Two rivals. One discriminating measurement.

HYPOTHETICAL CACHING EXAMPLE / No measurements or TRACE performance results are shown.

Report A

“Caching lowers latency.”

Report B

“Caching raises latency.”

Same offered workload: compare median latency AND score stale-response fraction ≤ 1%.

Rival H_1 · Reuse / write-fraction interaction

Read reuse and write fraction change the effect.
Illustrative toy rule: $\Delta L = 3 - 30r + 10u$ ms
At matched external conditions: more reuse → lower ΔL.

Rival H_2 · External background contention

Scheduled background work explains the contrast.
At matched background contention, reuse may have little effect on latency.

LOCK THE CONTRAST BEFORE MEASURING

Vary reuse; match write fraction, arrivals and scheduled background work.
Lock predictions; randomize cache-on/off; score latency AND freshness.

Possible outcomes

H_1 contradicted / H_2 contradicted / both compatible / insufficient evidence

r: repeated-key fraction of reads; u: write/invalidation fraction of all offered requests ($u < 1$).