FROM THE RESEARCH BENCH

What is incrementality?

Incrementality isolates the true causal revenue lift produced by commercial interventions relative to an unexposed counterfactual baseline.

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher

PUBLISHED
2 September 2026

READING TIME
10 min

WORDS
2,116

https://isoglu.com/insights/journal/what-is-incrementality/

CONTENTS

MANAGEMENT SUMMARY

Incrementality measures the true causal impact generated by a commercial intervention above what would have occurred organically without it. While observational attribution models credit transactions to observed ad exposures, they systematically conflate correlation with causation, mistaking high-intent buyers for incremental conversions. Grounded in experimental economics and causal inference, incrementality relies on randomized holdouts, ghost ads, and geo-experiments to establish a credible counterfactual. This guide formalizes incremental lift and return on ad spend, contrasts causal testing with correlational attribution, details an enterprise paid-search holdout experiment, and outlines an auditable protocol for commercial budget allocation.

Keywords: Incrementality · Causal inference · Marketing attribution · Randomized controlled trials · Ad effectiveness

Incrementality is the measurement of the true causal lift generated by a marketing campaign, sales incentive, or commercial tactic above what would have occurred naturally in the absence of that intervention. Rooted in experimental economics and statistical causal inference, it establishes whether an observed transaction was directly caused by commercial expenditure or merely coincided with an existing purchase intention.

Standard commercial reporting relies heavily on observational attribution models, such as last-touch, first-touch, or multi-touch algorithms. These systems routinely award credit to touchpoints that simply intercepts existing demand. A prospect searching for a brand name on Google or viewing a retargeting display banner is frequently a committed buyer who would have converted regardless of the advertisement. Crediting these conversions to paid channels creates an illusion of high marketing efficiency while disguising massive capital misallocation.

In commercial governance, measuring incrementality separates productive capital deployment from expensive confirmation bias. When leadership fails to demand counterfactual proof, growth budgets gravitate toward campaigns that claim credit for baseline organic demand, systematically inflating reported acquisition returns.

How is incrementality formally defined and calculated?

Causal incrementality is evaluated by comparing an exposed treatment group against an unexposed control cohort (the counterfactual baseline) over an identical time horizon.

The incremental conversion lift equation

Let Y T represent the observed outcome (conversions or revenue) in the treatment group exposed to the commercial intervention, and let Y C represent the outcome in an unexposed control group, normalized by the group size ratio (N T / N C):

Incremental Lift = Y T − (Y C × N C N T)

The Incrementality Ratio (I R) expresses the proportion of total observed treatment revenue that is genuinely incremental:

I R = Y T Y T − (Y C × N C N T)

Where I R = 1.0 indicates that 100% of observed sales were caused by the campaign, while I R = 0.0 indicates that the campaign generated zero incremental sales, functioning entirely as an unearned tax on baseline revenue.

Incremental Return on Ad Spend (iROAS)

While nominal ROAS measures gross reported revenue divided by ad spend, Incremental ROAS incorporates only causally verified revenue:

iROAS = Ad Spend Incremental Revenue = Ad Spend Y T − (Y C × N C N T)

Blake et al. (2015) ran “a series of large-scale field experiments done at eBay” and report the brand-keyword result as an extreme case rather than a rule: “as an extreme case, we show that brand keyword ads have no measurable short-term bene”fits. The economics come from who the spend reaches: “more frequent users whose purchasing behavior is not influenced by ads account for most of the advertising expenses, resulting in average returns that are negative.” The half usually dropped is the boundary: “for non-brand keywords, we find that new and infrequent users are positively influenced by ads.”

Gordon et al. (2019) compare observational methods against randomized benchmarks on Facebook, and the count is worth stating as a count: seven of 14 checkout-conversion studies had observational point estimates off by more than a factor of three. In their worked study the randomized benchmark was a 73% lift with a 95% interval of 49% to 103%, the naive exposed-versus-unexposed comparison estimated 316%, and exact matching on age and gender alone still estimated 222%.

Lewis and Rao (2015) examined 25 field experiments and put the difficulty in one number: “The median confidence interval on return on investment is over 100 percentage points wide.” The cause is variance, not method: “relative to the per capita cost of the advertising, individual-level sales are very volatile; a coefficient of variation of 10 is common.” An interval that wide cannot separate a highly profitable campaign from a loss-making one.

Lewis et al. (2011) identify time-based activity bias in three controlled experiments: ad exposure coincides with a period of heightened brand-relevant and cross-site activity for reasons unrelated to the ad, which makes it “difficult to find a suitable “matched control” using prior behavior” and shows “how and why observational methods lead to a massive overestimate of adfx in such circumstances.” That is a matched-control failure rather than a persuasion effect.

