Oracle Data Cloud

May 2018

WHEN KNOWLEDGE IS A SUPERPOWER.

Understanding CPG measurement methodology

CPG MEASUREMENT THAT SAVES THE DAY: UNPACKING WHAT MAKES AN EFFECTIVE METHODOLOGY

Ever feel like delivering campaign results requires a shield and sword? Inaccurate campaign results are like a villain conspiring to steal your budget. And with so much on the line (digital marketing spend is now outpacing TV) knowing how your campaign performs is critical to driving positive ROI for your brand.

CPG marketers have effective tools at their fingertips to measure the impact of their digital efforts, during and post campaign. As a pioneer in digital ROI measurement, and first to launch on all major media platforms, Oracle has evolved our measurement approach to incorporate methods critical to knowing how your campaigns perform and eliminate biases, noise, and false positives. For example, if your engagement metrics are falsely inflated you may think your audiences are engaging more heavily on a certain channel and then inaccurately allocate marketing spend to a channel that is ineffective for your brand.

FIGHT CAMPAIGN KRYPTONITE

A robust measurement solution is like a marketing super power—and it starts with how it's built. We believe there are four major components, or pillars, involved in delivering accurate, quality measurement.

- **Eliminate outliers**
  These are data points so far outside of the norms or trends in an experiment that they skew data analysis for no good reason.
- **Keep your test and control groups random**
  Every good data science experiment needs a test and control group, and in advertising technology it's a tall building to leap.
- **Find inconsistencies and reduce noise**
  There are inconsistencies that cause noise in an experiment. For example, if your sample size is too small, it will produce inaccurate results.
- **Add online activity data**
  Analyzing the amount of time and activity a user spends on a given platform is a primary driver to determining whether a person was exposed to an ad.

Combining these pillars gets us to the core concept of truly accurate advertising measurement: causal inference. This is a complex and nuanced science, which we'll explain for you, but the overall idea is simple. Causal inference answers the question: Did your intended audience pay attention to your ad and take action?

Each pillar is crucial to determining causal inference. To illustrate just how critical, we tested 27 campaigns across 4 media platforms to examine how each campaign behaved when we removed each component. Each time we removed a component then re-tested, we saw that taking away a component produced overinflated engagement metrics and false positives.

Removing outliers

Accurate measurement methodology relies on household-level purchase data. We often observe outliers—households that spend significantly more on a brand than expected in a given period of time. Outliers can be caused by a variety of technical and real-world phenomena like misidentifying what a household is. For example, multi-family units should be looked at as multiple households. Another example is loyalty cards, which are shared across consumers at a given retailer, such as when a consumer forgets his card or shares it with her roommate.

We began by excluding the outlier removal step from measurement in all 27 studies and comparing the results. Lift estimates calculated without first removing outliers deviated considerably from baseline lift. For a juice campaign, the lift estimate was more than four times that of the baseline.

Overall, lift estimates without outlier removal were, on average, 31% higher than the baseline. In other words, if you don't remove outliers your campaign results will be overinflated.

Keeping your test and control groups random

In any statistical experiment where you're trying to understand how or why one action causes another, you must create a test group and a control group. It's essential that each group is random. In advertising technology, it's almost impossible to randomize these groups. There are two reasons why:

1. Household data fragmentation causes the experimental groups to be contaminated. For example, multiple consumers living in the same household might use the same browser on a single device. If a mom and her teenage son are sharing a browser, this will create a distorted consumer profile. Two sets of online behavior and characteristics with different likes and interests were merged to create a muddy, inaccurate profile.

2. Almost every ad-serving engine has an optimization layer that gets better and smarter over time. This means the test group changes because the learning layer is applied, but the control group is not subjected to this optimization layer and thus behaves with its original characteristics.

So how do we solve for these discrepancies? We create a synthetic control group using a technique called entropy balancing. This assigns weights to households in the unexposed group to adjust for differences in pre-campaign spending, demographics, and online activity.

Typically, the exposed group (the audience exposed to an ad) exhibits dramatically different average sales than the control group, merely as a result of the targeting strategy of the campaign. For instance, in an ad campaign targeting brand buyers, the exposed group will likely show higher average brand sales than the control. The opposite effect might be observed in a campaign targeting buyers of a competitive brand.

