: Design and Planning



Experiment Design and Planning

Minimum **Detectable Lift**,
Total **Audience** Available,
Hold-out group size,
Exposed Audience
Needed, Experiment
Duration, etc

Run statistical power analysis to manage expectations, budgets and duration

- In a period of time, given **existing advertiser data**:
 - **Number of users** exposed to impressions, conversions, impressions and total spend
 - **Metrics derived:** **converter rate, total audience, eCPM, ad frequency**
- Parameters to tune:
 - Holdout size (**never 50%!!**), minimum detectable lift, test significance

Power Analysis Simulation: 2 months of data

Aggregate Statistics:

Converter rate: **XXX%**

Conversions per 100 users: **XXX**

Conversions per converter: **XXX**

Impressions per user: **XXX**

Experiment Users: **XXXM**

Scenario Simulation:

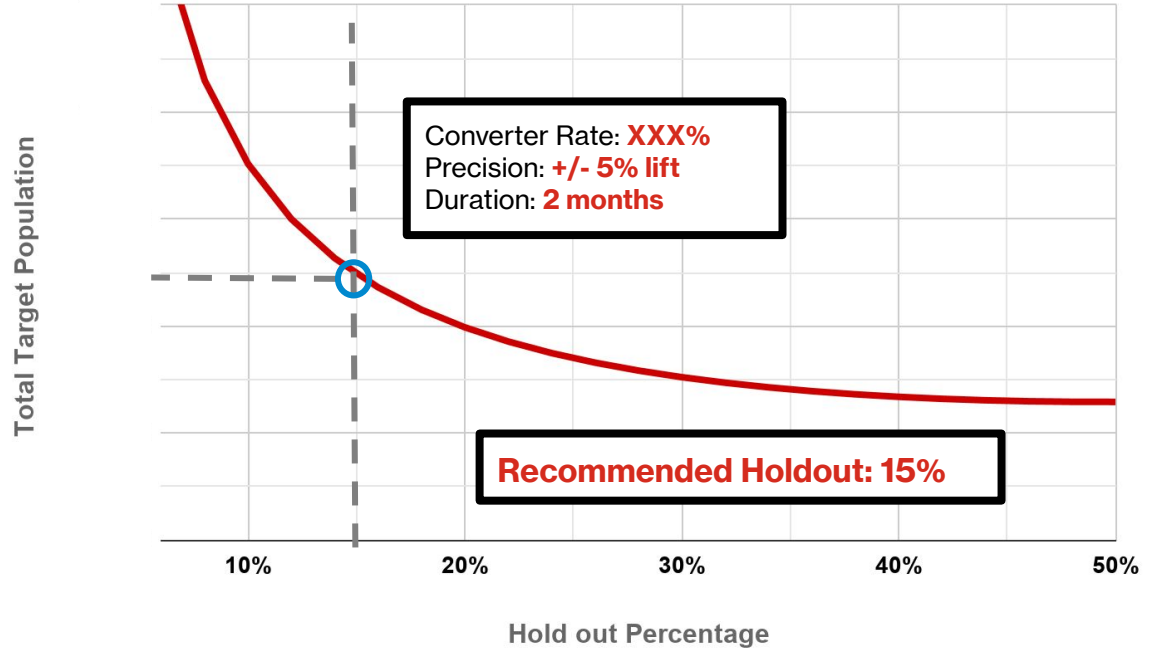
Converter Lift precision: **5%**

Expected Converter Rate:

Control: **<XXX%**

Test: **>XXX%**

Test Significance: **5%**



**Response: User
signup/acquisition**

Experimentation: Intervention Execution



Intervention Execution

Start **holding out users randomly** from showing Ads. **Log “Ghost” Ad** impressions for these users

The test actually runs

- Run quick **sanity checks**:
 - Verify **impressions and ghost impressions** are generated
 - Is **percentage of control users** as expected?
 - **Define guidelines** to hit a button and stop the test if needed

Experimentation: User Tracking and Metrics

Track user groups, metrics and mid-flight recommendations



User Tracking and Metrics

Track **user cohorts**.
Compute **aggregate metrics** to guide **high-level decisions** and campaign tuning

