

Method Disclosure in Published Marketing Results

Summary of a Cross-Sectional Analysis of 6,998 Published Case Studies from 369 Media-Placing Firms

Trevor Birba

Dark Wave Marketing Science

28 August 2026

darkwavemarketing.science/notes/method-disclosure-in-case-studies

Summary

658 marketing services websites were crawled and parsed. The analysis population is the 369 of those firms that place media on a client's behalf, whose sites carry 6,998 published case studies and 15,662 numeric claims. 3,386 of those case studies publish at least one numeric claim.

Most firms never publish evidence of a comparison condition. 256 of 369 (69.4%) show no comparison of any kind anywhere on their site. 62 (16.8%) show testing discipline without a named design. Only 51 firms, 13.8% of the population, ever describe running a named experimental design. At the level of the individual case study, 2,941 of 3,386 published case study pages (86.9%) show no comparison condition at all.

Capability does not produce disclosure (Figure 1). Narrowing to the 51 firms demonstrably running named designs, 76.6% of their 4,820 published numeric claims appear with no mention of a method in any form. Those firms are their own control, so whatever governs which results a firm elects to publish operates identically on the pages that disclose and the pages that do not, and cancels.

Nothing else on the page substitutes. Disclosed and undisclosed results have the same median magnitude, 51% against 50%, and the within-firm ordering is indistinguishable from chance. A reader cannot use the size of a number to judge whether it was measured.

Four properties required to interpret a percentage are near-universally absent. After correction for classifier error, a numeric baseline is missing from 99.2% of claims, the source of the figure from 94.7%, the period from 84.9%, and enumerated concurrent activity from 82.4%. Absent all four, a published percentage supports no inference: the reader cannot determine what the figure is a proportion of, over what interval, alongside what else, or who produced it. Four of the five conditions this paper sets out cost a sentence and require no measurement capability at all.

Competing interests. Dark Wave Marketing Science sells measurement services. A finding that measurement is poorly disclosed is self-serving on its face. That objection is raised first because it is the first a reader should raise, and it is answered by publishing the in-

strument, its failures, and the results that ran against the authors' expectations. See the note on the full paper at the end.

1. What this study does and does not observe

This is not a study of whether published marketing claims are accurate, and it could not be. The data behind a case study is not released, and in most cases no counterfactual appears to have been constructed. Every result here concerns what a published page permits a reader to conclude, and none concerns whether a number is true.

The distinction is the finding, not a hedge. A published number carrying no method is not a weakly supported causal claim; it is not a causal claim at all. "Bookings rose 47%" and "we drove a 47% increase in bookings" are different propositions requiring different evidence, and only the second requires a method. The industry publishes the first construction and is read as having asserted the second.

Every claim in the corpus was selected for publication by the firm that produced it, placed on a marketing page, under no confidentiality constraint and with every incentive to demonstrate rigor. Every rate reported here is therefore a generous estimate. A firm choosing what to display will not lead with the engagements it can say least about. When 76.6% of the claims published by demonstrably capable firms carry no method, that is a floor on non-disclosure, not an estimate of it.

2. Method in brief

The population is firms that support client advertising outcomes, meaning those evidencing paid media execution. Firms participating in no client advertising outcome at all are excluded: trade press, directories, associations and research vendors, which enter a domain crawl because they publish on the same subject. 369 firms qualify, of which 271 classify as advertising agencies and 98 as platforms and vendors. The two groups are statistically indistinguishable on every measure reported here, so the paper treats them as one population and reports the split as a robustness check.

Method language is scored in two tiers. Strong denotes a named design with a comparison group: control or holdout group, matched market, synthetic control, geographic test, exposed and unexposed cells, brand or conversion lift study, incrementality test, or an identified third-party measurement partner. Weak denotes testing discipline without a named design: A/B and split tests, stated statistical significance, confidence intervals, stated sample sizes.

Method counts only where it was applied, not advertised. A method is scored only where an execution or inference verb sits beside it. "We ran a matched market test across six DMAs" counts. A services menu listing "Media Mix Modeling and Incrementality testing" does not. Advertised capability is not evidence that a published number was measured.

Disclosure is a property of the page. A method stated anywhere on the page carrying a claim counts as disclosed, however far from the number it sits. This is the construction most favorable to the publishing firm and it is deliberate: the effect is large enough that the most generous available standard costs nothing and forecloses the obvious objection.

3. Findings

3.1 What the population publishes

	Firms	Share
Describe running a named design	51	13.8%
Testing language, no named design	62	16.8%
No comparison of any kind	256	69.4%

The classification is a site-level binary and the threshold is a single occurrence: a firm leaves the third row if it describes a comparison even once, anywhere across all of its pages. The bar is as low as it can be set while remaining a bar. Seven firms in ten do not clear it.

