THE INITIATIVE INTELLIGENCE GUIDE

A FRAMEWORK FOR EVALUATING THE IMPACT OF HEALTHCARE WORKFORCE INITIATIVES

CO-AUTHORS

AdventHealth

THE COALITION FOR PHYSICIAN & APP WELL-BEING

American Hospital Association™
Advancing Health in America

ATALAN

Providence | WELL BEING TRUST

The Importance of Measurement Discipline For Healthcare Workforce Wellbeing

In healthcare, we treat clinical rigor as an essential standard. We meticulously track prescribing, readmission rates, and surgical outcomes because we know that precise data saves lives. As we deeply commit to the wellbeing of the health workers delivering that care, our industry is undergoing a natural evolution—moving from celebrating the initial, heartfelt launch of wellbeing programs toward establishing sustainable, objective benchmarks. It is time for a vital paradigm shift: **the exact same rigor, data discipline, and systemic importance we bring to patient outcomes must be applied to workforce outcomes**.

This is where disciplined measurement becomes indispensable. Professional wellbeing is not a soft, secondary initiative; it is a core operational priority. We put a definitive stake in the ground on this issue in the *Impact Wellbeing™* guide, which we proudly partnered on with the CDC's National Institute for Occupational Safety and Health (NIOSH). Specifically, we highlighted that tracking and accelerating organizational maturity requires structured, objective metrics, similar to quality improvement methodology.[1]

In today's complex environment, precise measurement is critical to accurately allocate increasingly scarce resources and ensure that our organizational strategies are truly upholding our foundational oath to "first, do no harm." Increasingly, robust data is also becoming necessary to capture essential healthcare reimbursements. But beyond the financial spreadsheets and operational metrics, we must look at what is truly on the line.

In wellbeing work, we are protecting the very beating heart of healthcare: the people who have dedicated their lives to the healing of others. When they break, the entire system breaks. *The Initiative Intelligence Guide* offers the structured, decision-oriented framework required to move from well-meaning activity to proven operational impact. I urge everyone in healthcare to use this framework not solely to evaluate programs, but to safeguard the human beings who sustain our entire care system.

Stefanie Simmons, MD, FACEP
Chief Medical Officer
Dr. Lorna Breen Heroes' Foundation

DR. LORNA BREEN
HEROES' FOUNDATION

[1] *NIOSH [2024]. Impact Wellbeing™ Guide: Taking action to improve healthcare worker wellbeing. Washington, DC: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication 2024-109, (Revised 07/24)*

https://doi.org/10.26616/NIOSHPUB2024109revised072024

Table of Contents

THE PURPOSE OF THIS GUIDE

Healthcare systems are operating in increasingly complex environments. Workforce strain, operational pressure, financial constraints, changing technologies, shifting patient needs, and growing expectations around staff support have all increased the number and urgency of organizational initiatives.

Across health systems, leaders describe a familiar pattern: an urgent problem emerges, a thoughtful initiative is launched, and early feedback feels promising. But when it comes time to decide what happens next, the evidence is often incomplete.

Healthcare leaders are launching a wide range of efforts intended to improve day-to-day work and strengthen the workforce, including:

- Well-being programs
- Leadership trainings
- Coaching and support services
- Retention efforts
- Workflow and scheduling changes
- Digital tools
- Documentation changes
- New policies and operational pilots

These initiatives are introduced for good reasons. While the approaches in this guide can be applied to many types of organizational initiatives, the primary focus is on workforce and well-being efforts, which often present unique evaluation challenges. Unlike operational or financial initiatives, where outcomes may be easier to measure using traditional metrics, well-being initiatives frequently involve more subjective experiences, longer timelines, and complex organizational influences. As a result, organizations may struggle to demonstrate value using conventional evaluation approaches alone, even when staff experience meaningful benefit. Leaders are trying to solve real problems, respond to urgent needs, and improve outcomes for staff, patients, and the organization as a whole.

During the COVID-19 pandemic, for example, organizations moved quickly to support their workforce. Arpan Waghray, M.D., CEO of Providence's Well Being Trust, described launching programs to help ensure the 120,000 healthcare workers throughout Providence had seamless access to mental health support and proactive outreach through initiatives like No One Cares Alone. These efforts were widely valued, and staff expressed deep appreciation. But when leadership asked a simple question – did it work? – the answer was much harder to define.

"We had incredible anecdotal feedback," Dr. Waghray explained. "People told us it was invalable. But we didn't have a clear way to measure the overall impact on burnout or engagement at the system level."

> "We had incredible anecdotal feedback. People told us it was invaluable. But we didn't have a clear way to measure the overall impact on burnout or engagement at the system level." – Arpan Waghray, M.D., CEO of Providence's Well Being Trust

Indeed, in complex systems, even well-intentioned initiatives do not always work in the same way across settings, teams, roles, or individuals. An initiative that is valuable in one context may be less useful in another. Some may show early promise but need refinement. Others may benefit certain groups more than others. Still others may create hidden burdens that are not immediately visible.

Elisa Arespacochaga, Group Vice President of Clinical Affairs and Workforce at the American Hospital Association, has seen this challenge play out across health systems. She noted that organizations can invest in initiatives like workforce well-being with good intention, but struggle to define success in concrete, measurable terms, especially for outcomes that are inherently subjective. "Well-being is one of the hardest things to measure," she explained, pointing out that even widely touted interventions can show different impacts across individuals and teams.

Similarly, Herb Schumm, M.D., former Chief Medical Officer at Bon Secours Mercy Health and board member of the Coalition for Physician & APP Well-Being, described initiatives that appeared successful on the surface, such as leadership development programs or expanded support services, but only became truly actionable when organizations could connect them to measurable outcomes like engagement, retention, or utilization. In some cases, the absence of consistent measurement over time made it difficult to compare results or sustain momentum.