- Keep track of **top-level metrics**
 - Lift, incremental conversions, user ad frequency, CPIA, iROAS
- Pay special attention to **holidays** and fundamental user behaviour changes
 - For relevant holidays, **cumulative effects are realized** in these special periods (**ad stock realization**)
 - For irrelevant holidays sometimes **inventory becomes more expensive**

Experimentation: Performance and Recommendations

Testing is expensive, thus document learnings and insights



Incremental Performance and Recommendations

Document incrementality metrics, **CR lift**, **Cost per Incremental Action**. Provide **recommendations**, side effects and learnings

- Deliver **top-level metrics**
 - Including statistical tests and intervals
- Document lessons and **make recommendations**
 - What **can be learned from the test** beyond top-level metrics?
 - **Why** the test was (or was not) successful?
 - Make **business recommendations** and campaign management practices

Case Studies

How success looks like

Business Use Cases: Recap

Barajas and Bhamidipati (2021)

- **New advertiser** wants to *test waters* before fully deploying budgets
- Existing advertiser wants to **scientifically prove incremental value**
- Incrementality of **tactical practices** with external validity for future deployment
- Quarterly regular **strategic adjustments** (retrospectively)

Case Studies

New advertiser wants to test waters before fully deploying budgets

Onboarding Funnel: Key Findings



+200%

Statistically Significant Lift Effect on
Landing Page Visits (funnel entry)

12.4%

Consistently Positive Average Lift Effect
on **Purchase Confirmation (successful
end of the funnel)**

**Incremental Funnel
Entries and Successful
Endings**



Lift Results: Funnel Entries, Stages and Successful Endings

Lift on **Landing Page**, Funnel-entry Conversions

225% +/- 17%

Lift on **User Logins**, incremental quality (sticky) users

16.45% +/- 12%

Lift on **Complete Purchase Page Visit**, less funnel transitions

-6.58% +/- 8%

Takeaways:

- Incremental Funnel Entries, with **sticky more valuable users** (more logins)
- More straight Purchases (end of funnel) with **less transitions** (complete purchase page)

Lift on **Purchase Confirmation**, end-of-funnel conversions

12.4% +/- 10%



Case Studies

Existing advertiser wants to scientifically prove incremental value

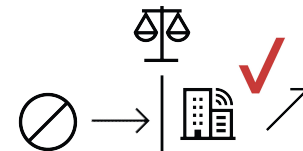
Insurance quotes: Key Findings

10.5%

Converter rate lift
(Statistically significant)

