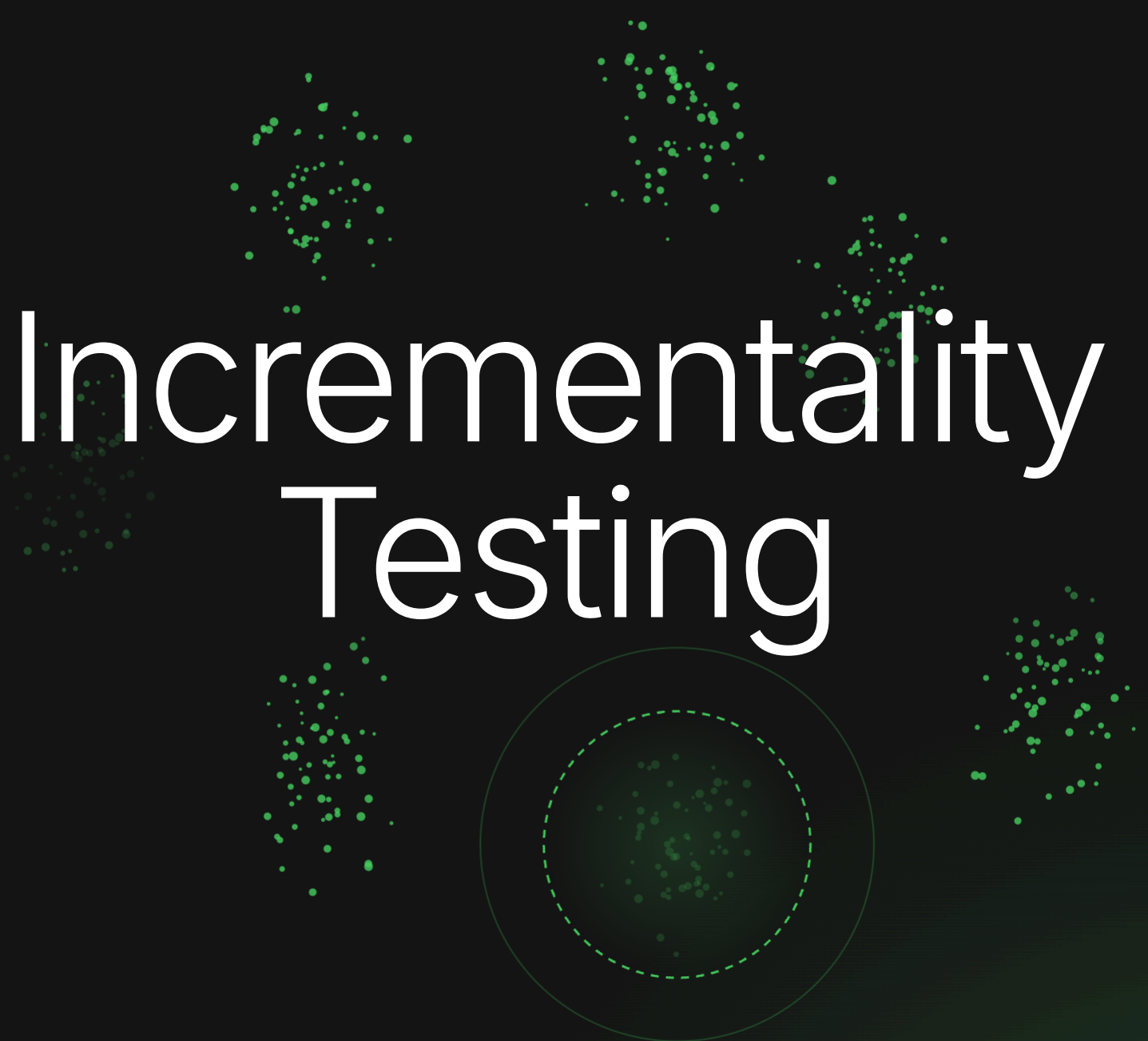


Incrementality Testing

The background features several clusters of small green dots scattered across the dark green field. In the lower center, there is a large circle with a solid green outer boundary and a dashed green inner boundary, creating a target-like effect.

Measure the true lift of your marketing channels and campaigns.

