

[MTH-001] VERSION 1.0

A Reproducible Standard for GEO Visibility

A public framework for measuring AI appearance rate and recommendation rank before and after a GEO engagement.

MATCHED EVALUATION DESIGN

PRE / POST

Independent measurement. Public formulae. Controlled replication evidence.

PUBLIC METHODOLOGY

STATIC PDF / NO ACCOUNT REQUIRED

[02] EXECUTIVE READOUT

What should a credible GEO effectiveness measurement disclose?

A defensible evaluation fixes the sampling rules before observation, measures the same estimands before and after the engagement, repeats unstable model runs and reports uncertainty alongside every effect estimate.

01	Define the estimand	State exactly what appearance and recommendation rank mean.
02	Match the observations	Use the same sampling logic and intent mix before and after service.
03	Report uncertainty	A point estimate without a sampling interval is not sufficient evidence.

[03] TEST DESIGN

The evidence chain.

Every conclusion starts with a defined claim, matched observations and a statistical rule selected before the outcome is known.

01

APPEARANCE RATE

AR = total brand appearances / eligible observations

02

RANK SCORE

RS = mean of $1 - (\text{rank} - 1) / K$; missing brand scores 0

03

EFFECT

Delta = post-engagement estimate - baseline estimate

04

INFERENCE

Matched resampling with repeated model observations

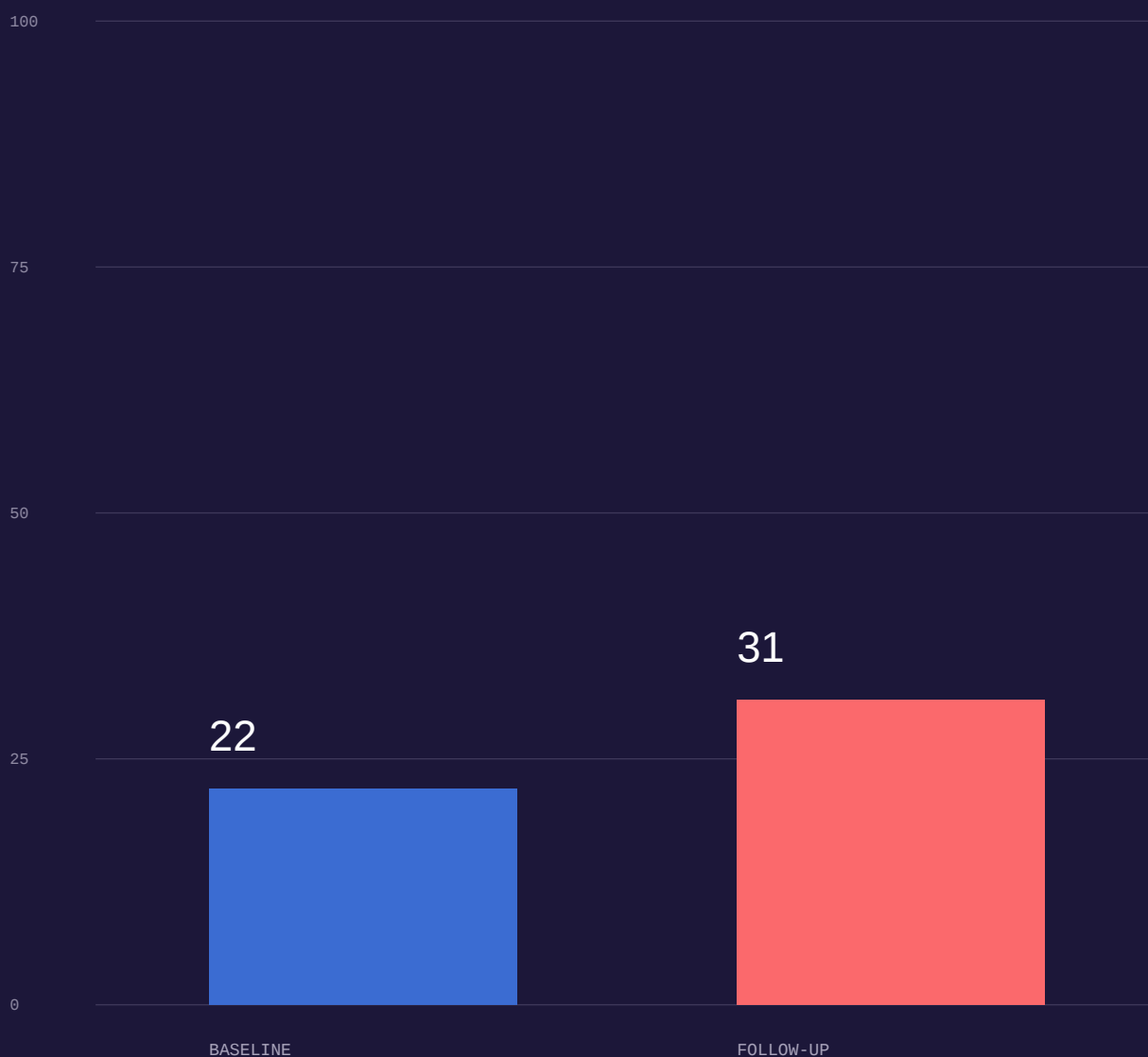
CONTROL

The query sampling rule and decision threshold are fixed before the follow-up result is observed.

[04] OBSERVED PATTERN

Illustrative matched pre/post estimate

% APPEARANCE RATE



DELTA = +9 PERCENTAGE POINTS

[05] STATISTICAL INTERPRETATION

EVIDENCE VERDICT

ESTIMATE + UNCERTAINTY

The effect is the matched change between periods. A production report must also show its confidence interval, observation count, missingness and sensitivity to reasonable query re-sampling.

WHAT THIS RESULT DOES NOT SAY

The public method does not require public release of customer queries or complete raw model outputs. Those items remain controlled, but must be preserved well enough for a paying client or authorised reviewer to replicate the analysis.

DECISION RULE

Interpret the pre-specified estimate, interval and test together. Do not convert statistical incompatibility into a stronger factual allegation than the evidence supports.

What can be concluded.

The Dragonfly Research publishes metric definitions, formulae, thresholds, assumptions, limitations and replication principles. Query samples, raw responses, timestamps, model versions and analysis files form the controlled evidence pack.

ALWAYS PUBLIC

- 01

Metric definitions and formulae
- 02

Decision thresholds and assumptions
- 03

Limitations and replication principles

CONTROLLED EVIDENCE PACK

- 01

Queries and raw model responses
- 02

Versions, timestamps and repeat logs
- 03

Analysis files for replication

THE DRAGONFLY RESEARCH

Your GEO provider should not grade its own work.

THE GEO WATCH