The proposition that the industry does not measure is not supported. Roughly three firms in ten show evidence of some comparison. The analytically interesting group is the 13.8% for whom capability is not in question.

3.2 The principal result

Across 4,820 claims published by the 51 firms that demonstrably run named designs, 76.6% appear on a page that does not mention a method in any form. Not the design, not a comparison group, not a statistical test, not a measurement partner, not a period over which a comparison was drawn.

Widening the qualifying threshold to any applied method (113 firms, 9,045 claims) gives 78.3%. The result does not depend on where the line is drawn, and it moves in the wrong direction for a capability explanation.

It replicates. An earlier crawl of 410 sites with 39 firms in the design-running group gave 77.2%. The corpus then grew by 53% and the qualifying group by twelve firms, and the estimate moved six tenths of a percentage point.

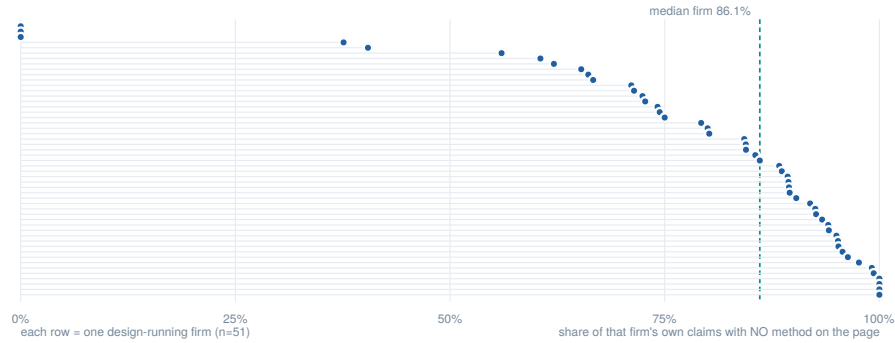


Figure 1. Non-disclosure rate for each of the 51 firms that describe running a named design, one point per firm, ordered ascending. The horizontal position is the share of that firm's own claims sitting on a page with no method. Capability is constant across every point, so the spread is variation in publishing practice alone. The median firm is at 86.1%. Four firms at the far left disclose on every claim, which establishes that the behavior is available.

Two explanations are unavailable by construction. The firm did not measure: excluded, because every firm in this comparison published a description of a design it ran. Client confidentiality: excluded, because these are marketing pages the firm chose to publish, and the method is not the confidential part. The client's data is, and the client is frequently named while the method is not.

3.3 Disclosure is a property of the firm

The distribution is not a spread (Figure 2). Among 223 firms publishing at least three case study pages, the median firm discloses a method on 0.0% of them, and 145 firms (65.0%) disclose on none.

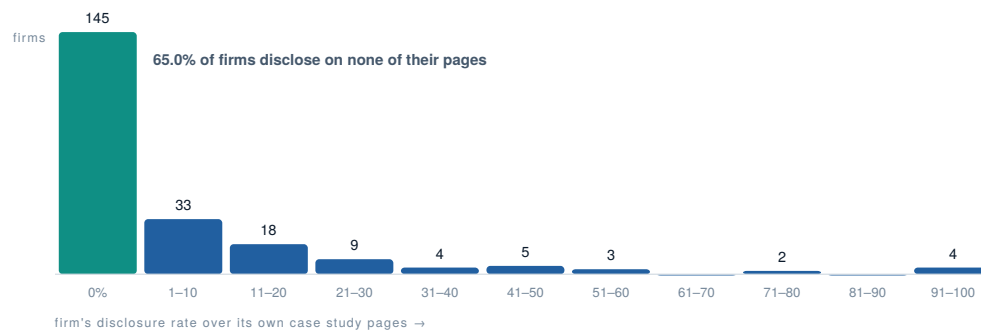


Figure 2. How the pooled rate is composed. Each bar counts firms by the share of their own case study pages that state a method, across the 223 firms publishing at least three. The distribution has no centre: 145 firms, 65.0%, sit in the leftmost bar and disclose on none of their pages. Most firms never disclose and a small number sometimes do.

Capability scales with firm size; disclosure does not (Figure 3). The share of firms that have ever described running a named design rises steadily with headcount, from 10.0% at one to ten employees to 38.5% above two hundred, which is what a fixed-cost account of

measurement capability predicts. The share of published pages stating a method shows no corresponding gradient. If disclosure were downstream of capability the two would track.