Overview

What is incrementality testing

- Incrementality measures the true impact of your campaigns and channels on conversions and revenue
 - It answers the question: would these results have happened anyway, without your advertising?
 - At Billy Grace, we measure this through a 30-day geo holdout test
-

Marketing intelligence as one connected system

- Incrementality works with our attribution models (MTA and UMM) as one system
 - It validates both models against real-world evidence
 - It feeds findings back to recalibrate UMM
-

How we set up and run a test

- A guided five-step process, from feasibility to applying the results
 - A 30-day holdout, monitored throughout
 - Google tests focus on conversions, Meta tests on traffic and demand
-

What incrementality makes possible

- A causal measurement to set against what MTA and UMM already show
 - A clearer view of which channels create demand and which capture it
 - Firmer ground for your next budget decision
-

Glossary

Confidence interval

A range around the estimated effect. A narrower range usually means a more precise estimate; a wider range means more uncertainty before acting on the number.

Contamination

Anything during the test that makes the holdout and the rest of the country harder to compare fairly (for example a promotion, a new campaign without geo exclusions, or a sudden spend shift).

Holdout region

The region where ads for the tested channel or cluster are paused.

Synthetic control

Our estimate of how the holdout would have performed if ads had stayed on, based on other regions that looked similar before the pause. We compare that expectation to what actually happened; the difference is the incremental impact (e.g. fewer sessions in Noord-Holland than peer regions suggest while Meta is off).

Minimum detectable effect (MDE)

The smallest true change your design can reliably detect, estimated in feasibility from your historical data.

Incremental lift/effect

The change we attribute to the pause, the outcome difference linked to turning the channel off in the holdout.

Marketing Triangulation

Using MTA, UMM, and incrementality together: incrementality validates click-path incrementality on lower-funnel tests (for example Google) and the effect of impressions on upper-funnel tests (for example Meta), with findings feeding back into UMM over time.

What is incrementality testing

Incrementality testing measures the true impact of a channel, or a cluster of campaigns within it. It shows the value that activity adds beyond what would have happened anyway and answers one simple question: would your results have been different if that marketing activity had not occurred?

There are several ways to measure incrementality. We use a geo holdout. We pause a channel in one region (the holdout) and keep it running everywhere else. For 30 days, we compare what happens in the holdout against what we would have expected if nothing had changed. That difference is the incremental effect.

This is a regional test, not a user-level A/B test. We pause ads across the whole holdout region rather than splitting users into groups, which keeps the comparison clean and the setup straightforward.



Marketing intelligence as one connected system

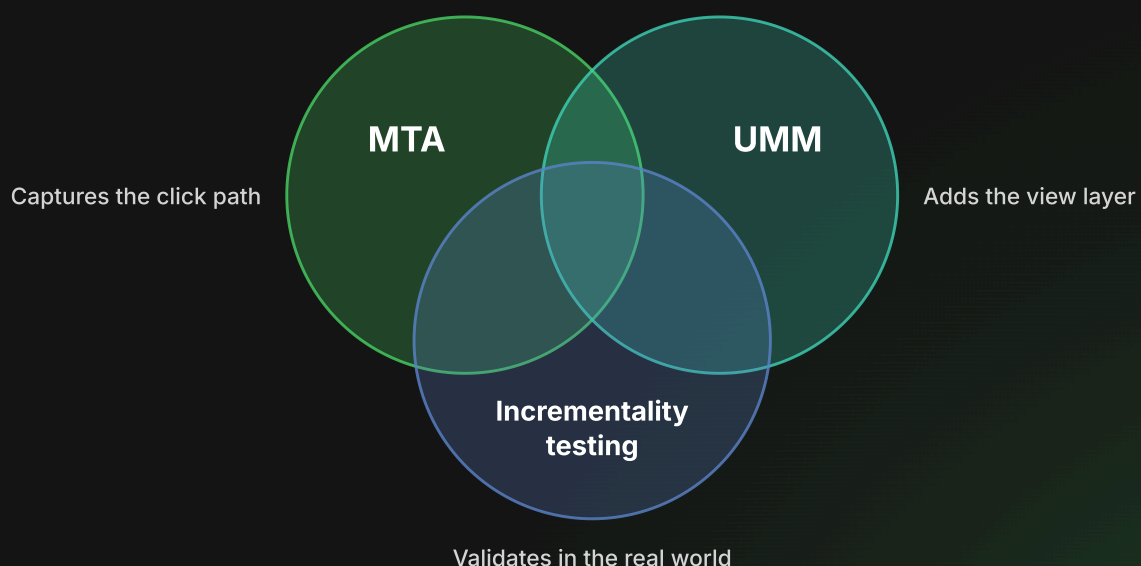
Multi-Touch Attribution (MTA), Unified Marketing Measurement (UMM) and incrementality testing are not three separate tools. Together they form one connected system (Marketing Triangulation), where each part checks and sharpens the others.

