

Why Did the Result Disappoint?

Turn a weak or uncertain finding into a disciplined question about what happened next

FOR

program teams, executives, boards, funders, evaluators, and partners responding to results that did not meet expectations

USE WHEN

a result is flat, negative, mixed, imprecise, or being explained differently by different people

INCLUDES

a competing-explanations map, worked example, response test, action ladder, and a 15-minute team exercise

The result arrives and it is not what anyone hoped. One person says the program was implemented poorly. Another says the measure missed the real change. Someone else points to the economy, the referral system, or the people who never answered follow-up. Each explanation may be plausible. None becomes true because it is reassuring.

A disappointing result is evidence that an expectation was not met under the conditions observed. It does not explain why.

The short answer

Separate the result from the explanations offered for it. Test delivery, measurement, participation, comparison, precision, context, and program logic as competing possibilities. Choose the next action that the available record supports, not the explanation the organization prefers.

A weak result is a question about the program, the evidence, and the conditions around both. It is not an automatic verdict on the people doing the work.

Build a competing-explanations map

Start with the result stated plainly. Then consider seven different explanations.

Delivery differed from the model

The service may not have reached the intended people, dose, quality, or timing. Staffing gaps, referral breakdowns, waiting lists, or local adaptation may mean the program evaluated was not the program described on paper.

The measure missed the intended change

An indicator can be reliable and still capture the wrong thing. Employment at exit may miss whether work lasts. A general wellbeing scale may be too distant from the change a brief program can reasonably produce.

The program pathway did not hold

The activities may have occurred as intended while the assumed link to the outcome was weak. Training can improve knowledge without changing hiring conditions. Navigation can improve access without creating available housing.

The result is too imprecise

A small or variable sample may produce an estimate compatible with both meaningful benefit and little benefit. “Not statistically significant” does not necessarily mean no meaningful effect occurred.

The observed group is incomplete

People who completed or responded may differ from those who left or could not be reached. A favorable or unfavorable rate can shift when the visible population changes.

The comparison does not answer the intended question

Before-and-after change may reflect wider trends. Comparison groups may have started differently. A target may have been aspirational rather than evidence-based.

Conditions changed

Policy, labor markets, service availability, public health, staffing, or participant needs may have changed during delivery. Context is not an excuse. It is a possible part of the explanation that should be examined.

A worked example

A housing-navigation program expected 60 percent of enrolled households to obtain stable housing within six months. The reported result was 34 percent.

The first meeting produces two explanations: staff did not follow the model, and the local housing market made the target impossible. The team pauses and maps the evidence.

Delivery records show that navigator contact remained high, but the average time from referral to first appointment doubled during a staffing transition. Participant interviews describe strong support after contact but difficulty reaching the program at the beginning. Rental-vacancy data show a substantial decline. Follow-up is available for most households, so missing outcomes are less likely to explain the result.

The evidence does not support a single cause. It does support two next moves: repair early access and revisit what the program can reasonably promise under current housing conditions. Ending the service because it “failed” would outrun the evidence. Claiming the market explains everything would do the same.

Distinguish diagnosis from defense

A diagnosis asks what evidence would separate plausible explanations. A defense chooses a preferred explanation and collects supporting anecdotes.

Use these checks:

- Did the explanation exist before the result, or appear only afterward?
- What evidence would count against it?
- Does it imply a feasible action?
- Can the current record distinguish it from alternatives?
- Who has information that is missing from the room?

If an explanation cannot be tested and changes no action, treat it as a hypothesis rather than a conclusion.

Choose a response that fits the explanation

If delivery differed, repair the operating conditions and make the next version traceable.

If the measure was poorly matched, improve the measure without erasing the disappointing result already observed.

If the program pathway did not hold, redesign the model or revise the outcome it can reasonably pursue.

If precision is the problem, decide whether a larger or longer study is worth the delay and burden.

If missing people are the problem, improve follow-up and narrow the claim.

If context changed, decide whether the program should adapt, whether expectations should change, and which conditions must be monitored.

Different diagnoses lead to different work. “Collect more data” is not a diagnosis.

What you can say

- “The program did not meet the stated outcome under the conditions observed.”
- “Delivery changed in these documented ways during the period.”
- “The current result cannot distinguish between these two explanations.”
- “We are changing the intake process because the record identifies an early access problem.”

What you cannot say yet

- “The program does not work,” when the design, delivery, comparison, or precision cannot support that conclusion.
- “The measure was wrong,” simply because the result was unfavorable.
- “External conditions caused the result,” without evidence that connects those conditions to the outcome.
- “More time will fix it,” without specifying what should change and what will be reviewed.

What to do next

Minimum useful step: State the result without explanation. List the three most plausible explanations and what evidence currently supports each.

Stronger step: Reconstruct delivery, measurement, participation, context, and comparison. Identify the smallest check that could separate the leading explanations.

Before a consequential response: Have the analysis and competing explanations reviewed by someone free to reach an unfavorable conclusion. Agree how the result will affect continuation, redesign, or transition.

A 15-minute team exercise

Write the disappointing result at the top of a shared page.

1. Give everyone two minutes to write possible explanations silently.
2. Group the ideas under delivery, measurement, logic, precision, missing people, comparison, and context.
3. For each explanation, name one fact that supports it and one fact that would count against it.
4. Mark explanations that appeared only after the result was known.
5. Choose one or two explanations the current record can examine.
6. Name the action each explanation would imply.
7. Record what must remain unresolved.

The aim is not immediate consensus. It is to turn disagreement into questions that evidence can actually help answer.

A note on blame

Results can become judgments about staff, participants, or grantees. That response makes honest learning less likely. People protect themselves when every unfavorable number threatens funding, reputation, or employment.

Avoiding blame does not mean avoiding accountability. It means locating the claim. Was the expected service delivered? Was the outcome appropriate? Were constraints known? Were problems reported? What responsibility sat with the program, funder, partner system, or wider environment?

Be generous toward people and exacting toward the explanations attached to the result.

When to bring in help

Outside support is useful when the result affects a major continuation decision, internal incentives make one explanation especially convenient, the analysis is technically uncertain, or funder and provider need a shared account without turning the conversation into a defense.

The useful output is not a more persuasive story. It is a bounded diagnosis: what the result establishes, which explanations remain credible, and what next check or action is proportionate.

Questions to carry forward

- What exactly disappointed relative to what prior expectation?
- Which explanations can the current record distinguish?
- What evidence would count against the preferred explanation?
- What response follows from the explanation the evidence actually supports?