FROM THE RESEARCH BENCH

More advertising observations do not guarantee better decisions

More advertising data can improve precision, but decision value also depends on variance, effect size, observation cost, and action thresholds.

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher

PUBLISHED
1 September 2026

READING TIME
8 min

WORDS
1,815

https://isoglu.com/insights/journal/more-advertising-observations-do-not-guarantee-better-decisions/

CONTENTS

MANAGEMENT SUMMARY

More advertising observations can narrow an interval without changing the action a manager should take. Johnson, Lewis, and Reiley study how data volume and experimental design affect power in advertising experiments and show why variance, effect size, and the relevant decision threshold matter. Wernerfelt, Tuchman, Shapiro, and Moakler use a randomized experiment involving more than 70,000 Facebook and Instagram advertisers to estimate the value of offsite conversion-optimization data. In their reported setting, median cost per incremental customer rises from $38.16 with purchase optimization to $49.93 with click optimization, a 31% increase, and the loss of offsite data disproportionately harms smaller advertisers. This article builds a test-value screen. It does not analyze a current campaign or provide a universal sample-size rule.

Keywords: Advertising experiments · Statistical power · Decision value · Offsite tracking data

More advertising data can make an estimate more precise without making the decision better.

The short answer is that another observation has decision value only when it can change a relevant choice, at an acceptable cost, under the variance and effect size of the test. A large sample is not a managerial conclusion. It is an input to a test whose useful output depends on the action threshold.

Johnson, Lewis, and Reiley study how data volume and experimental design affect power in advertising experiments. Their evidence supports designing tests around statistical power and decision value rather than assuming that more observations automatically create better managerial information. Wernerfelt, Tuchman, Shapiro, and Moakler provide a complementary empirical boundary. In a randomized experiment involving more than 70,000 Facebook and Instagram advertisers, they estimate the value of offsite conversion-optimization data. In the reported setting, the median cost per incremental customer rises from 38.16 w i t h p u r c ha seo pt imi z a t i o n t o49.93 with click optimization, a 31% increase. The sources answer different questions, but together they show why data volume and data value should not be treated as synonyms.

Why must advertising measurement design begin with commercial decision thresholds?

Suppose a manager is choosing between two advertising actions. The first action is more expensive but may produce higher-quality conversions. The second is cheaper to observe but may optimize a weaker signal. As the dataset grows, the estimated difference becomes more precise. The manager still needs to know how large the difference must be before the action changes.

Table 1 Why must advertising measurement design begin with commercial decision thresholds?

Test field	Question	Why it matters
Decision	Which action will change if the evidence moves?	Without an action, precision has no stated use.
Minimum relevant effect	How large must the difference be to matter?	A statistically detectable effect can be commercially immaterial.
Variance	How noisy is the outcome under the design?	More observations help differently at different noise levels.
Power	What probability of detecting the relevant effect is needed?	The test should be sized around the effect that matters.
Observation cost	What does another observation or longer follow-up require?	Information has time, attention, and opportunity costs.
Action threshold	At which result does the recommendation change?	The threshold converts an estimate into a decision rule.
Permitted statement	What can the evidence support?	A test result is not automatically a universal marketing rule.

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

The screen keeps significance and usefulness separate. A very narrow interval around a tiny effect may not justify a change. A wider interval around a potentially important effect may justify more data, a staged decision, or a bounded test. Both are legitimate outcomes if the decision rule is explicit.

Figure 1 The advertising test-value screen

Review field	Required input	Permitted interpretation	Stop signal
Decision	Action that could change	"This test informs this choice."	The test has no stated decision.
Relevant effect	Minimum commercially meaningful difference	"This effect size matters at the stated threshold."	Any statistically detectable difference is treated as useful.
Variance and power	Outcome noise and detection target	"The design has power for the effect that matters."	Sample size is copied from another campaign.
Observation cost	Data, time, and follow-up burden	"More data is worth collecting under this cost."	Data volume is treated as free.
Action threshold	Result that changes the recommendation	"This estimate crosses or does not cross the rule."	A confidence interval is reported without a decision rule.
Claim boundary	Study, platform, horizon, and metric	"The evidence supports this bounded statement."	One experiment becomes a universal sample-size or ROI rule.

Author's decision framework grounded in Johnson, Lewis and Reiley (2017) and Wernerfelt, Tuchman, Shapiro and Moakler (2025). The thresholds and example prompts are synthetic; the reported platform findings remain bounded to their studies.

Why must statistical power focus on economically meaningful lift rather than micro-effects?

Johnson, Lewis, and Reiley study how data volume and experimental design affect power in advertising experiments. Large advertising experiments can produce more precise estimates, but the value of additional data depends on variance, effect size, and the relevant decision threshold. That is a methods point with a commercial consequence.

If the effect a manager cares about is large relative to the noise, a modest test may be informative. If the effect is small or the outcome is highly variable, a much larger test may still leave the action uncertain. The additional observations have different value in those two settings. A sample-size number detached from variance and the minimum relevant effect cannot answer the decision question.

Power also depends on the experimental design. The unit of randomization, the outcome definition, the follow-up window, and the comparison all affect what an observation contributes. A longer window can reveal a meaningful outcome, but it can also delay the action. A more granular outcome can expose heterogeneity, but it may increase noise. The design should therefore be described with the decision, not only with the number of exposed accounts or impressions.

The correct question is not “How many observations do we have?” It is “What action could this design detect well enough to change?” That wording makes the threshold visible.

How can commercial experimentation create organizational value beyond point-estimate precision?

