# *Designing Experiments to Measure Incrementality on Facebook*


C.H. Bryan Liu, Elaine M. Bettaney, Benjamin Paul Chamberlain

ASOS.com & Imperial College London

@liuchbryan

# *Facebook advertising has risen dramatically in importance*

In 2017, $204bn was spent on online advertising [1].

Amongst that, $40bn was spent with Facebook targeting 2.13bn monthly active users [2].

We find advertising on Facebook allows us to continue inspire fashion-loving 20-somethings to express their best selves and achieve amazing things.

# *The need to measure incrementality*

Over time, we moved away from measuring our ad campaign's performance using ROI / ROAS in favour of incrementality.

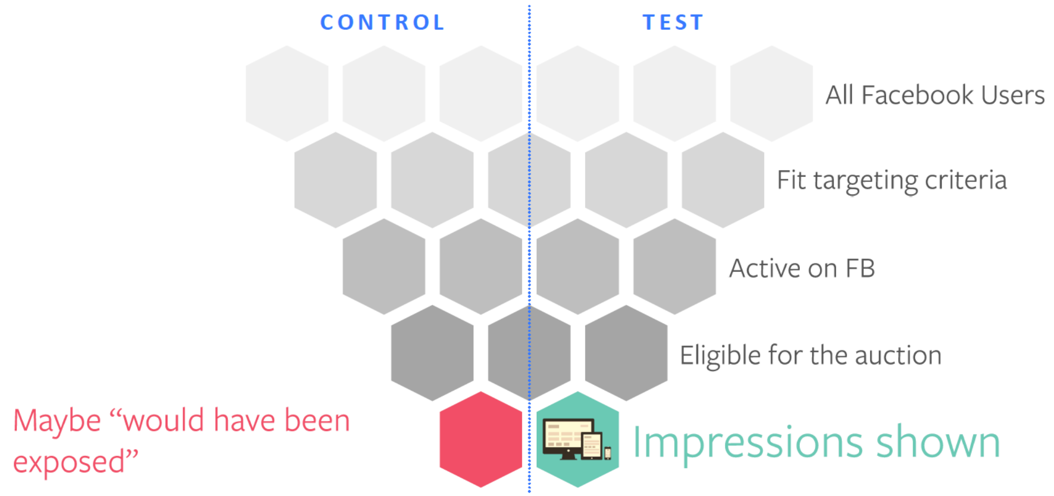
This allows a fairer evaluation of a campaign's performance, and hence enabling better attribution.

# They offer incrementality measurement via lift studies

A controlled experiment where the target population is split into two:
- Test group: Adverts shown
- Control group: Adverts not shown

Difference in a metric (e.g. total sales, number of app installs) between the test and control groups is the incrementality of the campaign.

Test vs. Control Comparison @ Facebook




Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# ISN'T THAT JUST AN A/B TEST?

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# A/B test with a few twists

An unreached audience

Not every one in the test group gets to see an ad!

A scaled control group

The group size of the control group is scaled to match that of the test group for a meaningful comparison.

A Facebook-controlled test/control split

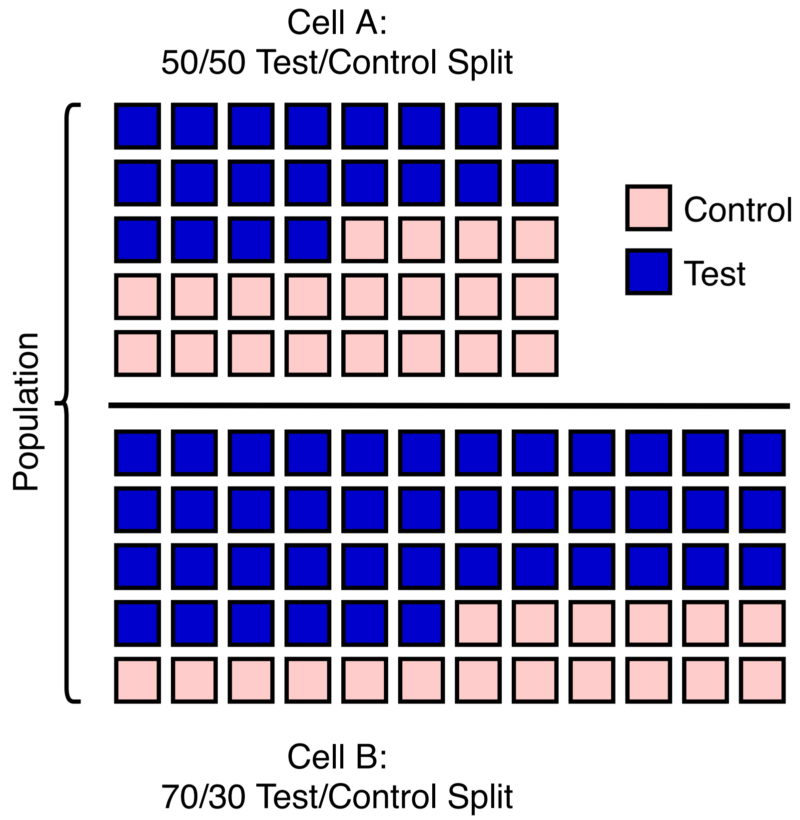
Advertisers have no access to user-level data for data protection reasons - only the aggregated level result is reported.

# Comparing marketing strategies via multi-cell tests

Facebook can split the target population into multiple *cells*. Different campaigns + test-control splits can be run in each.

These can be used to compare strategies where the target audience exhibits a selection bias [3] (e.g. comparing campaigns that vary the bid size based on customer lifecycle).