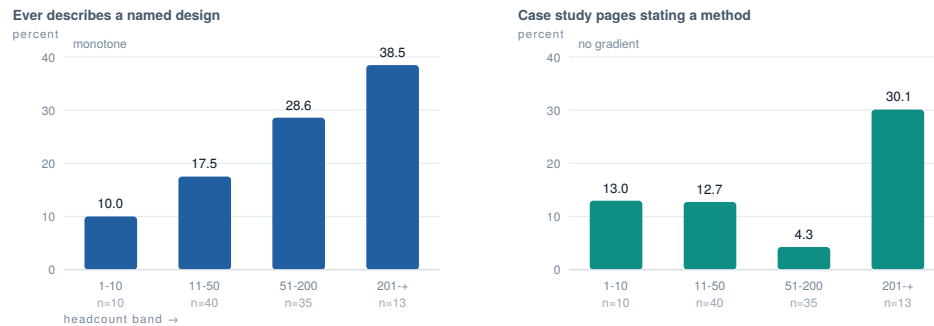


Figure 3. Capability and disclosure against firm size, on the 98 firms with a headcount. Left, the share of firms in each band that have ever described running a named design. Right, the share of those firms' case study pages stating a method. Both are percentages on the same scale and are read independently. The left panel rises with size, as a fixed-cost account of capability predicts; the right does not. If disclosure followed from capability the two would have the same shape.

Client vertical does not explain it either (Figure 4). Page-level disclosure ranges from 30.4% in CPG and retail to 2.6% in nonprofit and public, a twelve-fold spread that invites the reading that some industries oblige their agencies to be rigorous. Holding the publishing firm constant, across the 144 firms publishing in two or more verticals, every vertical's deviation from that firm's own average falls within a few points of zero, and CPG, the highest raw rate in the corpus, deviates negatively. Between-firm variation accounts for 51.9% of page-level variance in disclosure; between-vertical variation accounts for 4.8%. Verticals differ because different firms publish in them.

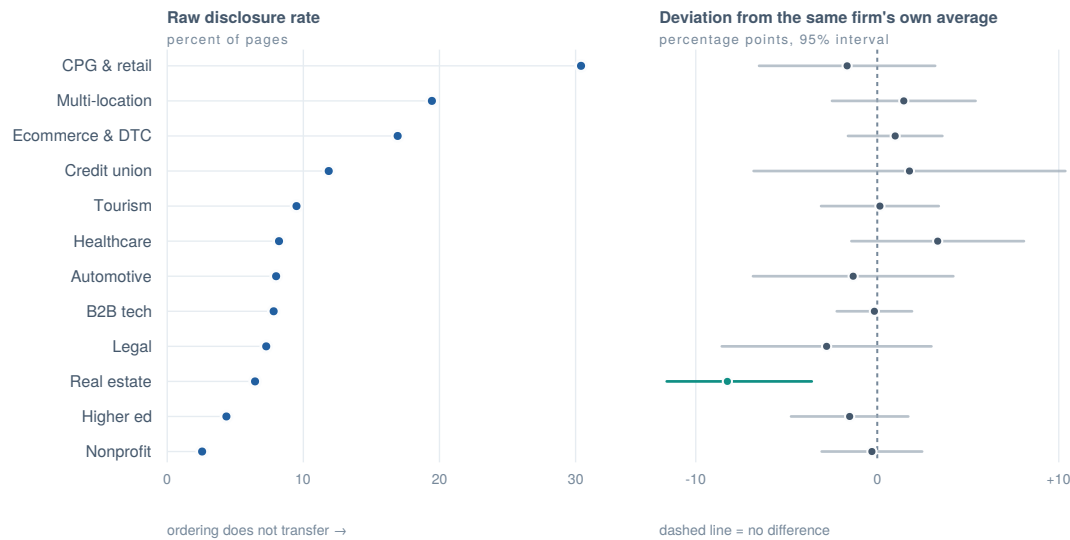


Figure 4. Whether client vertical explains disclosure. Left, the raw page-level disclosure rate per vertical, a twelve-fold spread. Right, the same verticals as deviation from the publishing firm's own average, across the 144 firms publishing in two or more verticals, with 95% intervals; the dashed line is no difference. Rows are in the same order in both panels. The left-hand ordering does not reappear on the right, and every interval except real estate, on 31 pages, crosses zero.

3.4 A rejected hypothesis

The preregistered expectation was that weaker methodological language would accompany larger claimed effects. It is not supported. Percent-unit claims from the design-running firms have median magnitudes of 51.0% on disclosing pages and 50.0% on non-disclosing ones. Within firms, among the 29 with at least five claims on each side, disclosing pages carry the lower median in 11 of 29, below chance and not at it.