DeAnna Santana-Cebollero, Ph.D., Executive Director of Physician Well-Being & Engagement at AdventHealth, described a similar challenge when scaling a system-wide initiative called Finding Meaning in Medicine, a structured storytelling-based program designed to foster connection and reflection among clinicians. While the program was widely valued by those who participated, sustaining engagement across a large, distributed workforce proved difficult, especially without clear data on outcomes. "We were tracking participation," Dr. Santana-Cebollero noted, "but not whether it was actually reducing burnout or improving engagement. Without that, it was hard to make the case for continued investment."

As a result, organizations are often left with a familiar set of questions:

- Is this initiative helping?
- How do we know?
- Who is it helping or unintentionally harming?
- Is it helping enough to justify continued investment?
- Should we continue it, improve it, scale it, target it differently, or stop it?

> "We were tracking participation. But not whether it was actually reducing burnout or improving engagement. Without that, it was hard to make the case for continued investment."
> – DeAnna Santana-Cebollero, PhD, Executive Director of Physician Well-Being & Engagement at AdventHealth

These questions are harder to answer than they may first appear.

Health system leaders agree: there is no shortage of activity. The challenge is knowing what is actually making a difference, and being able to demonstrate it in a way that supports real decisions.

Evaluation is not one single thing. Different decisions require different kinds of evidence. In some cases, leaders need rich feedback about staff experience. In others, they need clear descriptive data about uptake, reach, or trends over time. In others, they need stronger evidence about whether an initiative likely contributed to a meaningful change.

Without a clear framework, teams can easily run into predictable problems:

- Collecting data that do not answer the most important question
- Choosing measures before clarifying the purpose of the evaluation
- Attributing impact that is not fully supported by the evidence
- Failing to involve the right stakeholders early
- Overlooking which groups are benefiting, which are not, and which may need support but aren't receiving it
- Continuing initiatives that are not delivering sufficient value
- Stopping promising initiatives before their impact is properly understood

In other words, the challenge is often not a lack of activity; it is a lack of structure. Many organizations try to retrofit evaluation after the fact, when key data may already be missing or when success was not clearly defined to begin with.

That is the problem this guide is designed to solve.

WHY USE THIS APPROACH?

Why healthcare systems need to evaluate initiatives

This guide is intended to help healthcare leaders, Human Resources teams, operational leaders, and program owners approach initiative evaluation in a more structured and decision-oriented way.

It does this by helping organizations move through two critical steps.

1. Clarify the purpose of the evaluation

Before choosing measures, methods, or analyses, organizations first need to define what decision the evaluation is meant to support.

For example, the purpose may be to:

- Better understand a problem
- Assess whether an initiative fits a real need
- Monitor implementation and evaluate whether the initiative is helping
- Decide whether to continue, refine, scale, target differently, or stop it

Clarifying that purpose matters because it shapes everything that follows. It determines what kind of evidence is needed, what data will be most useful, and what conclusions the evaluation should reasonably support.

2. Identify the strongest practical way to assess the initiative

Once the purpose is clear, the next step is to determine how the initiative can realistically be evaluated given the data, timing, and comparison options available.

This guide helps organizations answer three foundational questions:

- **What kind of evidence do you need?**
- **What kind of data do you already collect, have access to, or need access to?**
- **When do you have data from, relative to the initiative?**
- **What kind of comparison do you have?**

Together, these questions help organizations choose an evaluation approach that is realistic, useful, and credible, rather than defaulting to a method that is either too weak to inform action or too ambitious for the data available.

When healthcare systems evaluate initiatives well, they are better able to:

- Identify which efforts are worth continuing or expanding
- Improve initiatives that show promise but need refinement
- Stop or redesign efforts that are not delivering meaningful benefit
- Understand which groups are benefiting and which may be left out
- Build a stronger case for leadership support, continued funding, or scale-up
- Estimate whether likely benefits justify the investment

This is especially important in healthcare, where even modest improvements can have meaningful operational and financial implications. A successful initiative may reduce turnover, improve engagement, lower burnout, reduce documentation burden, improve workflow efficiency, or increase uptake of useful resources. Those effects can translate into real value for the system. Conversely, an initiative that is ineffective, poorly targeted, or burdensome may quietly consume resources without delivering sufficient return.

In that sense, evaluation helps organizations do more than ask, *Did people like this?* It helps answer the questions that matter most for leadership decision-making:

- Is this initiative helping?
- Are we making an adequate impact?
- For whom is it helping, and under what conditions?
- Is it worth continuing, changing, scaling, or stopping?
- What is the likely return on this investment?

WHAT DO WE MEAN BY “INITIATIVE”?

When we say **initiative**, we mean any organized effort to effect change.

That could be a program, policy, tool, training, resource, process change, or pilot. It could be large or small. It could be aimed at patient safety, staff well-being, retention, workflow, efficiency, engagement, or another organizational goal.

Initiatives could include but are not limited to:

- New policy
- Patient safety initiative
- Workflow change
- Training or coaching offer
- Digital tool or platform
- Well-being resource
- Staffing model
- Communication campaign
- Support service
- Scheduling change
- Leadership practice or management intervention
- New process for documentation, handoff, or referral
- Pilot effort introduced in one unit, site, or team

In other words, if your organization is introducing something new, changing an existing practice, or investing in a resource in the hope that it will improve outcomes, it can likely be evaluated.

This broad definition is important because evaluation is often thought of as something reserved for large formal programs or research studies. In practice, many of the changes that shape staff experience, workflow, costs, and outcomes are smaller operational or policy decisions that still deserve careful evaluation.

An initiative does not have to be large, expensive, or branded to matter. If it is intended to influence behavior, experience, processes, or outcomes, then it is reasonable to ask whether it is working, for whom, and under what conditions.

This guide is designed to support evaluation thinking across that wide range of initiatives.

How to Evaluate an Initiative
A Practical Decision Guide

Start with your goal and purpose. What do you hope this project will achieve? Then follow the steps to determine what your evaluation can show and what to do next.

Flow Steps

DEFINE THE DECISION

What decision is this evaluation meant to support?

QUESTION 1

What kind of evidence do you need?

- Understand experiences
- Track what's happening
- Evaluate impact

QUESTION 2

When do you have data from?