Wernerfelt, Tuchman, Shapiro, and Moakler estimate the value of offsite conversion-optimization data in a randomized experiment involving more than 70,000 Facebook and Instagram advertisers. In their reported setting, the median cost per incremental customer rises from 38.16 w i t h p u r c ha seo pt imi z a t i o n t o49.93 with click optimization, a 31% increase. The reported result is a platform and outcome boundary. It is not a general claim that every additional conversion signal has the same value.

The source also reports that loss of offsite data disproportionately harms smaller-scale advertisers, and that purchase-optimized ads generate more long-term customers per dollar in the reported six-month follow-up. These findings make the data-value question concrete. The information changes what the platform can optimize for and changes the reported customer-acquisition outcome in that setting.

The distinction is important. More observations from a weaker signal can improve precision around that signal without making the signal more valuable. Better offsite conversion information can change the target being optimized. The two forms of improvement should not be collapsed into “more data.”

The permitted sentence is bounded: in the reported randomized experiment, offsite purchase-optimization data was associated with lower median cost per incremental customer than click optimization, and the reported loss of offsite data harmed smaller-scale advertisers disproportionately. The stronger sentence, “purchase optimization always improves advertising ROI,” is outside the source boundary.

Why do statistical significance and economic return represent divergent decision screens?

A test can pass a statistical screen and fail an economic screen. For example, an estimated difference may be precise enough to reject a null of no difference while remaining below the minimum change that justifies creative production, budget movement, or an operational switch. The result is real under the statistical model but not necessarily decision-changing.

A test can also fail a statistical screen while remaining worth continuing. If the possible effect is large enough to matter and the cost of another observation is low, more data may have positive expected decision value. The manager should state that the evidence is insufficient for the current action rule, not turn the uncertain estimate into a negative finding.

This is why the action threshold belongs beside the confidence interval. The interval describes uncertainty about the estimate under the design. The threshold describes what the organization is willing to do. They answer related but different questions.

How should marketing executives structure a synthetic budget allocation decision sequence?

Take a synthetic test comparing two optimization signals. The team defines a minimum relevant reduction in cost per incremental customer and a six-month customer-quality check. The initial sample produces an estimate whose interval crosses the action threshold. The team has three options: stop and retain the current action, collect more observations, or run a bounded follow-up that improves the outcome definition.

The choice depends on variance, cost, and the consequence of being wrong. If the follow-up is cheap and the decision is reversible, more data may be worthwhile. If the data window delays a time-sensitive action or the test is expensive, the threshold may support a staged choice. Neither option follows from sample size alone.

The example is synthetic. It does not estimate a campaign result. Its purpose is to show how a test-value screen turns a statistical result into a decision record without pretending that the record is universal.

What disciplined governance sequence should govern ad test evaluation and scaling?

Use this sequence before an experiment result enters a campaign or measurement decision:

1. Name the action that could change and whether the change is reversible.
2. Define the minimum relevant effect in the outcome and time horizon that matter.
3. Describe the randomization unit, comparison, outcome, variance, and follow-up window.
4. Choose the power target for the relevant effect, not for a generic effect copied from elsewhere.
5. Price the observation, time, data, and opportunity costs of continuing the test.
6. Write the threshold that changes the action and the rule for an inconclusive result.
7. Keep precision, practical effect, economic value, and long-term outcome as separate fields.
8. State the platform, advertiser population, metric, and horizon before generalizing.

This sequence does not make every experiment efficient. It makes the reason for collecting more data auditable. That is the relevant upgrade when a dataset is growing faster than the decision discipline around it.

Where are the empirical limits of advertising sample size and power calculations?

The two sources do not analyze a current campaign, provide a universal sample-size calculator, or show that more observations always create more economic value. Wernerfelt and colleagues report a bounded randomized experiment involving more than 70,000 Facebook and Instagram advertisers, with the stated optimization and six-month outcome boundaries. Johnson, Lewis, and Reiley provide a methods boundary about power, variance, effect size, and decision thresholds.

The stopping rule is concrete. Do not release “more advertising data will improve the decision” until the action, minimum relevant effect, variance, power, observation cost, and threshold are named. If those fields are missing, the dataset may still grow. Its decision value has not yet been shown.

The observation-volume boundary belongs beside the attribution model that needs a counterfactual and the incrementality illusion, as well as the methodological requirement that staggered difference-in-differences needs a cohort-time estimand to prevent negative weighting in multi-period policy evaluations.

Johnson, G. A., R. A. Lewis, and D. H. Reiley. (2017). When Less Is More: Data and Power in Advertising Experiments. Marketing Science, 36(1), 43-53. DOI

Wernerfelt, N., A. Tuchman, B. Shapiro, and R. Moakler. (2025). Estimating the Value of Offsite Tracking Data to Advertisers: Evidence from Meta. Marketing Science. DOI

END OF ESSAY

HOW TO CITE THIS ESSAY

Isoglu, S. (2026, September 1). More advertising observations do not guarantee better decisions. Sinan Isoglu. https://isoglu.com/insights/journal/more-advertising-observations-do-not-guarantee-better-decisions/

KEEP READING

From the research bench

What is statistical conclusion validity? A significant result can still be wrong

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

From the research bench

Conjoint analysis estimates preference, not demand

https://isoglu.com/insights/journal/conjoint-analysis-estimates-preference-not-demand/

From the research bench

A new-product forecast needs more than one method

https://isoglu.com/insights/journal/a-new-product-forecast-needs-more-than-one-method/

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

Online version: https://isoglu.com/insights/journal/more-advertising-observations-do-not-guarantee-better-decisions/