We are interested in measuring the difference between incrementalities attained by the campaigns. This is not reported by Facebook.*



# The current gap in literature

Advertising on Facebook

Controlled Experiment Design

# AGENDA

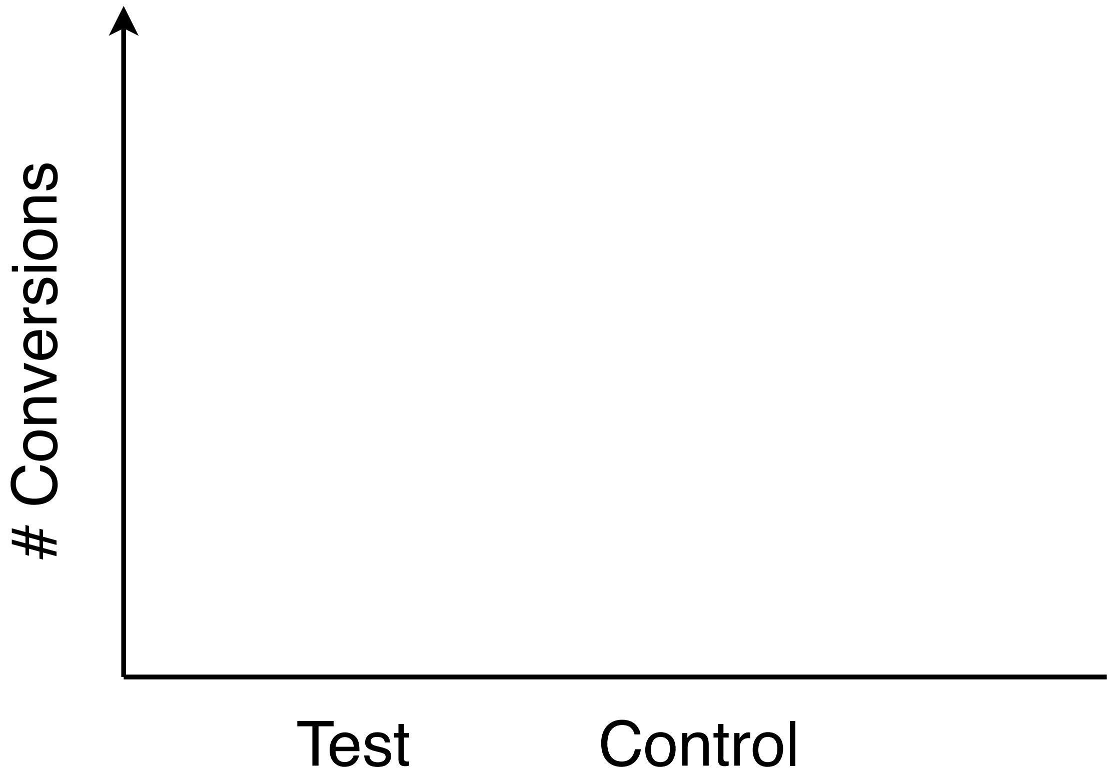
1. How does Facebook calculate incrementality and lift? (A quick overview of Gordon et al. [4])
2. How can we estimate the test power and sample size required to run a Facebook lift study?
3. Can we generalise the test power and sample size required calculations to a multi-cell study?

# Calculating Incrementality & Lift in a Facebook Lift Study

A Quick Overview

# *How does Facebook do it?*

We use # conversions as the running example, though the calculations apply to other metrics.



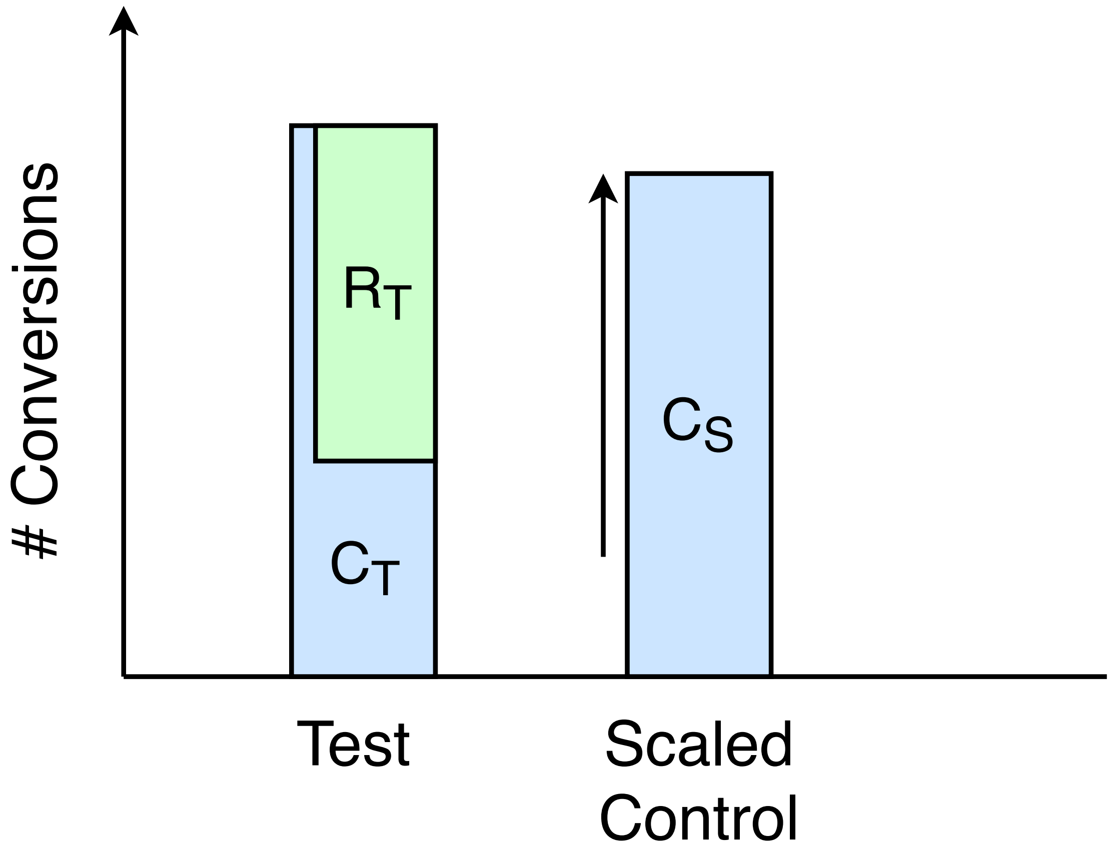
BRYAN LIU

# How does Facebook do it?

We use # conversions as the running example, though the calculations apply to other metrics.