- One point in time
- Before & after
- Over time

QUESTION 3

What kind of comparison do you have?

- No comparison
- Natural comparison
- Deliberate comparison

DECIDE WHAT HAPPENS NEXT

Continue | *Refine* | *Scale* | *Target* | *Stop*

INITIATIVE EVALUATION STEPS

Before Evaluation: *Define the purpose and outcome this initiative needs to support*

Before selecting measures, methods, or data sources, it is essential to **define the purpose of the initiative and the decision the evaluation is meant to inform.**

Healthcare organizations launch initiatives to address real operational and workforce challenges: retention, burnout, workflow inefficiencies, documentation burden, staffing strain, and gaps in support. But once an initiative is underway, a familiar question emerges: **is this initiative delivering enough value to justify continued investment?**

That question should guide the evaluation from the outset.

In practice, many organizations only recognize the importance of defining the decision after encountering its absence. Dr. Waghray described launching large-scale support efforts for frontline leaders that were widely appreciated but difficult to sustain over time.

"We had strong anecdotal feedback. People found it helpful," Dr. Waghray explained. "But we hadn't defined what success would look like in measurable terms, and that made it harder to justify continuing at the same scale."

As a result, the organization didn't stop the initiative, but they had to adapt it. They narrowed its focus and introduced more tailored measurements to better understand where it was creating value.

Dr. Santana-Cebollero described a similar situation with Finding Meaning in Medicine, where a structured storytelling and peer connection program showed strong qualitative enthusiasm but struggled to sustain participation over time. "It was an amazing program," she said, "but it required time outside of clinical work, and we didn't have the data to show leaders why it was worth that investment."

Evaluation is most useful when it is tied to a concrete organizational purpose and corresponding decision. Without that clarity, teams often collect data that are difficult to interpret, disconnected from leadership priorities, or insufficient to guide action. A clear decision focus helps determine what should be measured, whose perspectives matter, and what kind of evidence will be useful.

What decision is the evaluation meant to inform?

Ms. Arespacochaga emphasized that one of the challenges across healthcare initiatives is a lack of clear agreement about what success means for the organization and the teams involved.

"You have to be clear about what you can and cannot impact through the initiative," Ms. Arespacochaga noted. "So you can demonstrate that what you're doing is making a difference."

> "You have to be clear about what you can and cannot impact through the initiative, so you can demonstrate that what you're doing is making a difference." – Ms. Elisa Arespacochaga, Group Vice President of Clinical Affairs and Workforce at the American Hospital Association

Without that clarity, different stakeholders often evaluate the same initiative through different lenses, making it harder to align on what the data truly means for the stated goals and what decision it should inform.

In practice, organizations are usually trying to support one or more of five decisions:

1. **Clarify the problem.**
In some cases, the immediate need is not to evaluate a solution, but to better understand the underlying challenge. Where is the friction? Who is most affected? What appears to be driving the problem?

2. **Assess need and fit.**
Sometimes the question is whether a proposed initiative addresses a meaningful need, for whom it is relevant, and whether it fits the context in which it will be used.

3. **Monitor implementation.**
Some evaluations are designed to determine whether an initiative is being used as intended, by the intended groups, and under the conditions required for success.

4. **Evaluate whether the initiative is addressing the stated need.**
In other cases, the goal is to assess whether the initiative appears to improve outcomes that matter to the organization, such as workflow, staff experience, retention, efficiency, or another strategic priority.

5. **Guide what happens next.**
Often, the ultimate purpose is a leadership decision: whether to continue, refine, scale, target differently, or discontinue the initiative.

Stakeholder alignment is part of evaluation planning

A strong evaluation plan does not begin only with the question, **What do we want to learn?** It also begins with, **Who needs to be involved in defining success and acting on the results?**

Many initiatives cut across organizational functions. A digital tool may rely on IT or the EHR implementation team. A staffing or support initiative may affect Human Resources. A workflow change may require operational leadership. An initiative aimed at workforce well-being may also affect patient experience, adoption, and sustainability.

If the right stakeholders are not involved early, organizations risk defining success too narrowly, overlooking key constraints, missing relevant outcomes, and limiting buy-in for the findings. For that reason, stakeholder alignment should be treated as a core part of evaluation design, not as a separate implementation issue.

Relevant stakeholders may include:

- **IT and EHR team**, when a platform, system integration, or technical workflow is involved
- **Human Resources**, when the initiative affects staff experience, retention, policy, or workforce strategy
- **Operational leaders**, when workflows, staffing models, or service delivery may change
- **Clinical leaders or frontline staff**, when day-to-day practice will be affected
- **Well-being, quality, or patient experience teams**, when the initiative may influence those priorities
- **Finance or executive leadership**, when the evaluation may inform resource allocation, continuation, or scale
- **Communications,** when messaging, stakeholder engagement, or rollout strategy will influence adoption and implementation success
- **Legal or compliance**, when there are privacy or other related concerns

Why early stakeholder involvement matters

Involving the right stakeholders early improves both the quality of the evaluation and the likelihood that its findings will be used.

It helps organizations:

- Identify outcomes that matter across functions
- Surface operational dependencies and risks early
- Strengthen internal alignment and credibility
- Build buy-in for the evaluation process and its results
- Improve the likelihood that findings lead to action

This matters because an initiative may appear successful from one perspective while creating challenges from another. A tool that improves efficiency for one team may increase burden for another. A program that is well received by staff may still face sustainability challenges if operational leaders, IT, or finance were not involved early enough to shape the approach. No single stakeholder will see every potential challenge or downstream implication of an initiative. Involving the right perspectives early can help identify risks, avoid blind spots, and prevent significant time and rework later on.

Questions to answer before moving forward

Before selecting an evaluation approach, leadership teams should be able to answer the following questions:

- *What decision is this evaluation meant to support?*
- *Are we trying to clarify a problem, assess fit, monitor implementation, evaluate outcomes, or guide next steps?*
- *Who needs to be involved in defining success?*
- *Which groups will be affected by this initiative?*
- *Who controls the systems, workflows, approvals, or resources required for success?*
- *Who owns the data needed for evaluation?*
- *Who will need confidence in the findings in order to support continuation, redesign, or scaling?*

A stronger foundation for the rest of the evaluation

Answering these questions at the outset creates a stronger foundation for everything that follows. It sharpens the purpose of the evaluation, improves cross-functional alignment, and increases the likelihood that the results will be relevant, credible, and decision-ready.