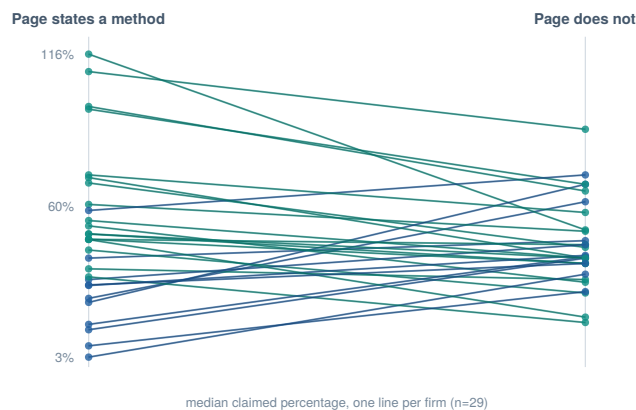


Figure 5. The magnitude comparison made inside each firm. Each line is one of the 29 design-running firms publishing at least five percent-valued claims on both sides, joining its median disclosed claim to its median undisclosed one. The hypothesis predicts every line tilts upward to the right. Navy marks the 11 firms where it does, teal the 18 where it tilts the other way. Within a firm, the presence of a method carries no relationship to the size of the number published.

Figure 5 makes the comparison inside each firm. The null is the more consequential result. Were large claims reliably the undisclosed ones, a reader could apply a protective heuristic and discount large numbers. No such heuristic exists. Claim magnitude carries no information about whether the claim was measured, which leaves disclosure itself as the only available signal.

3.5 The absence is uniform

The four interpretive properties are missing at between 91% and 99% of claims in every vertical and at every tier of methodological capability. Pages that disclose a method are no better on the other four properties than pages that do not, which establishes that method disclosure and interpretive completeness are separate behaviors, not two symptoms of one underlying diligence. There is no subgroup in this corpus in which a reader is adequately served.

This is the basis for describing the gap as one of convention. Stating what a percentage is calculated from, over what period, alongside what other activity, and from what source requires no experiment, no additional budget and no capability the firm does not already hold. Only a comparison condition requires additional work. The costless conditions are the ones missing.

4. What this study cannot say

Accuracy is not observed; no claim here is that any published number is false. Published pages are observed, not what a firm tells its client. The frame is a regionally and vertically weighted corpus assembled by search and client-side enumeration, and is not representative of the industry. Claims are selected by their publishers, which is why the within-firm comparison carries the argument and the pool-level rates do not. Automated classification carries measured error, reported per property in the full paper, and the corrected figures are lower than the raw ones because the classifier misses genuine disclosures more often than it invents them. The population boundary is a judgment, inspectable but not a validated classifier.

5. The criteria the authors apply

No standard is proposed for general adoption. These are the conditions this firm requires of a reported number before standing behind it, published on the principle that a paper about disclosure should disclose its own.

1. **A numeric baseline.** What the figure is a proportion of, in quantities a reader could check.
2. **A comparison condition.** What the result is measured against, or an explicit statement that no comparison was constructed.
3. **A stated period,** and whether it was fixed before the result was observed.

4. **Concurrent activity, enumerated**, where several activities ran together.
5. **The source of the figure**. Platform-reported, firm-calculated, or independently produced.

Four of the five cost a sentence. The second is the one requiring work.

A note on what this summary omits

This is a condensation. The full paper runs to twenty-seven pages and the difference is entirely apparatus, which a reader evaluating these figures is entitled to inspect:

- **The claim-eligibility gate**, with all 15,534 refused numbers broken out by recorded reason, because the gate is the component of the instrument most likely to be wrong and the one a reader cannot otherwise see.
- **Window sensitivity**, published as a curve. The measured disclosure rate moves by a factor of ten depending on where a character window around each claim is drawn, which is why this paper reports the page-level ceiling and not a windowed figure.
- **Inter-rater reliability** per criterion, reported as Cohen's kappa rather than raw agreement, with the two criteria that were rewritten mid-study because independent raters could not apply them consistently.
- **Every instrument failure found during the study**, including one that would have manufactured the expected result: a prospecting filter that selected on the outcome variable. Each is recorded with what it produced before correction.
- **Computational disclosure**: both AI vendors and the exact pinned model versions used for adjudication, their temperature and prompt, and an explicit statement that no model wrote, selected or interpreted any figure.
- **The operational definitions**, printed as the literal regular expressions, because a description of a pattern is not a definition of it.
- **A data and code availability statement**.

A summary that omitted the existence of that apparatus would be asking for the same trust this paper declines to extend to anyone else. The full paper is published without registration at <https://darkwavemarketing.science/notes/method-disclosure-in-case-studies>, and the instrument and derived data are available on request at trevor@darkwavemarketing.science.