Facebook tracks and reports the following:

$C_T$: # conversions in the test group
$C_S$: # conversions in the scaled control group*
$R_T$: # conversions in the reached audience of the test group



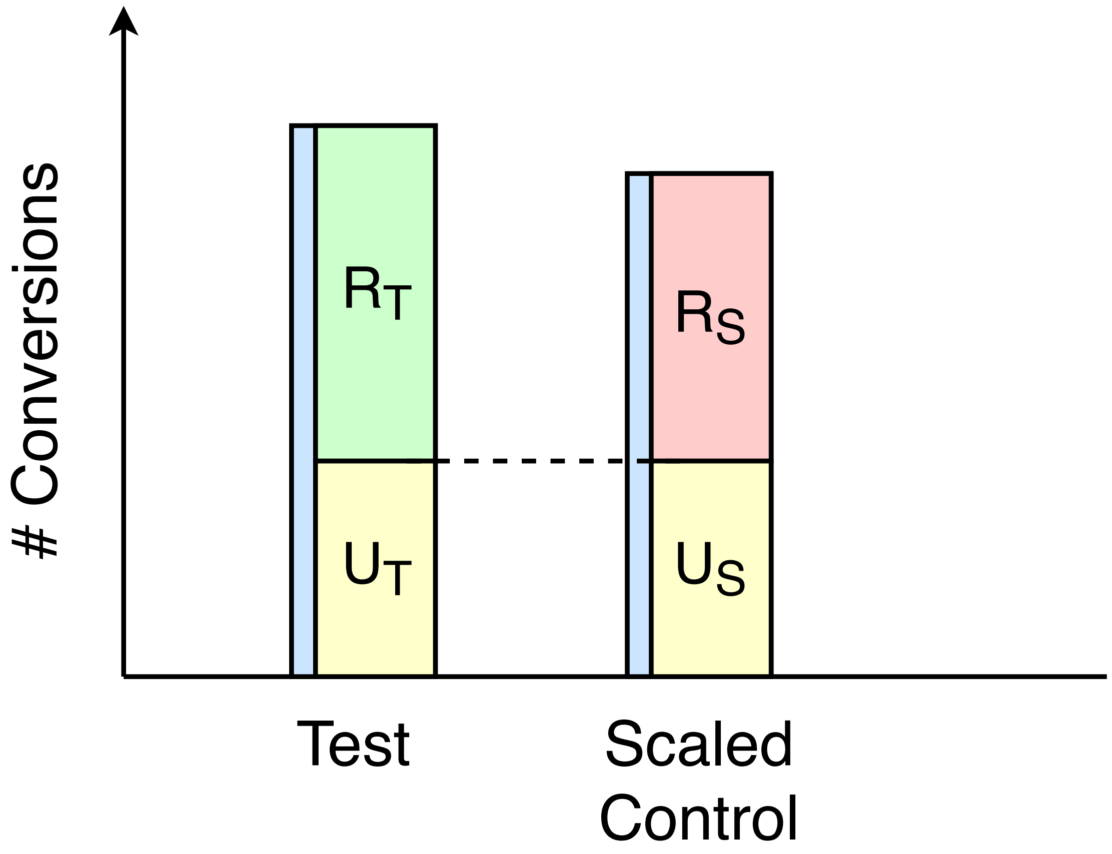

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# How does Facebook do it?

$C_T$: # conversions in the test
$C_S$: # conversions in the scaled control
$R_T$: # conversions in the reached audience of the test group

# conversions contains contribution from the reached and unreached audiences ($U_T$ & $U_S$).

The conversion rate for unreached audiences & control are assumed to be the same. $\Rightarrow U_T = U_S$.



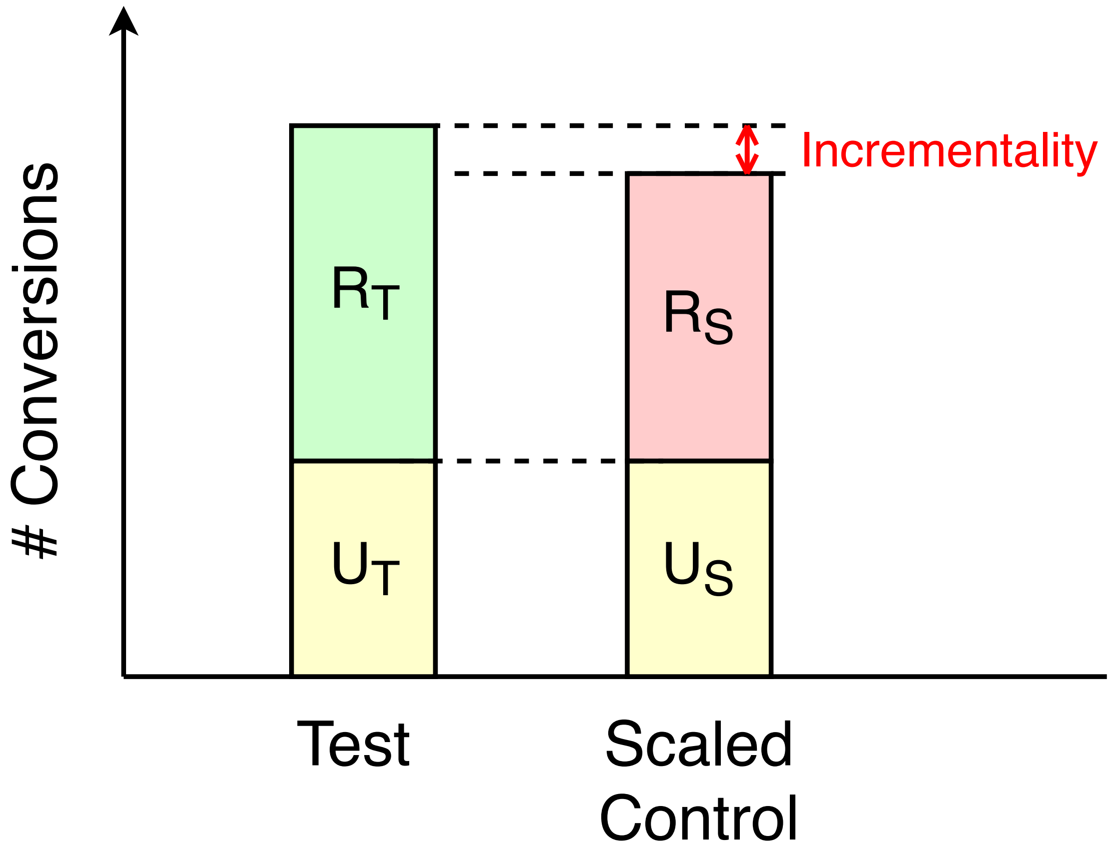

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# *How does Facebook do it?*

$R_T$, $U_T$: # conversions in the reached/unreached audience of the test group
$R_S$, $U_S$: # conversions in the reached/unreached audience of the scaled control group

The *incrementality* is the extra # conversions.

