FOR

executive directors, program directors, public-agency managers, county administrators, board members, and anyone deciding how much rigor a decision warrants

INCLUDES

a proportionality framework, a stakes-and-evidence decision table, a two-column say/cannot-say table, and a 15-minute team exercise

How Much Evidence Is Enough?

A practical framework for deciding when to keep investigating and when to act on what you have

Somewhere in your organization, right now, two reasonable people are having a disagreement that sounds like this:

One says: “We do not have enough evidence to make this decision. We need more data before we move forward.” The other says: “We have been studying this for a year. At some point we have to act.”

Both of them may be right. Additional evidence is worthwhile when its value exceeds the cost of waiting, and that judgment depends on what is at stake.

This is one of the most common practical questions in program work, and it rarely gets a clear answer. People fall back on instinct, habit, or whoever argues most persuasively in the meeting.

The short answer

There is no universal standard for “enough evidence.” The right standard depends on the stakes of the decision, the reversibility of the choice, the risk of harm, and the cost of waiting.

Low-stakes, reversible decisions can proceed with lighter evidence. High-stakes, hard-to-reverse decisions deserve stronger evidence. The appropriate standard matches the evidence to what is at stake.

If you find yourself unable to decide how much evidence is enough, the problem is usually that you have not clearly defined the stakes. Once the stakes are on the table, the answer often becomes more obvious than you expected.

Why this is harder than it looks

The difficulty is that “more evidence” always sounds responsible. Nobody gets criticized for asking for another study. The person who says “let us wait and learn more” sounds cautious and scientific. The person who says “we know enough — let us act” sounds impatient.

But caution has costs, too. While you wait for more data, the current approach continues. If the current approach is not working well, every month of delay is a month of worse outcomes for the people you serve. If you are considering expanding a program that is working, every month you spend studying instead of expanding is a month that additional families or individuals go without the service.

There is also a subtler problem: asking for more evidence can be a way of avoiding a difficult decision. If the evidence might point toward an uncomfortable conclusion — cutting a popular program, acknowledging that a major investment is not working — the call for “more data” can function as a postponement rather than a genuine need.

The practical question is: “Is the value of getting more evidence worth the cost of waiting?”

The concept: a proportionality framework

SIX FACTORS THAT DETERMINE HOW MUCH EVIDENCE IS ENOUGH

When your team disagrees about whether you have enough evidence to act, walk through these six factors together. They provide a structured way to make the implicit reasoning explicit.

1. Stakes. How much money, how many people, and how serious are the potential consequences? Adjusting the schedule for a weekly workshop is a small-stakes decision. Defunding a statewide housing program that serves ten thousand families is an enormous one. The larger the stakes, the more evidence you should expect before acting.

2. Reversibility. Can the decision be changed later if it turns out to be wrong? Trying a new group-session format is easily reversed — if participants do not respond well, you change it back next month. Closing a program site is much harder to undo. When a decision is reversible, lighter evidence is often sufficient because mistakes can be corrected. When it is not, you should be more confident before proceeding.

3. Risk of harm. Could a wrong decision hurt vulnerable people? This factor deserves its own weight, separate from stakes. A workforce program that switches to a new curriculum and gets disappointing results can switch back. But if participants lose access to a service during the transition — if they miss a critical enrollment window or lose a housing placement — the harm may not be reversible for them, even if it is reversible for the organization.

4. Public scrutiny. Will the decision be challenged by legislators, media, or the public? Decisions that face scrutiny need evidence that can withstand questioning. An internal program adjustment can rely on staff observations. A public claim submitted to a legislative committee needs stronger backing. This is not about appearances — it is about being able to defend the reasoning honestly.

5. Uncertainty. How much is currently unknown, and how much does the unknown matter? When uncertainty is high on a question that matters, more evidence has greater value. When you already know the essential facts and the remaining questions are secondary, additional evidence may not change the decision.

6. Cost of waiting. What happens if the decision is delayed while more evidence is gathered? If waiting has no meaningful consequence, investing in more evidence is relatively cheap. If waiting means families go without services, a funding window closes, or a promising pilot loses momentum, the cost of delay is real and should be weighed against the cost of acting on incomplete evidence.

A DECISION TABLE: STAKES VERSUS EVIDENCE STANDARD

This table suggests what is proportionate. It is a guide for conversation, not a rulebook.

DECISION CHARACTERISTICS	EVIDENCE STANDARD	WHAT THAT LOOKS LIKE IN PRACTICE
Low stakes, easily reversed, minimal harm risk	Descriptive evidence is usually sufficient	Staff observations, participation records, brief participant feedback. Act, observe, and adjust.
Moderate stakes, somewhat reversible, limited public exposure	Descriptive evidence with some outcome tracking	Track outcomes for participants, look for patterns, and compare with a baseline when available.
Significant stakes, partially reversible, some public scrutiny	Correlational or quasi-experimental evidence	Compare outcomes across groups, control for obvious differences, and state clearly what the analysis can support.
High stakes, hard to reverse, high public scrutiny, risk of harm	Strong quasi-experimental or experimental evidence	Use comparison groups, address selection bias, and build a design that can withstand challenge.

DECISION CHARACTERISTICS	EVIDENCE STANDARD	WHAT THAT LOOKS LIKE IN PRACTICE
Any stakes, but the cost of waiting is very high	The best evidence available now, with a plan for more	Use what you have, be transparent about its limits, and build a learning plan for the next decision cycle.

TWO CONCRETE EXAMPLES

Example 1: Changing a group-session format. A youth-mentoring nonprofit is considering whether to switch from weekly one-hour group sessions to biweekly two-hour sessions. Staff believe the longer format allows deeper discussion. Some participants have said the current sessions feel rushed.

