The Hypothesis Comes First

A design-stage checklist for activations whose results will be published

Trevor Birba

Dark Wave Marketing Science

1 September 2026

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

The hypothesis

Every activation is a claim about cause. Written down before the money moves, in a form that can come back false, it is a hypothesis. Written down afterward, in whatever form the numbers happened to take, it is a description.

A hypothesis has three parts and all three are settled in advance.

The intervention	What will run, where, and for how long
The predicted effect	Which metric moves, in which direction, by how much
The threshold that decides	The result that would prove the plan wrong, and what happens then

The third part is the one that gets left out, and it is the one that does the work. A prediction that cannot fail is not a prediction. If no observable result would have changed the recommendation, the activation was never being measured, it was being justified.

Strong	Adding CTV in six Ohio metros will raise new-customer orders by at least 8% against eleven matched holdout markets, read over the eight-week flight plus a two-week tail. Under 8% we do not renew.
Weak	CTV will build awareness and support the funnel.

The weak version is not false. It is unfalsifiable, which is worse, because it will be reported as a success whatever happens.

The five conditions below are what make the hypothesis testable. Each is decided before launch. Four of them cost a sentence and require no additional budget, no experiment and no capability a firm running media does not already have. Only the second asks for more work.

---

1. The denominator

What is the number a proportion of, and can you read it today?

A baseline assembled after the flight is a baseline that was chosen. Fix the quantity and the source before launch, and confirm you can pull it now, at the granularity the claim will need.

Strong	New-customer orders in the six test metros, from the client's order table, running at 1,240 a week across the eight weeks before launch.
Weak	Conversions, pulled at the end from whichever system had the cleanest numbers.

In the published corpus a numeric baseline is absent from 99.2% of claims. It is the most frequently missing of the five and the cheapest to supply.

---

2. The comparison condition

What is the result measured against, and is that settled now?

A holdout, matched markets, a fixed pre-period, or an explicit statement that no comparison was constructed. All four are acceptable answers. Choosing among them after the result is visible is not.

An instrument that can only register the effect where the intervention ran cannot return a null. It validates, it does not measure. A brand study fielded only in exposed markets, or a survey asking respondents whether the advertising influenced them, will produce a number in every condition including the condition where nothing happened. That is a property of the instrument rather than a finding about the campaign.

Strong	Eleven holdout metros selected on pre-period fit and frozen on 12 September, before any campaign data was seen.
Strong	No comparison was constructed. The figure is descriptive and is reported as such.
Weak	Year over year. Same period last year. Exposed respondents asked whether the campaign moved them.

The explicit no is a passing answer under this checklist. A descriptive number labelled as descriptive costs nothing and forecloses the objection. A descriptive number presented as a causal one is the failure this document exists to prevent.

---

3. The window

What period is being read, and is it locked before launch?

Fix the read window in advance, including how far past the end of the flight it extends. A window selected once the data is visible is a result selected once the data is visible.

Strong	Weeks one through eight of the flight plus a two-week tail, fixed at kickoff. The tail is included because carryover in this channel decays inside ten days.
Weak	During the campaign period. Q3. The period in which the effect appeared.

The period is absent from 84.9% of published claims in the corpus.

---

4. Concurrent activity

What else runs inside that window, and what happens if it lands in a test cell?

List it now: promotions, price changes, PR, a competitor launch, an email push, a store opening. Then decide in advance what happens when one of them appears in a test market, because deciding afterward is deciding with the result in view.

Strong	Four items enumerated at kickoff, with a rule: any test metro carrying the October promotion is dropped from the read, decided before the promotion calendar was final.
Weak	Other factors were taken into account.

Enumerated concurrent activity is absent from 82.4% of published claims.

---

5. The source of the figure

Who produces the number?

Platform-reported, calculated by the agency, or produced independently. Name which. A firm that buys the media and also reports the lift is grading its own homework, and saying so plainly costs less than being asked.

Strong	Calculated by us from the client's order table. The platform's own conversion count is reported alongside and differs by 31%, for the reasons in the appendix.

Weak	A number with no provenance. Two numbers from two systems, reconciled silently.

The source of the figure is absent from 94.7% of published claims.

The tie-out

The hypothesis is the header and the five conditions are what make it answerable. Settled before launch, they produce a result that can be written up in one pass, because everything the writeup needs was decided when the plan was.

Settled afterward, they cannot be recovered. There is no way to reconstruct a baseline that was never pulled, a holdout that was never held, or a window that was never fixed. The work either happened at the design stage or the claim is descriptive, and the honest move then is to say so.

Requests, questions and the full paper: trevor@darkwavemarketing.science