The *lift* is the incrementality relative to the # conversions in the reached audience of the scaled control. Facebook then runs a NHST to see if there is enough evidence for a non-zero lift.




Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# *Estimating the test power & sample size required*

*Filling in the missing piece*

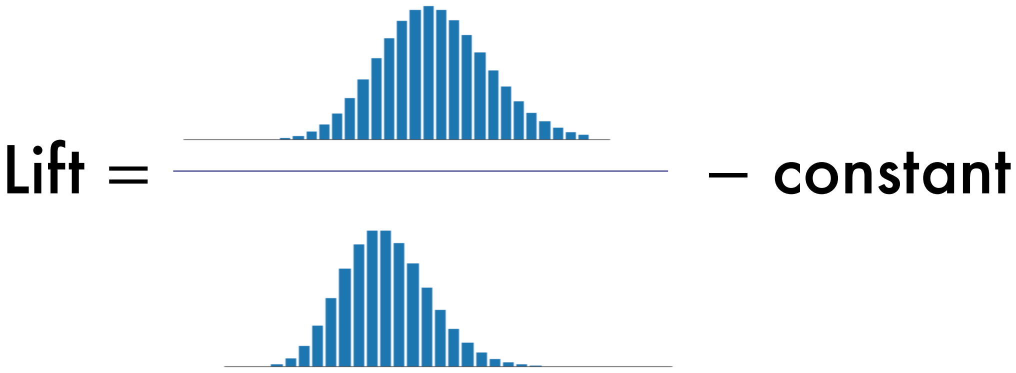
# We first require the distribution of the test statistic - lift

Two ways to do so:

1. Derive the analytical form of the CMF (assuming conversions follow a Poisson process):

$$F_L(l) \approx \sum_{k=0}^{\infty} \sum_{j=0}^{\left\lfloor \left(l+\frac{1}{r}\right)(rs)(k) \right\rfloor} e^{-(\lambda_T+\lambda_C)} \frac{\lambda_T^j \lambda_C^k}{j!k!}, \; l \in \mathbb{Q}$$

2. Simulating the distribution using Poisson samples.



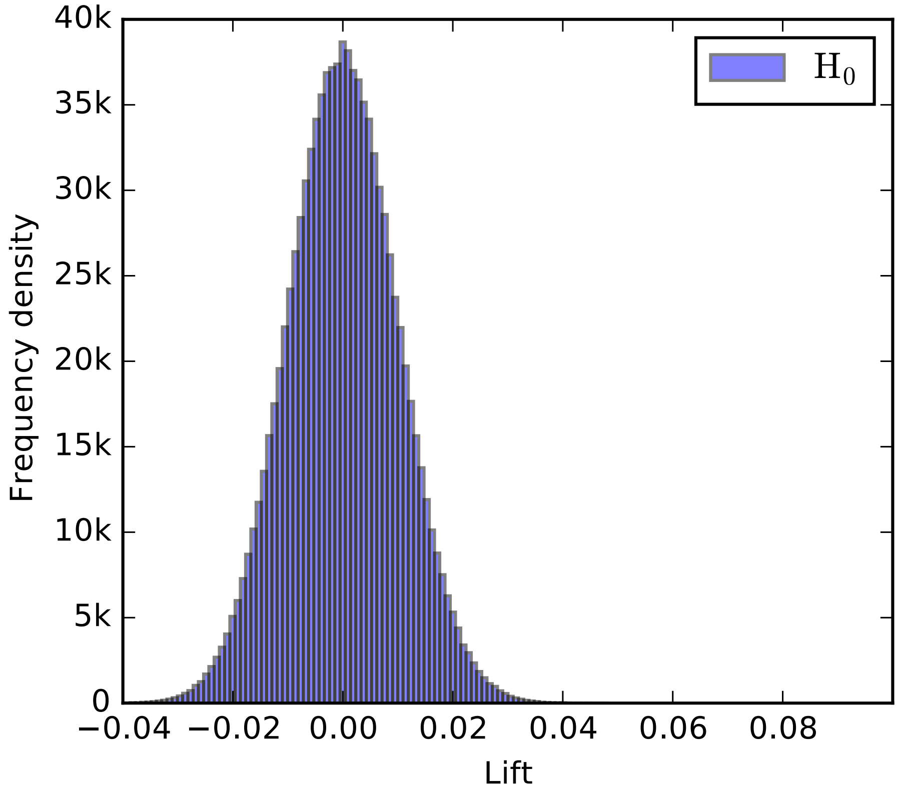

We show the two methods are equivalent, it is faster to simulate, and we can swap in other distributions to model the # of conversions or other metrics.

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# *Test power*

We can then estimate the test power w.r.t. a minimum detectable lift & a sample size as follow:

1. Derive the lift distribution under the null hypothesis, usually when expected lift = 0.



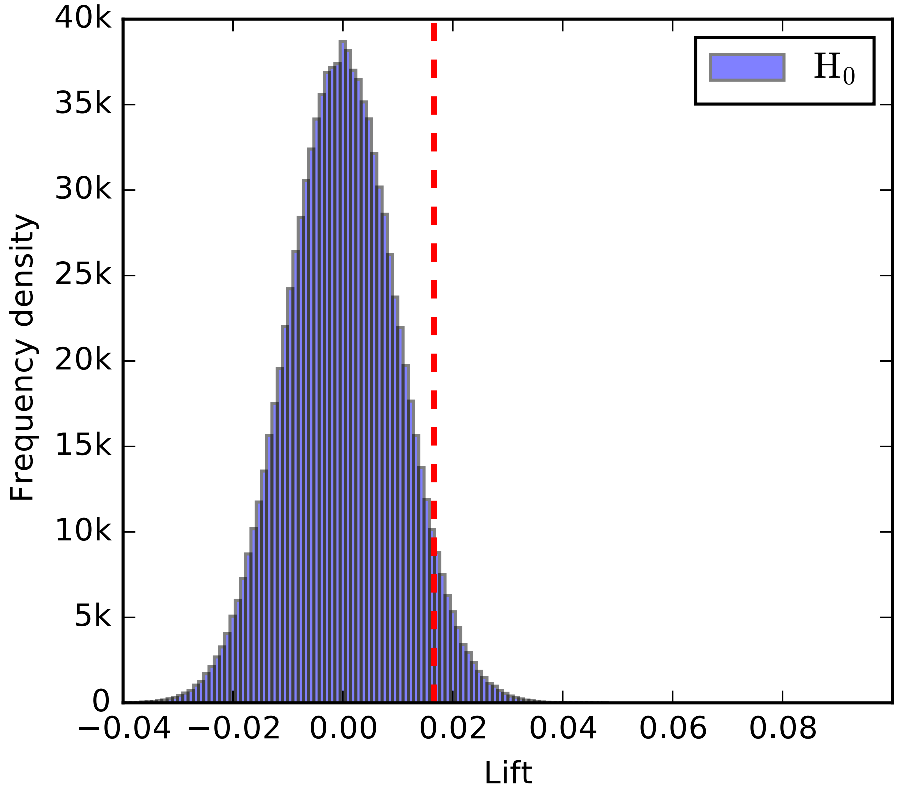

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# Test power

We can then estimate the test power w.r.t. a minimum detectable lift & a sample size as follow:

1. Derive the lift distribution under the null hypothesis, usually when expected lift = 0.

2. Find the critical value, usually the 95th percentile.



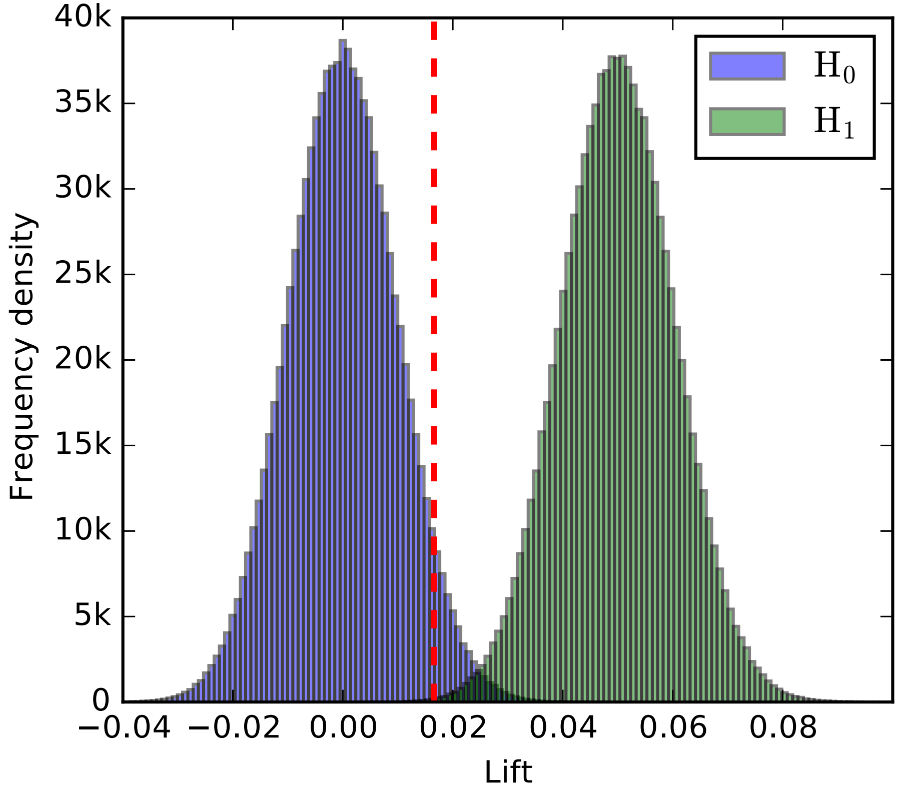
# Test power

We can then estimate the test power w.r.t. a minimum detectable lift & a sample size as follow:

1. Derive the lift distribution under the null hypothesis, usually when expected lift = 0.

2. Find the critical value, usually the 95th percentile.

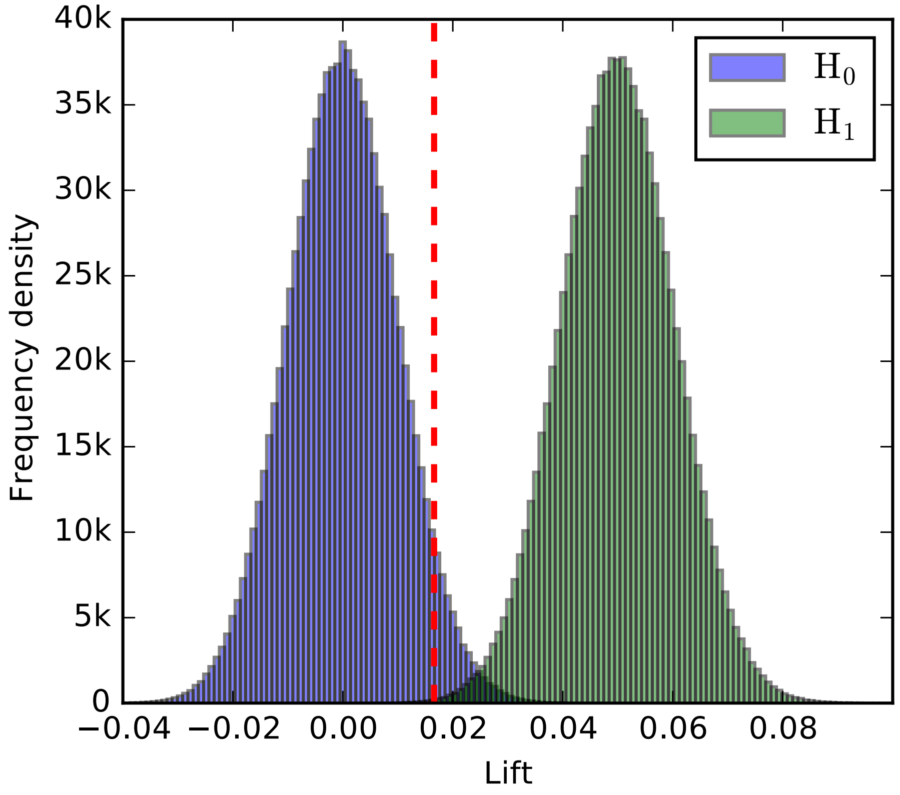
3. Derive the distribution under the alternate hypothesis (expected lift = minimum detectable lift), and find the proportion of the distribution over the critical value.




Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# Minimum sample size required

Given a target minimum detectable lift and test power (usually 80%), we can solve for the minimum sample size required using some root-finding methods.