Johnson et al. (2017) introduce ghost ads, a methodology that “facilitates this comparison by identifying the control-group counterparts of the exposed consumers in a randomized experiment.” Their claimed advantages are relative to the alternatives rather than absolute: against PSA and intent-to-treat A/B tests it “can reduce the cost of experimentation, improve measurement precision, deliver the relevant strategic baseline, and work with modern ad platforms that optimize ad delivery in real-time.” What this vault holds is a preprint, so read it as a method rather than a settled result.

Figure 1 The incrementality causal measurement hierarchy

Hierarchy tier	Methodology	Mechanism	Causal validity	Selection bias risk
Tier 1 (Gold Standard)	User-level RCT / Ghost Ads	Randomized holdouts with synthetic ad tags	High	Zero (statistically eliminated)
Tier 2 (Market-Level Causal)	Matched-Market Geo Testing	Synthetic control regions holding out spend	High to Moderate	Low (mitigated by pre-period matching)
Tier 3 (Econometric Calibration)	Calibrated Marketing Mix Modeling	Bayesian MMM constrained by holdout priors	Moderate	Moderate (requires experimental anchors)
Tier 4 (Observational Statistical)	Propensity Score Matching	Matched control groups on historical observables	Low to Moderate	High (vulnerable to unobserved intent)
Tier 5 (Flawed Attribution)	Multi-Touch Attribution (MTA)	Algorithmic weighting of observed touchpoints	Negligible	Critical (systematically claims organic lift)
Tier 6 (Commercially Misleading)	Last-Touch / First-Touch	Credits 100% of deal to arbitrary final click	Zero	Total (subsidizes bottom-funnel arbitrage)

Author's commercial measurement framework grounded in empirical field experiments from Blake et al. (2015), Gordon et al. (2019), Lewis and Rao (2015), Johnson et al. (2017), and Lewis et al. (2011).

Why do observational attribution dashboards report phantom returns?

The divergence between reported attribution metrics and actual business performance stems from three fundamental structural distortions in digital ad delivery:

Table 2 Why do observational attribution dashboards report phantom returns?

Distortion mechanism	Underlying behavioral cause	Dashboard consequence	Strategic commercial impact
Intent Harvesting	Retargeting bids aggressively on users already in checkout	Reports astronomical ROAS (e.g., 1,500%)	Starves top-funnel acquisition of growth capital
Activity Bias	Users browse actively across the web during high buying intent	Ads served during purchase surges claim causal credit	Mistaking correlation for advertising persuasion
Cannibalization	Paid search ads intercept users seeking direct brand domains	Paid clicks substitute 1-to-1 for free organic clicks	Converts free organic navigation into recurring cost
Cookie & Device Breakage	Cross-device paths cause false unexposed misclassification	Disregards long-cycle offline relationship building	Overweights short-cycle digital retargeting
Platform Self-Reporting	Ad networks grade their own performance with view-throughs	Total claimed conversions exceed total company sales	Inflates reported demand beyond real cash inflow

Table from this essay. Sources and interpretation are given in the article.

Understanding how attribution flaws distort broader economic metrics is vital. As detailed in What is CAC? and What is Contribution Margin?, calculating acquisition payback against un-incremented top-line figures understates true customer acquisition costs and overstates the marginal contribution generated by marketing teams.

Worked commercial example: Branded search holdout test

Consider an e-commerce enterprise spending $100,000 per month on branded Google Search ads. The ad platform dashboard reports stellar commercial success, claiming $800,000 in attributed revenue (an 8.0x ROAS).

To audit whether these returns are genuine or illusory, leadership executes a rigorous four-week randomized geo-holdout test. The market is split into two equal, demographically balanced geographic regions representing 50% of the customer base each.

1. Test design parameters

- Treatment Region (Ad Spend Maintained): $50,000 monthly ad spend.
- Control Region (Ad Spend Completely Paused): $0 monthly ad spend (100% holdout).
- Pre-period baseline sales between the two regions are statistically identical ($400,000 per month each).

2. Observed experimental outcomes during the four-week test

- Treatment Region Results (Ads Active): Paid Search Revenue: $400,000 Organic Search Revenue: $80,000 Direct & Other Revenue: $20,000 Total Realized Treatment Revenue: $500,000
- Control Region Results (Ads Paused): Paid Search Revenue: $0 Organic Search Revenue: $440,000 (users shifted seamlessly to organic links) Direct & Other Revenue: $25,000 Total Realized Control Revenue: $465,000

3. Calculating causal incrementality

The holdout test proves that 91.25% of the sales claimed by the ad platform were non-incremental ($365,000 of the claimed $400,000). When ads were turned off, consumers simply clicked the organic listing that sat two centimeters lower on the screen. While the dashboard reported an 8.0x return, the true causal return was 0.70x, destroying $15,000 in cash each month in the test market alone.