In short, evaluation should not begin with methods. It should begin with clarity: clarity about the decision, clarity about the desired learning, and clarity about who needs to be at the table.

An Illustrative Example

A health system launches a new retention initiative in response to staffing pressures. After several months, leaders ask whether it worked. One person shares positive staff comments. Another shares low participation rates. Someone else points to a small improvement in survey scores. No one can say whether the right outcomes were measured, whether the timing of the data makes sense, or whether any observed change was likely due to the initiative.

So the organization is left in a familiar position: making an important decision with incomplete evidence.

The risk is not only that ineffective initiatives continue. It is also that promising initiatives are abandoned too early, staff are asked to engage in programs that are not helping, and resources are spent without a clear return. A thoughtful evaluation approach helps organizations avoid this trap.

Choose the Right Type of Evidence for Your Goal

Different evaluation goals require different types of evidence

TYPE OF EVIDENCE	WHAT IT HELPS YOU UNDERSTAND	BEST USED WHEN YOU WANT TO	WHAT IT LOOKS LIKE IN PRACTICE
Qualitative *(Stories & Feedback)* 	Why people engage – or don't; how they experience the initiative	Understand staff experiences, barriers, or unmet needs Identify what is helping or frustrating users Learn why people engage or don't engage Refine or improve the initiative	Interviews, focus groups, open-ended survey responses, direct feedback
Descriptive *(What's Happening)* 	What is happening, how often, and for whom	Track uptake, participation, and usage Monitor implementation Identify trends or patterns over time Summarize reach across groups	Participation rates, usage metrics, survey scores, trend data over time
Impact *(Did It Work?)* 	Whether the initiative likely contributed to meaningful change	Evaluate whether outcomes improved Compare results across groups or time periods Support decisions about scaling, funding, or continuation Assess return on investment	Changes in retention, burnout, engagement, or performance compared across groups or time periods

QUESTION 1: WHAT KIND OF EVIDENCE DO YOU NEED TO UNDERSTAND THIS INITIATIVE?

Now that you are clear on the purpose of your evaluation, you can begin to determine your approach.

Different evaluation approaches answer different kinds of questions. Some help you understand people's experiences in detail. Some help you describe what happened in a measurable way. Others help you ask whether the initiative likely contributed to a meaningful change.

This is the most foundational question in the guide because it helps clarify the **main goal of the evaluation** and the **type of insight needed to support that goal**. In other words, it helps you decide what kind of answer you are actually looking for before thinking about data sources, timing, or comparison options.

Start with the goal of the evaluation

The kind of evidence you need depends on what you are trying to accomplish.

For example, your goal may be to:	‣ **Understand staff experiences, barriers, or needs** ‣ **Monitor reach, uptake, or implementation** ‣ **Determine whether outcomes appear to be improving** ‣ **Support a decision about whether to continue, refine, scale, fund, or stop the initiative**
Different goals call for different kinds of evidence:	‣ If the goal is to understand how an initiative is being experienced, why people are engaging or not engaging, or what needs to be improved, detailed qualitative feedback may be most useful. ‣ If the goal is to track participation, reach, use, or observed changes over time, descriptive quantitative data may be the best fit. ‣ If the goal is to assess whether the initiative likely contributed to a measurable improvement, stronger statistical evidence is usually needed.

Qualitative data, such as rich stories and detailed feedback

In many cases, qualitative feedback is the first signal that something is working or not working.

During the pandemic, Dr. Waghray and his team initially assumed that staff primarily wanted access to clinical mental health resources. But through open-ended feedback and direct outreach, a different pattern emerged.

"We went in thinking people valued one thing," he shared. "But what we heard consistently was something else: people wanted connection, support, and a sense that someone was paying attention to what they were going through."

Example: Dr. Schumm described how qualitative input – particularly through town halls and open discussion – surfaced perspectives that had not appeared in structured data collection. These conversations revealed deeper cultural dynamics and helped shape future interventions in ways that surveys alone could not.

That insight led to the expansion of a proactive outreach program designed to meet staff where they were, an adjustment that would not have emerged from survey scores alone.

Dr. Santana-Cebollero saw qualitative feedback directly reshape the Finding Meaning in Medicine program. Participants shared that while they valued the experience, they struggled to attend sessions outside of work hours. In response, her team expanded the model to include family-inclusive events, where spouses attended too, turning participation into shared experiences rather than additional time away from home. “That feedback changed how we thought about engagement,” she explained.

This type of evidence helps you understand people’s experiences, perspectives, and suggestions in their own words. It is most useful when your goal is to understand **how the initiative is being experienced, why it is or is not working in practice**, or **what should be improved**.

This is especially useful when your goal is to:

- Understand staff experiences, barriers, or unmet needs
- Identify what is helping or frustrating users in practice
- Learn why people are engaging, not engaging, or using the initiative differently
- Refine or improve an initiative based on frontline feedback

Turn to page 20 for more detail.

Quantitative description of uptake, reach, and observed changes over time

Descriptive data often surfaces patterns that are not immediately obvious and can challenge initial assumptions.

For example, Dr. Waghray described how his team introduced a “stress meter” tool to help staff identify their needs and access resources. When leaders analyzed usage patterns, they expected most users to select traditional self-help or clinical support options. Instead, a large proportion gravitated toward spiritual care resources.

That insight led to a system-wide expansion of spiritual health programming, an investment that would not have been prioritized without clear data on what staff were actually choosing.

In another example, Dr. Schumm found that participation challenges in employee surveys were not due to disengagement, but to survey fatigue. By simplifying the process and reducing frequency, participation improved significantly, leading to more reliable data and better-informed decisions.