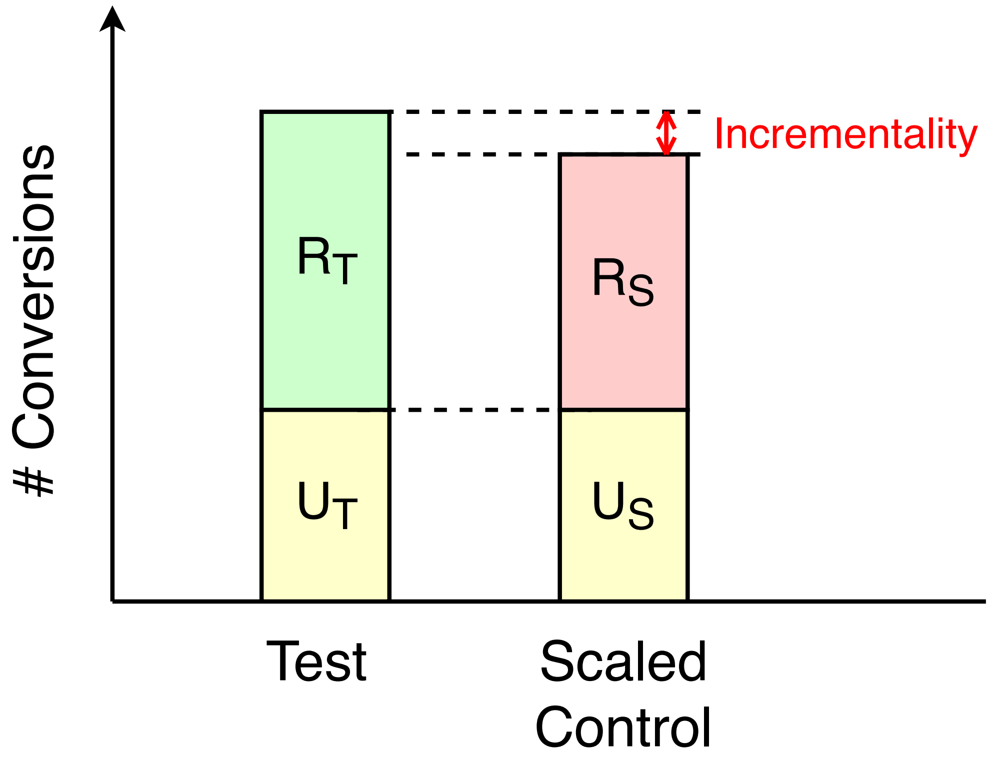

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# Generalising to multi-cell lift studies

*Same techniques, but a different test statistic*

# Comparing incrementalities

Recall we were finding the incrementality/lift of an ad campaign.

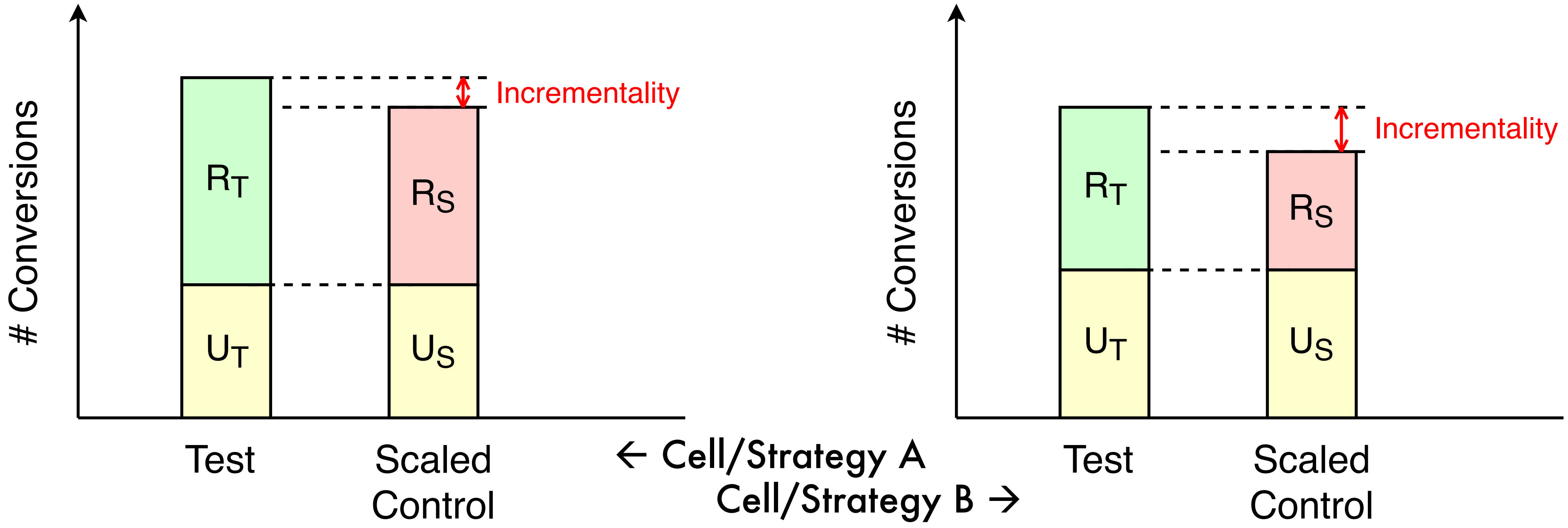



Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

BRYAN LIU

# Comparing incrementalities

In a multi-cell lift study, we are also interested in comparing the *difference* between incrementalities/lifts, and see if there is statistical evidence that it is non-zero.



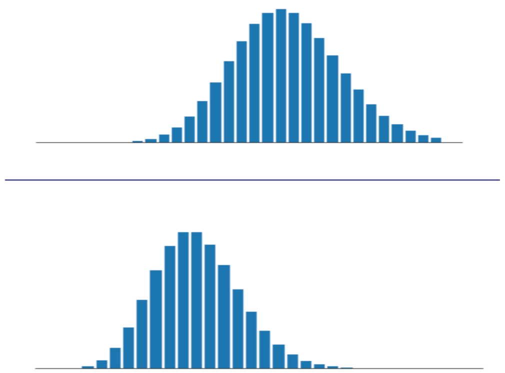

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# Test statistic

The difference in lift is defined as the (absolute, not relative) difference between the lift in Cell B and that in Cell A.

We simulate the distribution of the difference in lift, as it is very difficult to obtain an analytical form of the distribution.