This is a low-stakes, easily reversed decision. If the new format does not work, the program can switch back in a month. The risk of harm is minimal. No external audience is scrutinizing the choice.

What evidence is enough? Descriptive evidence: staff observations from the pilot sessions, attendance data comparing the two formats, and informal participant feedback. The organization does not need a quasi-experimental comparison of the two formats. It needs to try the change, pay attention, and decide whether it is working.

Example 2: Defunding a statewide program. A state agency is considering whether to end funding for a community health worker program that serves ten thousand people across fifteen counties. The program costs $8 million per year. A legislative committee has questioned whether the program is producing measurable health improvements or simply providing a popular service.

This is a high-stakes, hard-to-reverse decision. If the program is defunded and the decision turns out to be wrong, rebuilding the workforce and reestablishing community relationships will take years. Thousands of people would lose access during the gap.

What evidence is enough? Descriptive data alone is not sufficient. The agency needs evidence that addresses the most likely alternative explanation — that participants who have better health outcomes would have had them anyway. A quasi-experimental design comparing participants with similar non-participants is warranted. The cost of this evidence is proportionate to the $8 million annual decision.

Neither organization is doing evidence better or worse than the other. They are facing different decisions with different stakes.

THE DANGER OF BOTH EXTREMES

Too much evidence (paralysis). Some organizations and oversight bodies demand a level of evidence that the decision does not warrant. They insist on quasi-experimental designs for operational questions. They require multi-year evaluations before allowing a pilot to expand. They ask for randomized trials when the program serves sixty people.

The costs include time lost, staff burned out on data collection that serves no decision, participants burdened with surveys and follow-up calls that do not improve their service, and good programs stuck in pilot status for years because the evidence bar was set higher than the decision required.

Too little evidence (recklessness). Other organizations act on almost no evidence at all. They expand programs statewide based on testimonials. They claim impact based on outputs. They make large commitments based on a feeling that things are working, without checking whether the data supports that feeling.

The cost falls on the people served. When a program is expanded without evidence and turns out not to work, communities lose years and resources. When a claim of effectiveness is later challenged, the program's credibility may be damaged beyond repair — even if it was actually doing good work.

Matching the evidence to the decision requires an explicit account of the stakes.

What you can say and what you cannot say

WHAT YOU CAN SAY	WHAT YOU CANNOT SAY YET
"For this decision, descriptive evidence is proportionate to the stakes."	"We do not have enough evidence" (without specifying: enough for what decision?).
"The stakes of this decision warrant stronger evidence before we act."	"We always need more data" (as a way to postpone a difficult choice).
"We have enough evidence to proceed, and here is what it supports."	"The evidence proves we should do this" (if the evidence supports a weaker claim than proof).
"We are choosing to act now because the cost of waiting is greater than the cost of being wrong."	"We are confident this will work" (if you are choosing to act despite uncertainty, say so).
"This is a reversible decision. We will try it, track what happens, and adjust."	"We will evaluate later" (if there is no specific plan for what you will track and when you will review it).

A 15-minute team exercise: the "enough" conversation

Choose one decision your organization is currently facing – ideally one where there is internal disagreement about whether you know enough to act.

Write the decision at the top of a board or shared document. Then work through these questions together:

1. What is the decision, stated as a specific question?
2. On a scale from low to high, how would you rate the stakes? (Consider money, number of people affected, and seriousness of consequences.)
3. If this decision turns out to be wrong, how easy is it to reverse?
4. Could a wrong decision cause direct harm to participants or community members?
5. Will this decision face public scrutiny from legislators, media, oversight bodies, or the community?
6. What is the cost of waiting another three months, six months, or a year for more evidence?
7. Given your answers above, what grade of evidence is proportionate to this decision?
8. Does the evidence you already have meet that standard?
9. If not, is the gap worth closing — or should you proceed with what you have and be transparent about its limits?

If the team agrees that the evidence is already proportionate, say so clearly. If the team agrees that more evidence is needed, name exactly what is missing and set a deadline that comes before the decision closes.

Requests for more evidence can postpone a difficult decision

Some calls for more evidence arise from the wish to avoid a choice that will make someone unhappy.

Asking for another study can be easier than telling a program director that the numbers do not support expansion. Requesting a longer evaluation timeline can be easier than acknowledging that a program is not producing the results it promised. Saying "we need more data" can be easier than saying "the data we have is uncomfortable, and we need to talk about what it means."

This is human and understandable. But it shifts the cost to the people waiting for a decision. If a program is not working and the decision to change it is postponed for another round of data collection, participants spend another year in a service that is not helping them. If a program is working and expansion is postponed, the people who would have been served spend another year without it.

Being honest about how much evidence is enough includes being honest about whether the request for more evidence is genuine — or whether it is a way of putting off a conversation that needs to happen now.

When to bring in help

Outside support may be useful when:

- your team cannot agree on whether the current evidence is sufficient, and the disagreement is stalling a decision;
- an oversight body or policymaker is demanding a level of evidence that seems disproportionate to the decision, and you need help making the case for a better fit;
- the decision involves high stakes, limited reversibility, and real risk of harm to participants, and you want an independent assessment of what the evidence supports;

- you need to design a phased evidence plan — one that provides enough information for an immediate decision while building toward stronger evidence for future decisions;
- the cost-of-waiting question is politically sensitive, and an outside perspective would help the team have an honest conversation about it.

In those situations, an adviser can help your team walk through the proportionality framework so the standard for "enough" is explicit, shared, and defensible.