MTA captures the click path. It reads the clicks at the most granular level (channel, campaign, ad and keyword), so you can see which touchpoints drove a conversion. This is your tactical foundation, daily, granular, actionable.

UMM adds the view layer on top. It models how impressions from each channel drive sessions across all traffic sources, then integrates those view effects back into your MTA attribution. Now your attribution reflects both what users clicked and what they saw.

Incrementality validates both. It checks what happens in the real world when you pause spend in a region. Did sessions and conversions actually change? Did other channels move? It grounds both models in real evidence. When the test confirms the view effects, those findings recalibrate UMM for the next cycle.

All three work together, each one strengthening the others. That is the difference between a reporting stack and a measurement system.



When incrementality is useful

Incrementality testing is especially helpful when you are asking questions like these.

1 Are my campaigns driving results?

THE PROBLEM

Are the outcomes you see (sales, clicks, sign-ups) really due to the campaign, or would they have happened organically without it?

THE INSIGHT

Incrementality testing helps determine the **causal impact** of the campaign, so you can see whether spend is creating new demand or mostly capturing demand that was already there.

2 Which marketing channels deliver the most value?

THE PROBLEM

With several channels running at the same time, it is not always clear which channel is truly adding incremental growth.

THE INSIGHT

By **isolating the impact of one channel at a time**, you can see which channels add real extra growth and adjust budget with more confidence.

3 What is the optimal budget allocation?

THE PROBLEM

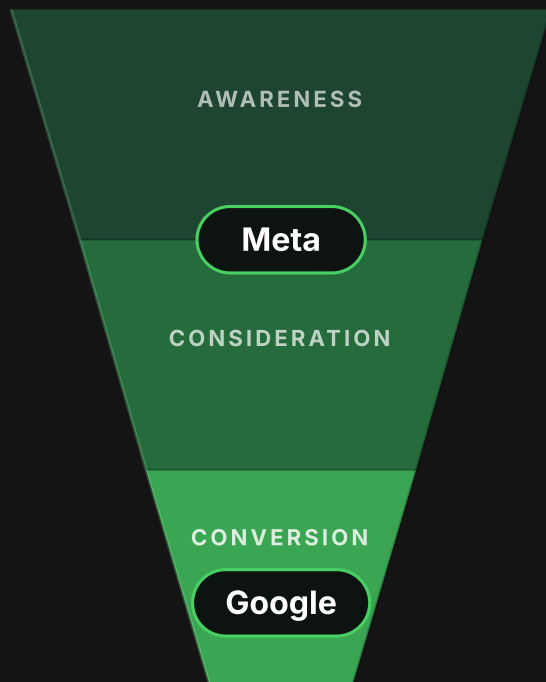
It is hard to know how to split the budget across channels for the best return.

THE INSIGHT

Tests can show where extra spend still moves the needle and where a channel may already be **saturated**, so you can rebalance and tap into underfunded areas with incremental potential.

Lower-funnel vs upper-funnel

The way we run an incrementality test is the same for all channels: we pause ads in one region and compare performance with the rest of the country. The main difference between lower-funnel and upper-funnel channels is what we measure and what kind of impact you should expect. Google and Meta are a clear example.



Same holdout method, different focus



Google

KPI · INCREMENTAL CONVERSIONS

Campaigns like Search, Shopping and Performance Max sit close to the moment of conversion, so the impact usually shows up in conversions. We also watch organic traffic, to see whether paid ads are replacing clicks that would have happened organically.



Meta

KPI · INCREMENTAL TRAFFIC & DEMAND

Meta campaigns sit earlier in the journey, around awareness and consideration, so conversions often land after the test window. We mainly measure traffic (sessions) to see whether Meta is driving demand. Conversions are still monitored, just not the main success metric.

How we set up and run a test

You set the geo exclusions in your ad platforms, pausing the selected channel in one region. **We handle the rest**, from feasibility through to interpretation.