Lift of a single cell =



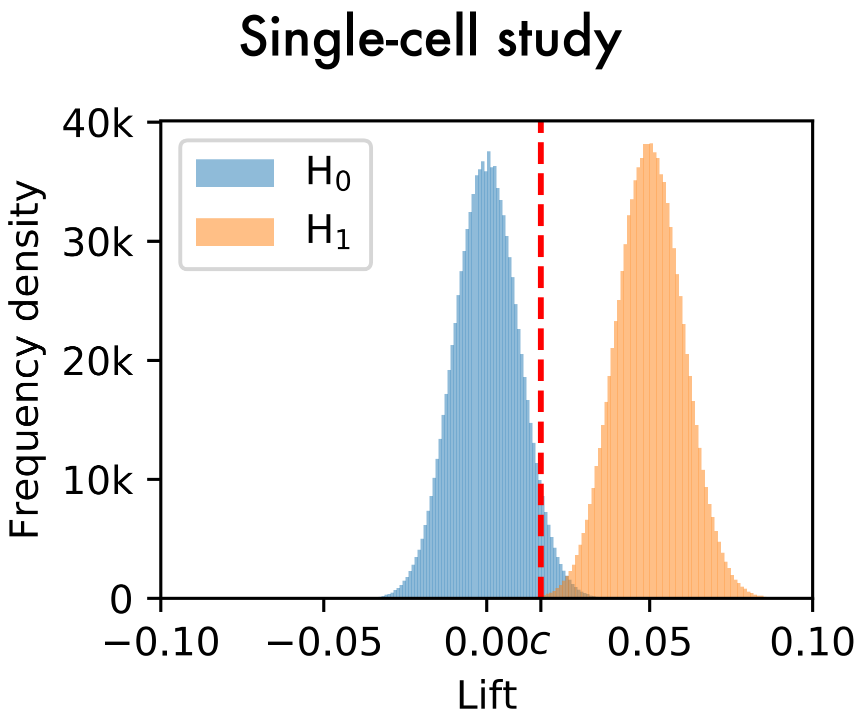
– constant

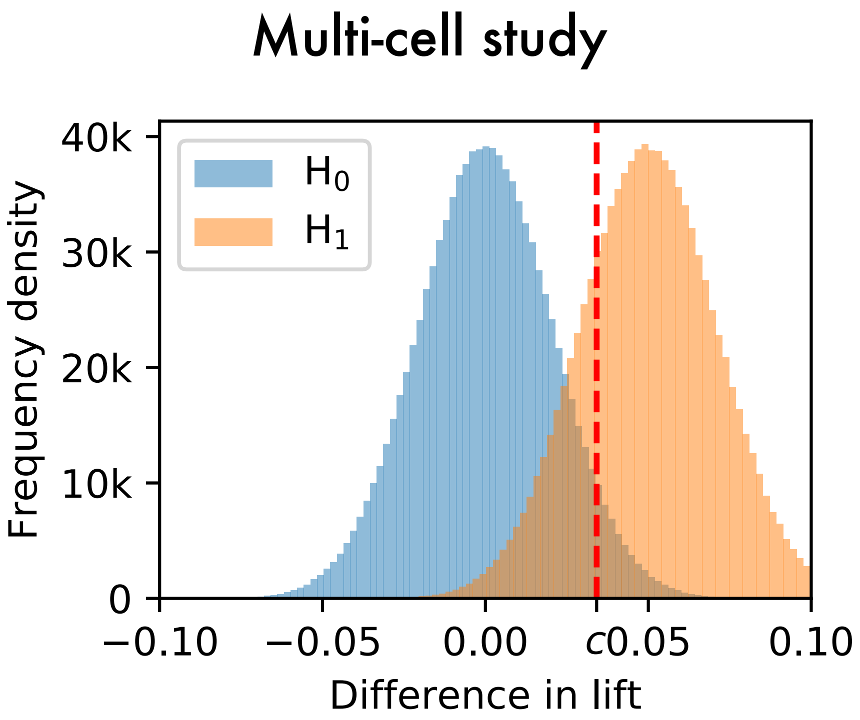
Difference between lift of cells =
Lift of Cell B – Lift of Cell A

BRYAN LIU

# The test statistic distribution is different!

The variance of the test statistic will increase even if the variance within each group stays the same, as multi-cell studies have more test/control groups.

