

[INV-001] AUGUST 2026

Pxxxx Case Study: Could These Results Occur by Chance?

A worked statistical examination of a provider case-study claim using synthetic data and a pre-specified null model.

REPLICATION GAP TEST

$$p < .001$$

Independent measurement. Public formulae. Controlled replication evidence.

INVESTIGATION - DEMONSTRATION

STATIC PDF / NO ACCOUNT REQUIRED

Is the published outcome compatible with an independent replication?

In this synthetic example, a claimed 80% appearance rate is compared with an independent replication probability of 10%. Seeing 32 or more appearances in 40 matched observations under that null is extraordinarily unlikely.

01	The claim is testable	The provider states a rate, a time window and a measurable outcome.
----	------------------------------	---

02	The replication does not agree	Matched observations produce a materially lower appearance probability.
----	---------------------------------------	---

03	The claim requires evidence	The gap is too large to dismiss as ordinary sampling noise under the stated model.
----	------------------------------------	--

[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

CLAIM

80% AI appearance rate after the engagement

02

REPLICATION

40 matched observations under controlled conditions

03

NULL

10% appearance probability from independent runs

04

TEST

Exact one-sided binomial tail probability

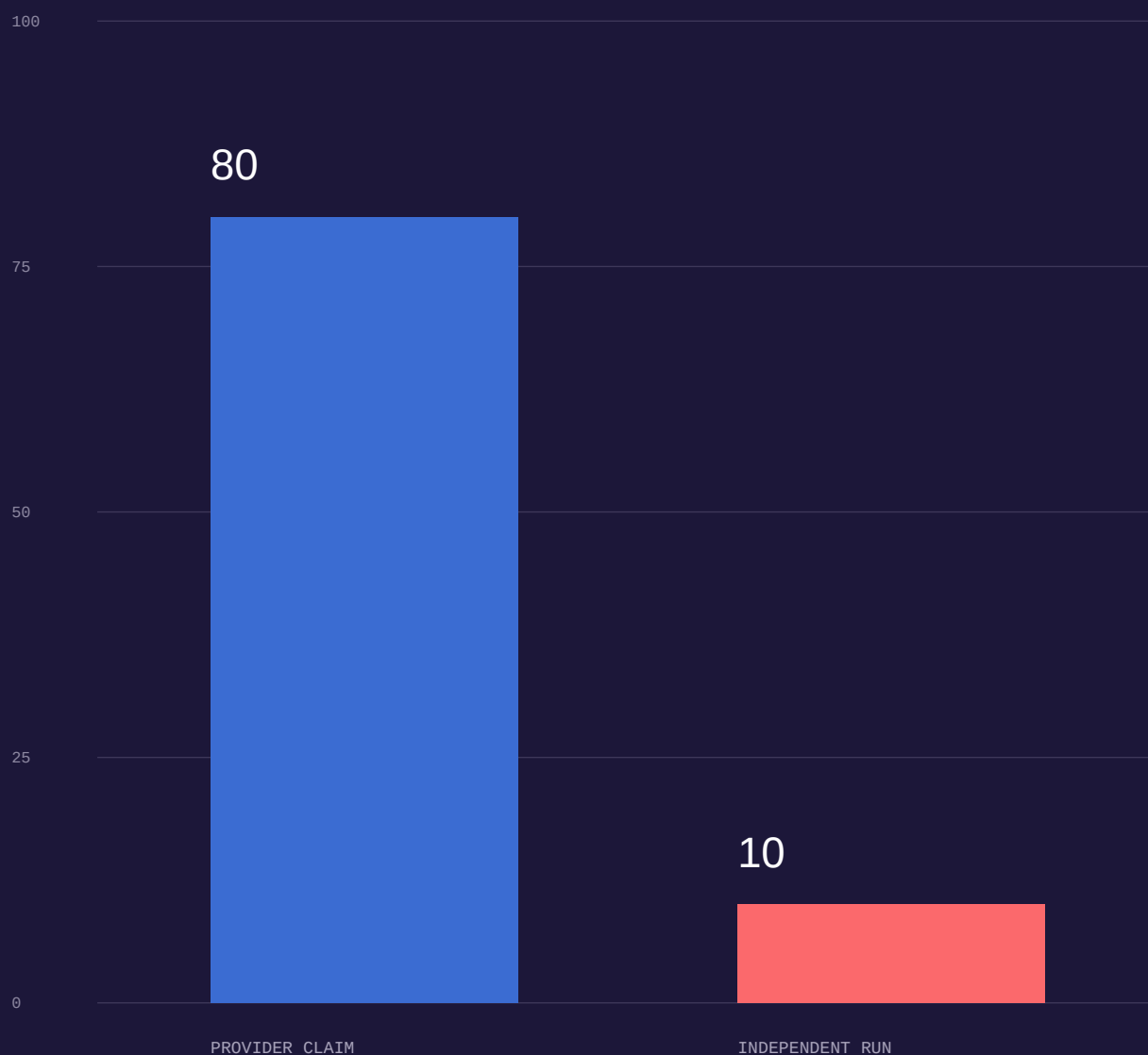
CONTROL

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

[04] OBSERVED PATTERN

Claimed rate versus independent replication

% APPEARANCE RATE

 $P = 3.40E-25$

[05] STATISTICAL INTERPRETATION

EVIDENCE VERDICT

NOT REPRODUCIBLE

Under this synthetic null model, the published-style result and the independent replication cannot reasonably be treated as measurements of the same stable process.

WHAT THIS RESULT DOES NOT SAY

A small p-value is not, by itself, proof of fabrication. It establishes incompatibility with the specified model and creates a need for the provider's raw evidence, query selection rules and run logs.

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 correct public conclusion is not that fraud has been proven. It is that the claim has failed an independent compatibility test and should not be relied on without a complete replication package.

ALWAYS PUBLIC		CONTROLLED EVIDENCE PACK	
01	Metric definitions and formulae	01	Queries and raw model responses
02	Decision thresholds and assumptions	02	Versions, timestamps and repeat logs
03	Limitations and replication principles	03	Analysis files for replication

THE DRAGONFLY RESEARCH

Your GEO provider should not grade its own work.

THE GEO WATCH