Connecting this reality to long-term valuation is essential. As shown in What is Customer Lifetime Value?, subsidizing acquisition with negative-ROI tactics permanently depresses cohort economics, regardless of how high gross retention appears on paper.

Which operational miscalculations undermine incrementality testing?

Table 3 Which operational miscalculations undermine incrementality testing?

Miscalculation	Root cause	Experimental failure	Corrective protocol
Running tests without sufficient sample size	Underestimating sales variance relative to small ad lift	Yields wide confidence intervals that span zero; inconclusive tests	Perform pre-test power calculations following Lewis and Rao (2015)
Treating pre-post changes as causal	Turning off ads nationally and measuring before vs after	Conflates seasonal demand swings and macro trends with ad lift	Always use concurrent control groups or synthetic geo-controls
Ignoring spillover and interference	Treatment exposure leaks into control regions via word-of-mouth	Violates SUTVA (Stable Unit Treatment Value Assumption)	Establish geographical buffer zones around test territories
Evaluating incrementality on short windows	Measuring conversions only during the active test period	Misses delayed conversions and ad-stock decay effects	Extend measurement windows post-intervention to track decay
Applying uniform incrementality across channels	Assuming paid social and brand search have identical lift	Over-funds brand search while under-funding true prospecting	Measure incrementality independently across discrete channel tiers

Table from this essay. Sources and interpretation are given in the article.

What auditable protocol establishes an incrementality testing program?

1. Classify channels by selection-bias risk. Segment commercial spend into high-bias channels (branded search, retargeting) and low-bias channels (broad prospecting, unbranded awareness).
2. Implement continuous holdouts on high-bias campaigns. Maintain a permanent 5% to 10% randomized exclusion group on all retargeting and branded search channels to monitor baseline cannibalization.

3 Deploy matched-market geo-testing for upper funnel. Utilize synthetic control algorithms to pair geographies with correlated historical sales, pulsing spend to evaluate market-level incrementality.

4 Instrument ghost-ad tracking where technically feasible. Utilize synthetic ad impression tags to isolate auction winners without incurring control ad costs, leveraging Johnson et al. (2017).

1 Establish iROAS and iCAC as primary governance metrics. Replace platform-reported ROAS and blended CAC on executive dashboards with causally discounted incremental metrics.

2 Reallocate capital from unearned demand to verified lift. Defund campaigns exhibiting incrementality ratios below 0.20, shifting capital to channels with positive causal marginal returns.

3 Calibrate marketing mix models against experimental priors. Feed empirical lift coefficients from randomized holdouts directly into econometric models to prevent attribution drift.

Where are the empirical limits of incrementality measurement?

Incrementality testing is a rigorous scientific protocol, not an effortless operational panacea. In business-to-business settings with long sales cycles, low transaction volumes, and complex buying committees, pure randomized controlled trials are often technically or economically infeasible.

Lewis and Rao (2015) also size the requirement: informative advertising experiments “can easily require more than 10 million person-weeks, making experiments costly and potentially infeasible for many firms.” So detecting a small lift is a budget question before it is a statistics question, and executives have to weigh the cost of measurement against the cost of deciding wrong.

The empirical foundations of this framework derive from leading econometric field research in digital advertising, specifically Blake et al. (2015), Gordon et al. (2019), Lewis and Rao (2015), Johnson et al. (2017), and Lewis et al. (2011).

author’s synthesis for defensible commercial capital management.

Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. Econometrica, 83(1), 155-174. DOI

Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. Marketing Science, 38(2), 193-225. DOI

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. Journal of Marketing Research, 54(6), 867-885. DOI

Lewis, R. A., & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. The Quarterly Journal of Economics, 130(4), 1941-1973. DOI

Lewis, R. A., Rao, J. M., & Reiley, D. H. (2011). Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In Proceedings of the 20th International Conference on World Wide Web (pp. 157-166). DOI

END OF ESSAY

HOW TO CITE THIS ESSAY

Isoglu, S. (2026, September 2). What is incrementality? Sinan Isoglu. https://isoglu.com/insights/journal/what-is-incrementality/

KEEP READING

From the research bench

An uplift claim needs a specification before it needs a number

https://isoglu.com/insights/journal/an-uplift-claim-needs-a-specification/

From the research bench

What are parallel trends? The assumption behind difference-in-differences

https://isoglu.com/insights/journal/what-are-parallel-trends/

From the research bench

What is external validity? A result needs a transfer boundary

https://isoglu.com/insights/journal/what-is-external-validity/

Sinan Isoglu, MBA (Quantic)
Commercial growth leader, lecturer and doctoral researcher