This type of evidence helps you describe what happened during the initiative in a clear and measurable way, such as reach, participation, uptake, scores, or changes over time. It is most useful when your goal is to **monitor implementation, track who is being reached, or summarize what is happening and how often**.

This is especially useful when your goal is to:

- Monitor uptake, participation, or use
- Understand who is being reached and who is not
- Track trends over time
- Provide clear descriptive summaries for operational or leadership review

Turn to page 21 for more detail.

Quantitative evidence about whether the initiative made an impact

Stronger evaluation becomes especially important when organizations need to make higher-stakes decisions about investment, scale, or continuation.

Dr. Schumm described a leadership development program where evaluation data clearly demonstrated impact. Participants demonstrated stronger engagement, lower burnout rates and less attrition, justifying continued investment in the program.

In contrast, other initiatives with similarly positive anecdotal feedback, but without comparable data, faced greater scrutiny and were harder to sustain.

> "We could look at the data and see that people who went through the program were more engaged and less burned out, with lower attrition rates. That made it much easier to justify continuing and expanding the program." – Herb Schumm, M.D., former Chief Medical Officer at Bon Secours Mercy Health and board member of the Coalition for Physician & APP Well-Being

Dr. Santana-Cebollero described an example where stronger evaluation made a difference. A mission-based interview process for physicians was initially introduced to assess alignment with organizational values prior to hiring. Over time, her team collected longitudinal data – tracking retention, satisfaction, and alignment at multiple time points, post-hiring. "We now have several years of data showing that when physicians understand our mission upfront, they are more likely to stay," she explained.

This contrast illustrates a key principle: perception of value may be enough to start an initiative, but evidence of impact is often required to scale it.

This type of evidence helps you go beyond description and ask whether the initiative likely contributed to a measurable change. It is most useful when your goal is to **evaluate whether the initiative is helping**, or to support decisions about whether to **continue, expand, fund, scale, or stop** it.

This is especially useful when your goal is to:

- Determine whether outcomes improved in a meaningful way
- Assess whether observed changes were likely related to the initiative
- Compare results across groups, sites, or time periods
- Support higher-stakes decisions about investment, continuation, or scale-up

Turn to page 22 for more detail.

In practice, you may need more than one type of evidence

Many leaders emphasized that the most useful evaluations combine multiple types of evidence. "You need both," said Dr. Waghray. "The quantitative data tells you what is happening at scale. The qualitative feedback tells you why."

Together, these insights provided a much clearer picture than either type of data alone and helped justify continued investment in the program.

At a system level, Ms. Arespacochaga noted that this combination is especially important for complex issues like well-being, where subjective experience and organizational outcomes need to be understood together. An initiative may improve measurable outcomes without being experienced positively by staff. Understanding both perspectives helps organizations identify when an initiative is directionally effective but still needs refinement to improve adoption, experience, or long-term sustainability.

Example: In a proactive mental health outreach program, **quantitative data revealed** that a significant number of staff were struggling but had not sought help before the program; **qualitative feedback explained** why.

The anonymity and flexibility of the new program made it feel safe and accessible.

These options are not mutually exclusive. In many real-world evaluations, organizations need more than one type of evidence because they have more than one goal. For example, an organization may want to know:

- Whether outcomes changed
- How staff experienced the initiative

In these cases, combining evidence types is often the strongest approach. Qualitative data can help explain **why** an initiative succeeded or struggled, while quantitative data can help show **how broadly it was used** and **whether outcomes changed**.

Unexpected findings are often where the most valuable insights emerge. Dr. Waghray noted that many initiatives did not work uniformly across groups, even when they were well received broadly. "The same messaging didn't resonate the same way with all groups," he explained. "We had to iterate and customize based on where people were."

These differences highlighted how the same structured program could resonate differently depending on role, context, and local leadership support. These differences are not a failure of the initiative; they are opportunities to refine and target it more effectively.

Example: Dr. Santana-Cebollero observed similar variation across roles and settings within Finding Meaning in Medicine. In some regions, physician participation was limited, while advanced practice providers (APPs) engaged more actively and even requested more frequent sessions.

Ms. Arespacochaga emphasized that for these insights to best inform leadership decisions, they must be connected to their impact on broader mission priorities for the organization such as patient care, safety and quality, financial performance, and workforce engagement.

Once you know what kind of evidence you need, the next questions help you figure out whether your available data and comparison options can support that goal.

Rich stories and detailed feedback

Use this when you need in-depth, open-ended information in people's own words.

This option is best when you want to understand the **details behind people's experiences** — for example, how staff experienced a new initiative, what made it helpful or frustrating, why they did or did not engage, and what barriers or unexpected effects came up. This type of evidence is usually collected through interviews, focus groups, written comments, or other open-ended responses. It is especially useful when you want depth, nuance, and concrete examples rather than counts or averages.

This is best for questions like:	• What was the experience like for staff to use this? • Why did people engage with it or avoid it? • What problems, barriers, or benefits came up in practice? • What are the main themes in staff feedback?
This type of evidence is good for:	• Improving or refining a program, tool, policy, or resource • Identifying barriers to adoption • Understanding staff needs and concerns • Collecting compelling quotes or examples for leadership
This type of evidence does not tell you:	• How common a view is across the whole workforce • Whether a change caused measurable improvement
Examples:	• Interviewing nurses about their experience using a new handoff tool • Running focus groups with staff about why a well-being resource is underused • Reviewing open-ended comments about a new scheduling policy

Description of uptake, reach and observed changes over time

Use this when you want to describe what happened in a clear, measurable way.

This option is best when you want to summarize uptake, usage, participation, scores, or trends over time. It helps show whether something went up, down, or stayed the same, and how widely a program, tool, policy, or resource was used. This type of evidence is often presented as percentages, averages, counts, charts, or before-and-after summaries. It is useful for monitoring and reporting, but it does not show whether the initiative itself caused the change.