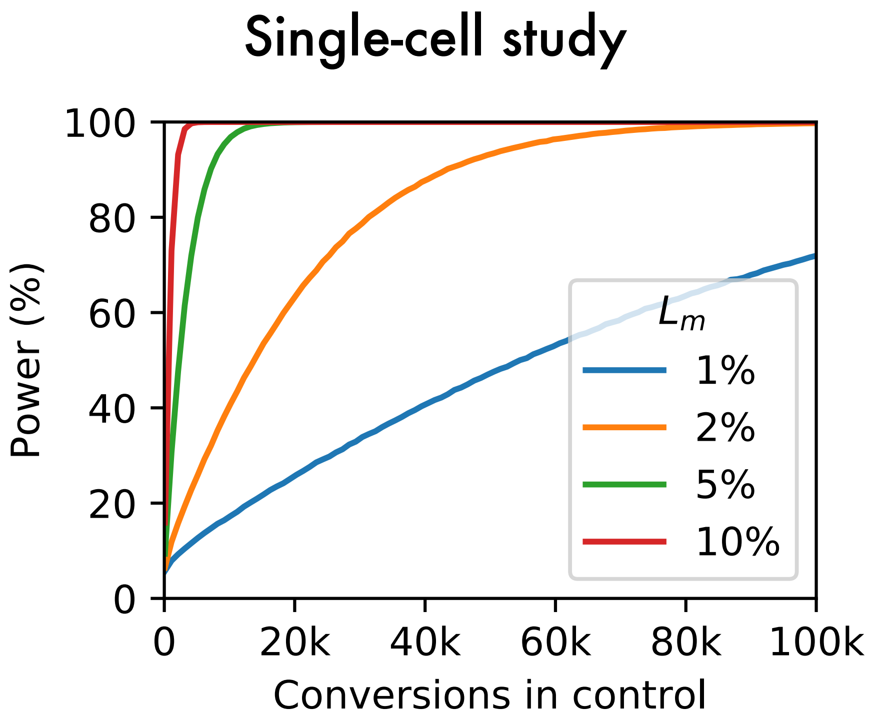


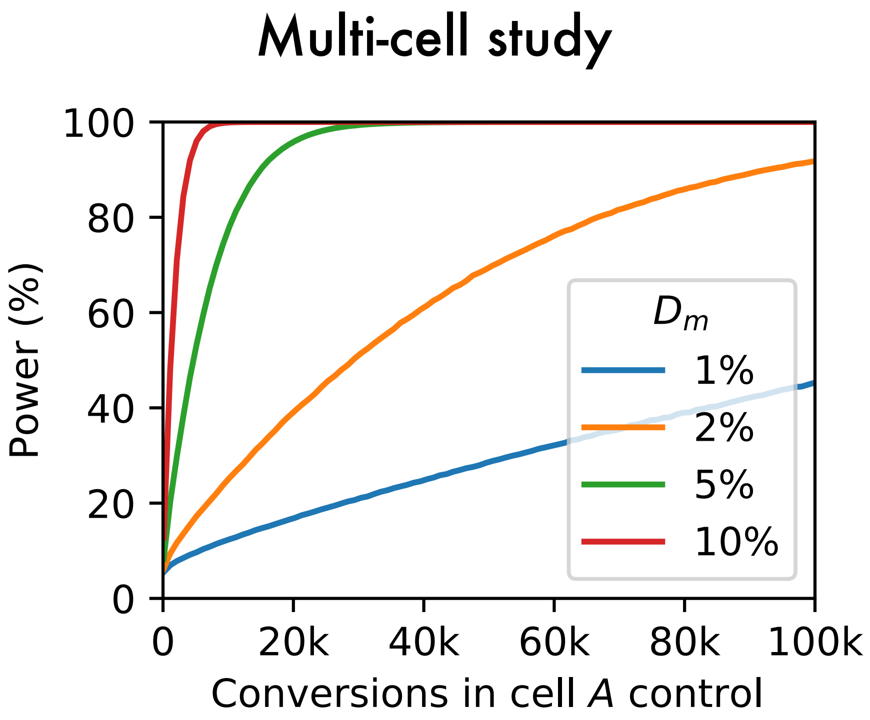




Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# The test statistic distribution is different!

Hence, a common pitfall in multi-cell studies is to use the test power and minimum sample size derived for single cell lift studies. Otherwise, the way to calculate test power is similar.



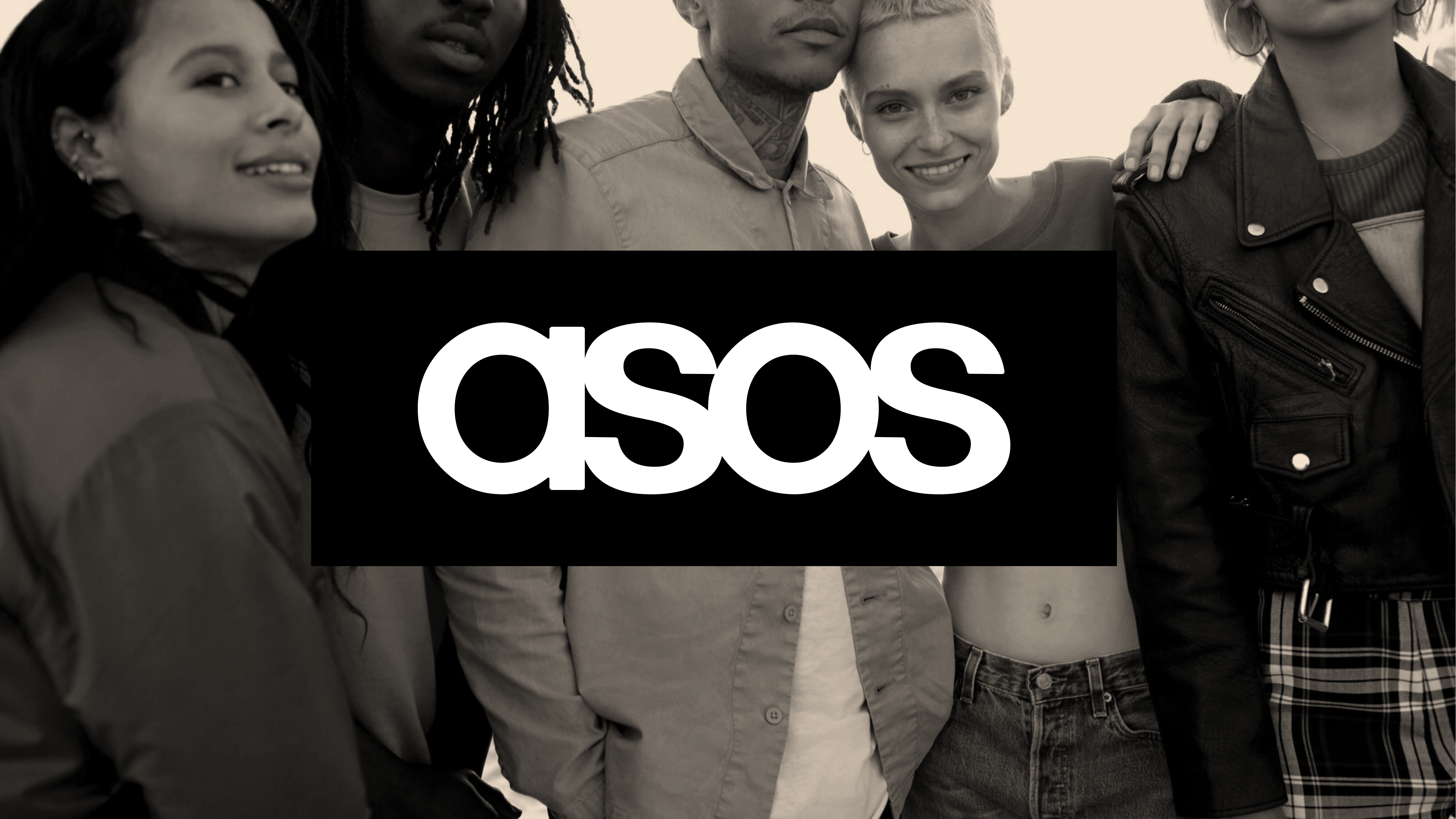

Liu, Bettaney, Chamberlain. Designing Experiments to Measure Incrementality on Facebook.

# SUMMARY

1. We went through how Facebook calculate incrementality and lift. (A quick overview of Gordon et al.)

2. We fill in the gap in literature by providing a test power & minimum required sample size calculation.

3. We generalise the test power and sample size required calculations to multi-cell studies, and show these quantities are very different to that of single-cell studies.


asos