01

Feasibility

02

Run

03

Monitor

04

Estimate

05

Result

STEP 01

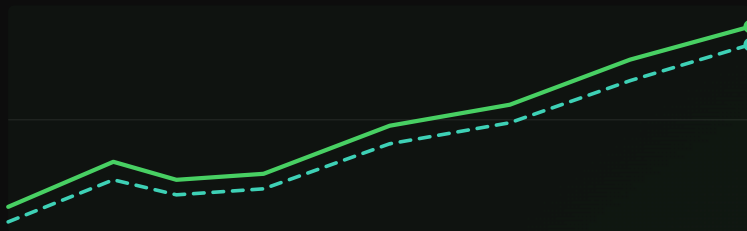
Feasibility check

Before recommending a holdout, we run simulations on your historical data (at least 90 days) to find regions and channels where a 30-day test can give a reliable answer. We estimate the minimum detectable effect, the smallest change the test could pick up, and cross-check against your attribution to confirm the channel is large enough to measure.

EXAMPLE

If Google Branded Search is **7%** of your last-30-day conversions, feasibility might show Noord-Holland needs a **10%** drop to detect an effect while Zuid-Holland needs only **5%**. Zuid-Holland is then the better region to test.

— Candidate region - - Synthetic control



Pre-test fit · the control tracks the region closely before the pause

STEP 02

Run the experiment

You pause or geo-exclude the channel in the selected region for **30 days**, keeping it active everywhere else. Pick a stable period, a sale or a big planned spend change can blur the comparison. Align with your team so new campaigns in the holdout use the right exclusions.

How we set up and run a test

STEP 03

Monitoring during the test

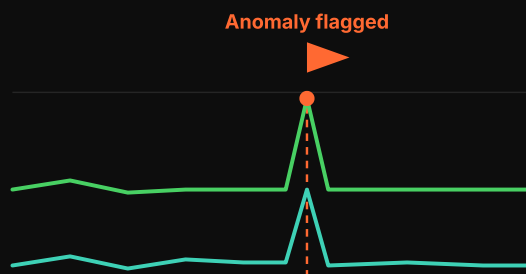
While the test runs, we watch spend and impressions for anything unusual, such as unexpected spikes or changes on clusters you did not mean to pause. If we see **contamination** that breaks the comparison, we may suggest ending the test early or extending it so the read stays trustworthy.

STEP 04

Estimation

Once the test ends, we estimate what would have happened if ads had stayed on, using a **synthetic control** built from regions that tracked the holdout closely beforehand. The gap between expected and actual is the incremental effect, with an effect size and confidence interval.

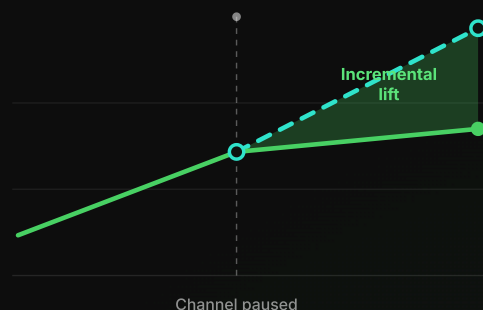
— Spend — Impressions



Monitored daily — anomalies surfaced as they happen

— Expected (synthetic control)

— Actual (paused region)



The gap between expected and actual is the incremental lift

STEP 05

Reading the result

We evaluate the effect, its confidence interval and significance to see whether the channel drove real incremental impact. How we use it depends on what you tested:

- **Conversion-focused holdouts:** whether attributed conversions were truly incremental or mostly shifting between channels.
- **Upper-funnel holdouts:** a read on sessions and cross-channel demand that can recalibrate how UMM models the effect of impressions for the next cycle.

Designing a test that gives a clear answer

Incrementality testing follows a basic idea: pause in one region and compare to the rest of the country. Good planning is what makes the difference. Even with a solid design, the wrong region, timing or changes mid-test often lead to inconclusive or misleading results. Knowing those pitfalls upfront makes feasibility and monitoring easier to interpret.

A well-designed test is worth the effort; a rushed one often ends inconclusive or misleading. A few common issues:

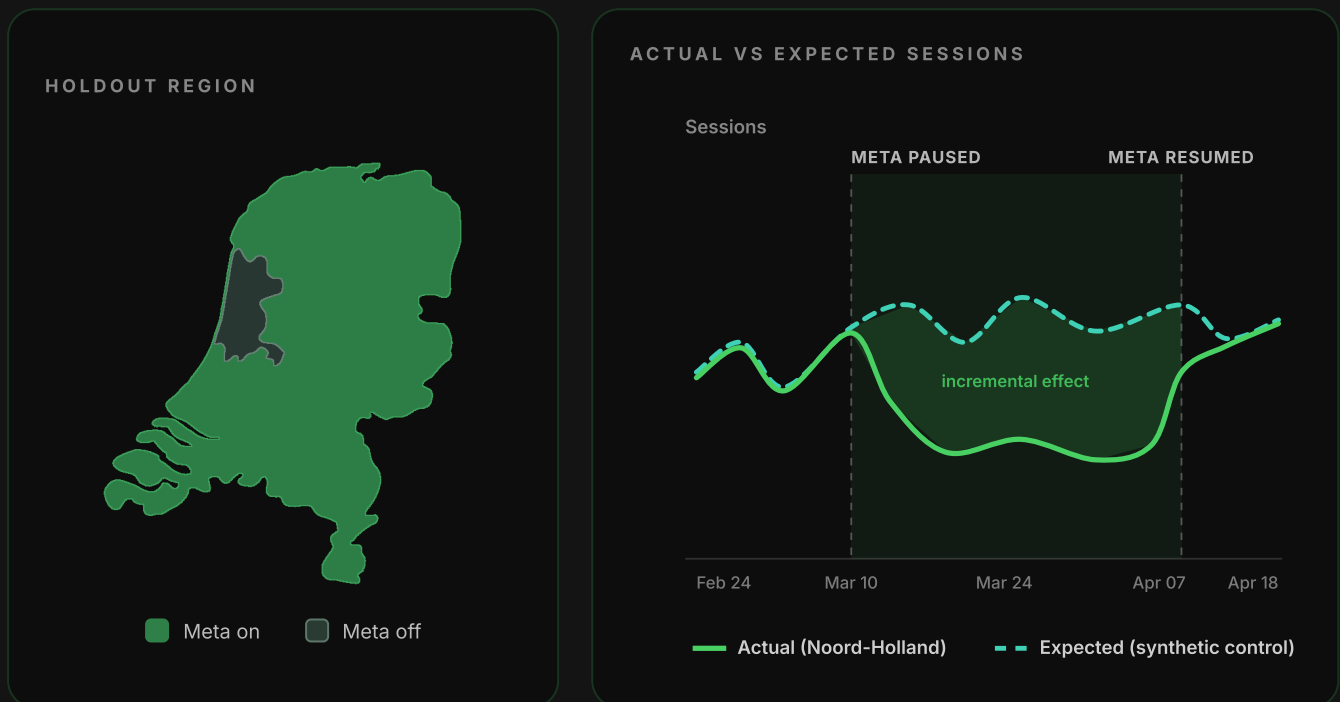
ISSUE	THE RISKS	WHAT WE DO TO HELP
Region or channel too small	The pause may not produce a change you can detect.	Feasibility estimates the minimum detectable effect (MDE) before you pause spend.
Busy period (sales, promos)	Regions may react differently to the same event.	We recommend stable periods .
Changes during the test	New campaigns, promos or spend shifts can blur the comparison.	Monitoring during the holdout for spend and impression anomalies.
Wrong metric for the channel	For example, judging Meta only on in-window conversions.	We align the primary KPI with the channel.

We run **feasibility first**, so you have a clearer picture before turning ads off, and we keep an eye on the test while it runs.

Example: a Meta holdout in Noord-Holland

Suppose you are running Meta ads nationally and want to understand the incremental value of that spend.

- **Holdout region:** in Noord-Holland, you pause all Meta ads for the duration of the test.
- **Rest of the country:** you continue running Meta as usual. This is your baseline.



- For 30 days, we compare Noord-Holland against what we would have expected if Meta had stayed on, built from regions that behaved like it before the pause (the synthetic control).
- We also watch how Google, organic and direct sessions move, to see whether demand spilled into other channels. That cross-channel read is something UMM and incrementality can validate together.
- If sessions fall clearly below the expected line while the rest of the country holds steady, Meta was likely driving incremental traffic. If the gap is small or uncertain, the result may be inconclusive, or Meta's role in that period limited.

This is where the value of a test shows. Google often carries high attribution because it sits close to the purchase, while Meta can look quieter even though it **built the demand** that Google then turns into a sale. The channel that looks strongest in its own platform reporting is not always the one creating that demand.

A real-world check on your measurement

Incrementality gives your measurement a real-world check, on top of what MTA and UMM already show.

A geo holdout adds evidence from an actual pause, on top of what the models and platform reporting show. It shows which channels create demand and which capture it, and gives you firmer ground for your next budget decision.

Speak to your Billy Grace contact if you'd like to explore running an incrementality test.