\$XX

Cost per Incremental Conversion over 8 weeks of experiment



9.5%

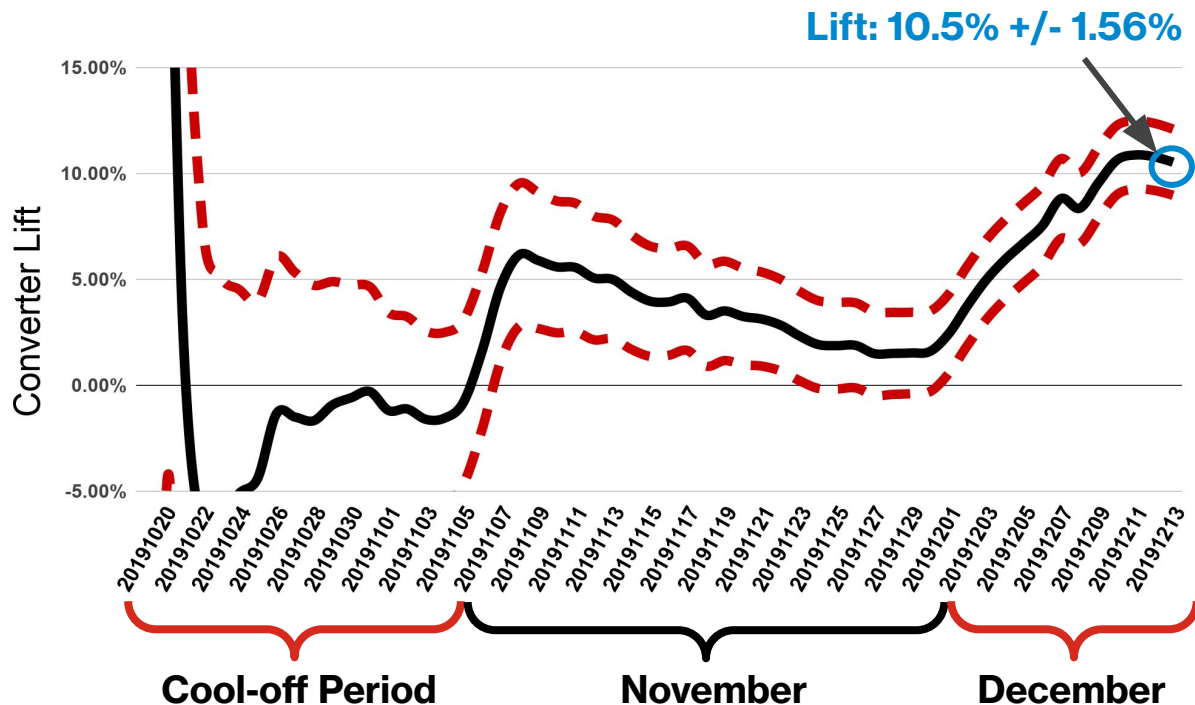
Lift in Avg conversions per user (Statistically significant)

87%

Lower Cost per Incremental Conversion than post-click CPA

Key Insights: Incrementality Real Study

Industry standard
last-touch attribution
undervalued native
display programmatic
advertising by 87%

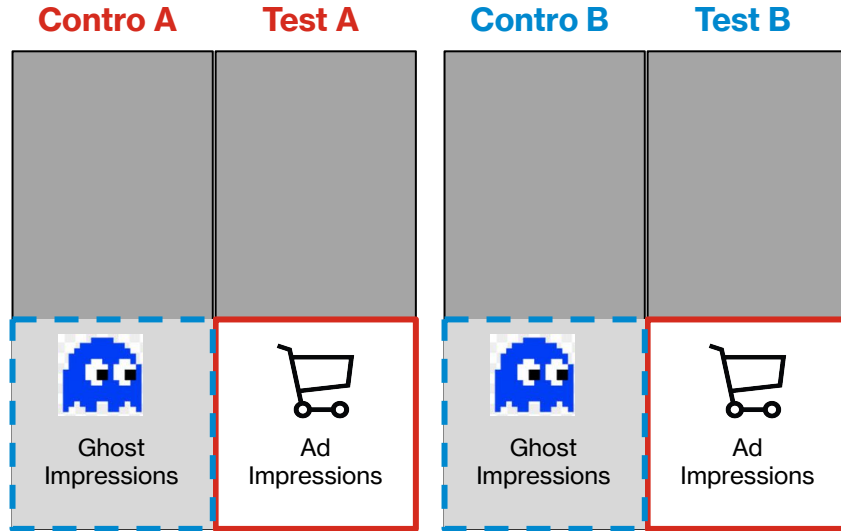


Cumulative converter lift progression. 95% confidence intervals are displayed. Conversion definition: insurance quote. Duration: 8 weeks, 10/18/2019 - 12/12/2019. Holdout size: 10%.

Case Studies

Incrementality of tactical practices with external validity for future deployment

A/B test + Incrementality test: **Multi-cell testing**



A benefit from double blind designs, an A/B split test fits easily to run a multi-cell test

User split is required to test A and B tactical practices

Tactical Practices on user Signups: Key Findings

4.5%

Tactic A cell reached
Statistically Significant
Converter rate lift

2.2%

Tactic B cell reached
Statistical Significant
Converter rate lift

**Tactic A
outperforms
Tactic B**

\$XX

vs

\$XX

**Cost per Incremental
Converter Tactic A**

**Cost per Incremental
Converter Tactic B**

What if we can't randomize user ids?

We'll review identity challenges and trends in the next and last part of the tutorial....