> "Oracle Data Cloud has become Heineken's data and analytics backbone. They have become the yardstick by which we measure our digital-marketing spend value."
>
> – Ron Amram, Vice President of Media, Heineken USA

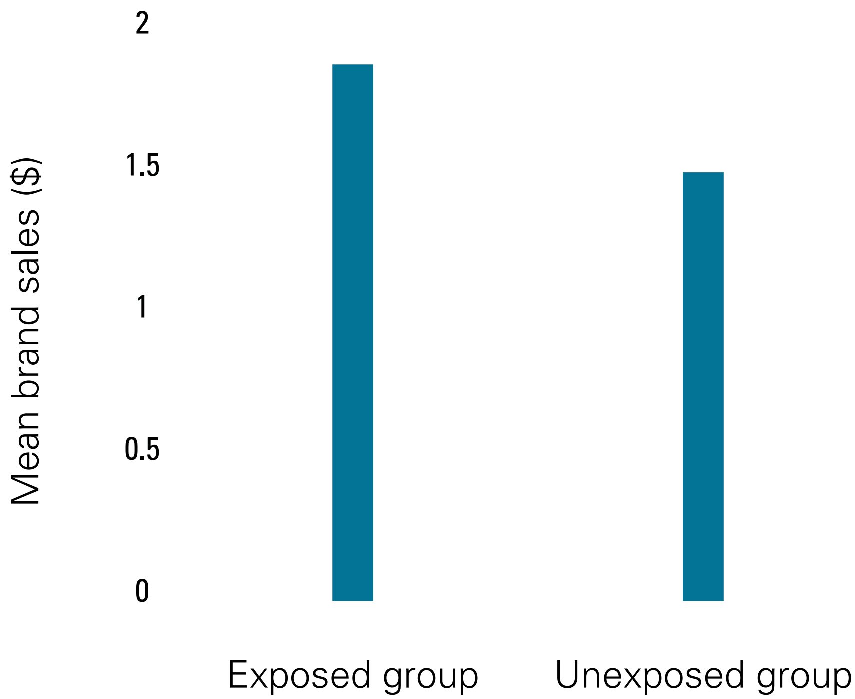

FIGURE 1

Figure 1 shows the average brand sales of a salty snack campaign for both groups in the pre-campaign period. In this case, the difference between the two groups means sales in the pre-campaign period were 33¢ per household. A similar difference would be expected during the campaign period. Therefore, an accurate lift estimate accounts for this discrepancy to isolate the true effect of exposure from pre-existing differences between the two groups.

Without entropy balancing, the lift estimates were on average 175% higher than the baseline.

Open web studies, or studies happening outside of walled garden consumer platforms, had an average overestimation of 326%. One extreme example shows that lift estimates for a juice campaign were five times larger than the baseline lift.

Smoothing the remaining inconsistencies after randomizing control groups

An outcome model is used to neutralize any remaining major discrepancies between the test and control group that entropy balancing did not catch. These models reduce bias and variances when calculating sales lift. For example, ensuring the sample size isn't too small to create an accurate measurement is called noise reduction.

Deploying an outcome model after balancing essentially brings the test and controls closer increasing the likelihood that the only difference between them is the presence of the advertisement.

For example, before deploying the outcome model, the groups may have similar income, past-purchase history, and age. After deploying the outcome model, the groups may have similar income, past-purchase history, age, interests, and web history. The more similar the groups look—with ad exposure the only difference—the more accurate your results will be.

Turning off the outcome model pillar led to inflated sales lift in 78% of tested campaigns.

If we turn off the outcome model, we see inflated lift estimates on 78% of the tested studies. On average, there was a 19% increase in lift estimates compared to baseline lift, with differences ranging from -49 to 126%.

Adding online activity data

Online activity data is a critical ingredient to a causal measurement framework. The actual amount of time and activity a user spends on a given platform is a primary driver to determining whether a person was exposed to an ad. We consider online activity to be crucial information to balance on in the entropy balance step, but how important is it?

In our analysis, 18 out of 27 studies showed 30% or higher deviation from baseline lift after excluding the activity data.

Online activity data was important in removing biases inherent in the nonrandom assignment of households to the unexposed and exposed groups. Therefore, inclusion of online activity data in a causal measurement framework is critical to ensuring the accuracy of lift estimates.

Measuring up

By now you might be wondering which pillar is the most impactful on measuring sales lift. The answer is all of them. Our ROI measurement studies show that removing any piece of the whole leads to bloated sales-lift estimates and significantly distorted campaign results.

Business decisions based on inaccurate results can badly damage your advertising strategy and your brand. Overestimating lift can make a media platform seem more desirable and cause the advertiser to overpay for advertising. Underestimating lift can result in an advertiser inappropriately and preemptively moving funds away from a campaign or platform in favor of one that is actually less effective for your budget and puts you further away from success.

A strong measurement solution should deliver fast, accurate, and reliable insights every time so you never feel like you need that shield again.

Want to be an advertising superhero? Contact us to ditch that shield and put on a cape with Oracle ROI inflight and post-campaign measurement and ensure your advertising measures up in every campaign.

> "Oracle Data Cloud is great in driving a lot of competitive advantages, when it comes to results that really prove and help establish advertiser budgets—then prove that advertising works."
>
> – Gunnard Johnson, Head of Measurement Science and Insights, Pinterest
>
> Pinterest