This is best for questions like:	‣ How many people used it? ‣ How often was it used? ‣ Did scores or outcomes go up or down? ‣ How did use or outcomes change over time? ‣ Did use differ for certain subgroups?
This type of evidence is good for:	‣ Tracking reach and uptake ‣ Monitoring implementation ‣ Showing before-and-after patterns ‣ Creating dashboards and leadership summaries ‣ Spotting promising trends
This type of evidence does not tell you:	‣ Whether the initiative caused the change ‣ Whether the same change would have happened anyway
Examples:	‣ 18% of eligible staff enrolled in a new coaching program ‣ Average documentation time fell from 5 hours to 4 hours after launch of a new tool ‣ Use of an employee support resource increased steadily over six months

Evidence of impact

Use this when you need to understand whether a program, policy, tool, or resource contributed to a measurable outcome or change.

This option is best when leaders need to make a bigger decision. For example, whether to continue, expand, fund, or stop an initiative. It is used when you want to go beyond simply describing what happened and instead ask whether the observed change was likely related to the initiative rather than other factors happening at the same time. This kind of evaluation usually requires stronger data, such as a comparison group, before-and-after data, repeated time points, or a rollout structure that allows meaningful comparison.

This is best for questions like:	‣ Did this change likely improve outcomes? ‣ Did it make a bigger difference than would have happened otherwise? ‣ Did one group improve more than another after receiving it? ‣ Is there enough evidence to justify continued investment?
This type of evidence is good for:	‣ Supporting scale-up or funding decisions ‣ Evaluating impact ‣ Comparing groups or sites ‣ Testing whether an initiative likely contributed to improvement
Examples:	‣ Did units that adopted a new staffing model show lower turnover than similar units that did not? ‣ Did burnout scores improve more in staff who received coaching than in those who did not? ‣ Did a new documentation tool reduce charting time beyond the usual trend?

QUESTION 2: THINKING ABOUT WHEN THIS INITIATIVE STARTED, WHEN DO YOU HAVE DATA FROM?

Once you are clear on what kind of evidence you need, the next step is to look at the data you have available and when those data were collected relative to when the initiative began. **IMPORTANT NOTE**: There is likely more data already being collected that you may not be aware of or have prior access to. This is an opportunity to engage a broader group of stakeholders to improve the impact and reach of your initiative. For example, Quality may have data you aren't thinking about that correlates to what you want to understand.

This question is about the **timing and structure of your available data relative to the initiative**. It helps you understand what your data can actually show. In other words, it helps you determine whether you can describe one point in time, examine change before and after the initiative, or look at trends across multiple time points.

The timing of your data affects **what you can show and how strongly you can show it**. In general, having data from before and after an initiative – especially across multiple time points – allows you to make stronger claims than having data from only one point in time.

For example, you may have:

Data from: one time period	This means you have data from only one point relative to the initiative, usually during or after it began. This allows you to describe what things looked like at that moment, but not whether they changed over time. *Turn to page 25 for more detail.*
Data from: before and after the initiative began	This means you have at least one time point before the initiative started and at least one time point after it began. This allows you to examine whether outcomes changed after the initiative began compared with before. *Turn to page 26 for more detail.*
Data from: multiple points over time	This means you have repeated data, such as weekly, monthly, quarterly, or repeated survey data, spanning before and/or after the initiative. This allows you to look at patterns and trends over time, not just a single before-and-after difference. *Turn to page 27 for more detail.*

To choose where to start, ask yourself: **When, relative to the initiative, do we actually have data?** Begin with the option that best matches your available data, then turn to the detailed section for that option.

Importantly, these options are **not separate from the evidence goals in Question 1**. Instead, they can be combined with any of them to help determine the strongest possible analytic approach.

Once you know when your data were collected, the next question helps you determine what kinds of comparisons are possible.

While this section focuses on timing, it's important to note that feasibility also matters. In many healthcare organizations, the teams responsible for accessing and analyzing data are distributed. For example, the electronic health record (EHR) analytics team that manages clinical and operational data may be separate from the IT team responsible for systems and infrastructure. Human Resources, quality, and finance teams may also own different data sources.

As a result, even when data exist, accessing them may require coordination across multiple teams, approvals, or competing priorities. Health system leaders should consider not only what data are theoretically available, but also what is realistically accessible given internal bandwidth, timelines, and resources.

In practice, this means asking a few additional questions early:

Do we have the capacity to pull and analyze this data?

How long will it take?

Who needs to be involved?

And is the effort required aligned with the importance of the decision we are trying to make?

Option 1: Data from one time period

Use this when you only have data from one point in time relative to the initiative, usually during or after it began.

This option is best when your available data provide a single snapshot. For example, you may have one post-launch survey, one quarter of usage data, one reporting period after implementation, or one snapshot of burnout, turnover, or satisfaction. This setup allows you to describe what things look like at that point in time. If you also have another group to compare with, you may be able to compare groups at the same time point. However, on its own, this structure usually does not allow you to show whether things changed after the initiative began.

This is best for questions like:	‣ What does participation, usage, or satisfaction look like right now? ‣ How are staff experiencing the initiative at this point in time? ‣ What proportion of eligible staff used the resource during this reporting period? ‣ How do outcomes differ across teams, roles, or sites at this moment?
This type of data structure is good for:	‣ Describing a current state or snapshot ‣ Monitoring implementation at a single point in time ‣ Summarizing uptake, reach, or outcomes during one period ‣ Comparing groups at the same time point, if comparison data are available ‣ Generating an initial picture when more data are not available
This type of data structure does not usually allow you to:	‣ Examine change over time ‣ Show whether outcomes improved after the initiative began ‣ Determine whether the initiative likely caused a change ‣ Distinguish initiative effects from pre-existing differences unless you have a strong comparison
Examples:	‣ A single post-launch staff survey about a new well-being resource ‣ One quarter of usage data for a new digital tool ‣ One reporting-period snapshot of turnover, burnout, or satisfaction after implementation ‣ A one-time set of interview or focus group data collected after rollout

Option 2: Data from before and after the initiative began

Use this when you have at least one time point before the initiative began and at least one time point after it began.

This option is best when your data allow you to compare outcomes before the initiative with outcomes after it started. For example, you may have one survey before launch and one after, one measure of documentation time before a workflow change and one after, or one pre period and one post period for outcomes like turnover, satisfaction, or usage. This setup allows you to examine whether things changed after the initiative began compared with before. It is one of the most common and practical evaluation structures.

This is best for questions like:	‣ Did usage, satisfaction, or outcomes change after the initiative started? ‣ Were scores better after launch than before? ‣ Did documentation time, uptake, or other outcomes improve after implementation? ‣ Did staff report different experiences after the initiative compared with before?
This type of data structure is good for:	‣ Examining whether outcomes changed after the initiative began ‣ Conducting practical pre-post evaluations ‣ Showing before-and-after differences ‣ Supporting stronger descriptive or statistical analyses than a single snapshot ‣ Providing an accessible starting point for evaluating possible initiative effects
This type of data structure does not necessarily allow you to:	‣ Rule out other factors that may have affected outcomes at the same time ‣ Determine whether the initiative alone caused the observed change ‣ Show whether trends were already moving in the same direction before the initiative ‣ Capture more complex patterns over time unless additional time points are available
Examples:	‣ One survey before launch and one survey after launch ‣ One pre-implementation and one post-implementation measure of charting time ‣ One period of turnover data before a new staffing model and one period after ‣ Interviews conducted both before and after implementation to understand changes in staff experience

Option 3: Data from multiple points over time

Use this when you have repeated data, such as weekly, monthly, quarterly, or repeated survey data, spanning before and/or after the initiative.

This option is best when your data allow you to look beyond a single snapshot or one before-and-after comparison and instead examine patterns over time. For example, you may have monthly turnover data, weekly usage data for a new tool, repeated pulse surveys, or regularly tracked operational metrics. This setup allows you to see whether outcomes were already changing before the initiative started, whether there was a noticeable shift after launch, and whether any changes were sustained over time.

This is best for questions like:	‣ Were outcomes already changing before the initiative started? ‣ Did the trend shift after implementation? ‣ Was there a sudden or gradual change after launch? ‣ Were any changes sustained over time? ‣ Did uptake or outcomes fluctuate across different phases of implementation?
This type of data structure is good for:	‣ Examining trends over time ‣ Distinguishing one-time fluctuations from more stable changes ‣ Understanding whether outcomes were already improving or worsening before launch ‣ Assessing whether changes appear sustained, delayed, or temporary ‣ Supporting stronger evaluations of whether an initiative may have contributed to change
This type of data structure does not automatically allow you to:	‣ Prove causation on its own ‣ Fully rule out other events or system-wide changes happening at the same time ‣ Substitute for a comparison group when one is needed ‣ Guarantee strong conclusions if the data are inconsistent, sparse, or poorly aligned with rollout timing
Examples:	‣ Monthly turnover data tracked for a year before and after a new staffing model ‣ Weekly usage data for a newly introduced digital resource ‣ Repeated pulse surveys measuring staff well-being over time ‣ Operational performance metrics tracked quarterly across implementation

QUESTION 3: WHAT KIND OF COMPARISON DO YOU HAVE?

Once you know what kind of evidence you need and when your data were collected, the next step is to think about whether you have another group to compare with, and how that group was selected.

This question helps you understand how strong your evaluation can be. In other words, it helps you assess whether you can only describe what happened among people who received the initiative, or whether you can compare their outcomes with a group that did not receive it.

This matters because comparison is one of the main ways evaluators judge whether an initiative likely made a difference. If you only have data from people who received the initiative, you may still be able to describe what happened or look at change over time. But if you also have another group to compare with, you are in a stronger position to ask whether the initiative may have contributed to any observed differences.

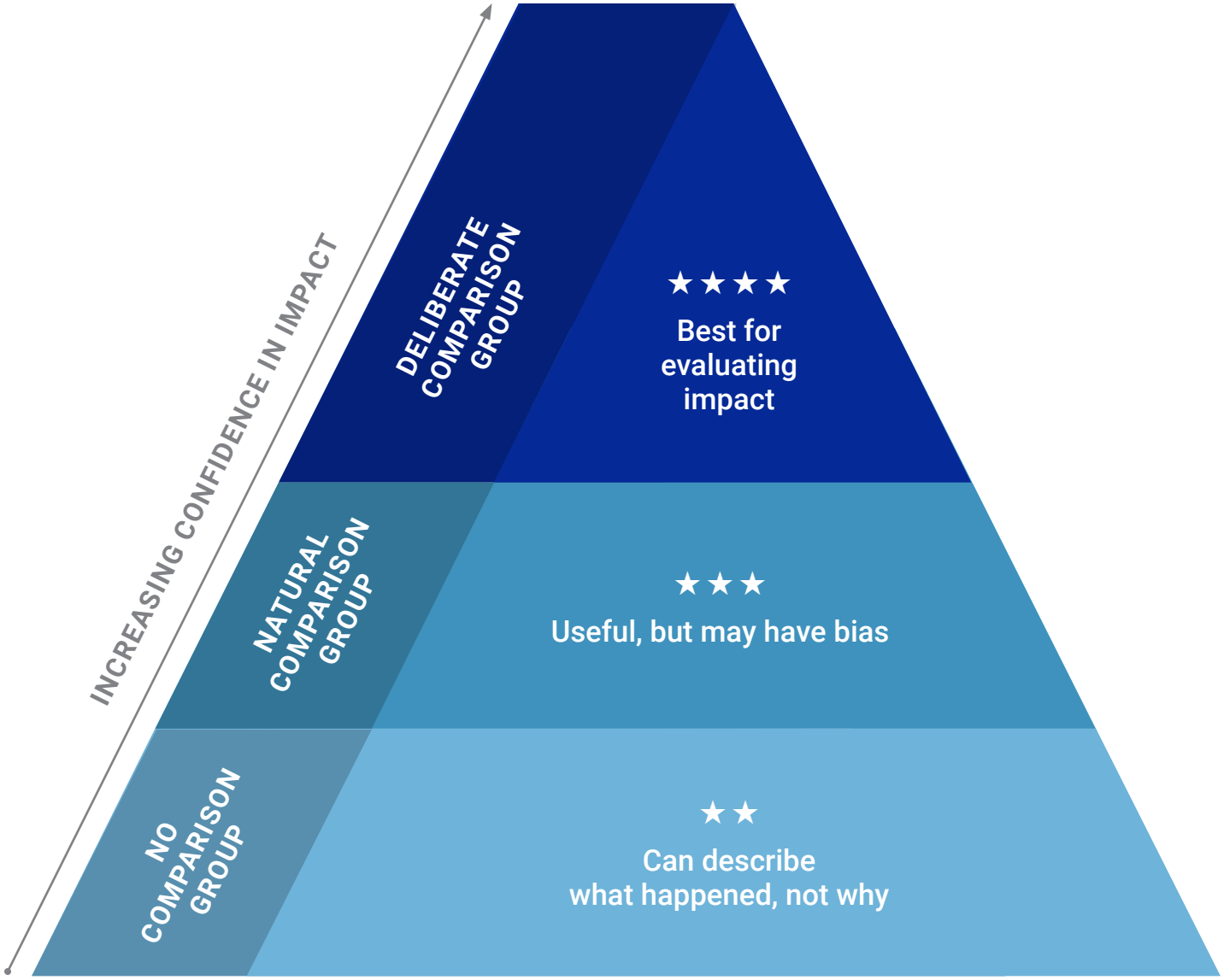
For example, you may have:

No comparison group	You only have data from the people, teams, or sites that received the initiative. This allows you to describe what happened in that group, but it is harder to tell whether the initiative itself caused any observed change. *Turn to page 30 for more detail.*
A naturally occurring comparison group	You have another group to compare with, but the groups were formed through normal operations rather than by design. This gives you a useful real-world comparison, but the groups may differ in important ways that affect the results. *Turn to page 31 for more detail.*
A deliberately assigned comparison group	You have another group to compare with, and the groups were assigned on purpose using a structured process to support a fair comparison. This usually provides the strongest basis for testing whether the initiative likely made a difference. *Turn to page 32 for more detail.*

Importantly, these options can be combined with any of the evidence goals in Question 1 and any of the data timing structures in Question 2. They do not tell you what kind of evidence you want or when your data were collected. Instead, they help determine how strong your comparison is, and therefore how strongly your evaluation can support conclusions about impact.

To choose where to start, ask yourself: **Do we have another group to compare with, and if so, how was that comparison created?** Begin with the option that best matches your situation, then turn to the detailed section for that option.

Strength of Evidence Depends on Your Comparison



Option 1: No comparison group

Use this when you only have data from the people, teams, or sites that received the initiative.

This option is best when everyone received the initiative, the whole hospital or system changed at once, you only collected data from participants or users, or there is no reasonable "other group" to compare against. In many real-world healthcare settings, creating a formal comparison or "control" group may not be practical, appropriate, or aligned with operational priorities. Not every initiative can, or should, be evaluated like a randomized controlled trial. In these situations, you can still learn useful things. For example, you may be able to describe participation, satisfaction, uptake, or outcomes within the group that received the initiative. If you have data from before and after, or across multiple time points, you may also be able to examine whether things changed over time. However, without another group for comparison, it is harder to know whether those changes were due to the initiative itself or to something else happening at the same time.

This is best for questions like:	‣ What happened among the people or teams who received the initiative? ‣ How did participation, usage, or outcomes look after rollout? ‣ Did outcomes improve over time within the group that received the initiative? ‣ What was staff experience like among those exposed to the initiative?
This type of comparison is good for:	‣ Describing what happened within the initiative group ‣ Summarizing uptake, participation, satisfaction, or outcomes ‣ Examining change over time if pre-post or repeated data are available ‣ Monitoring implementation when no reasonable comparison group exists ‣ Generating useful early evidence in real-world settings
This type of data structure does not usually allow you to:	‣ Tell whether the initiative caused the observed change ‣ Rule out other explanations, such as broader organizational changes or outside events ‣ Know whether the same change would have happened even without the initiative ‣ Make as strong a case for impact as designs with a comparison group
Examples:	‣ A hospital-wide documentation change rolled out to all staff at once ‣ A well-being resource offered to the entire workforce, with no non-user comparison group ‣ A post-launch survey completed only by staff who participated in the initiative ‣ Burnout or satisfaction data tracked over time within one group that all received the initiative

Option 2: A naturally occurring comparison group

Use this when you have another group to compare with, but the groups were formed through normal operations rather than by design.

This option is best when one unit used the new approach and another did not, one hospital started earlier than another, one department chose to use the resource and another did not, or the initiative was offered to some groups but not others for practical reasons. This kind of comparison can be very useful in real-world settings because it gives you something to compare against without requiring a formal experiment. However, the groups may already differ in important ways before the initiative began. For example, they may differ in staffing, leadership, workload, culture, or readiness for change. That means any differences in outcomes may be partly due to the initiative and partly due to those pre-existing differences.

This is best for questions like:	‣ Did outcomes differ between groups that received the initiative and groups that did not? ‣ Did one site improve more than another after starting the initiative? ‣ How did adoption, satisfaction, or outcomes compare across naturally different units or departments? ‣ Is there suggestive evidence that the initiative may have contributed to better results?
This type of comparison is good for:	‣ Making use of real-world differences in rollout or participation ‣ Providing a stronger basis for interpretation than having no comparison group ‣ Supporting practical evaluation when random assignment is not feasible ‣ Comparing sites, units, teams, or departments under routine conditions ‣ Generating useful evidence about possible impact in operational settings
This type of comparison does not necessarily allow you to:	‣ Assume the groups were equivalent before the initiative started ‣ Rule out pre-existing differences between groups ‣ Conclude with high confidence that the initiative alone caused the outcome difference ‣ Remove the need to think carefully about timing, context, and group characteristics
Examples:	‣ One hospital adopted a new workflow while another similar hospital did not ‣ One department used a new support resource and another department did not ‣ A phased rollout where some clinics started earlier than others for operational reasons ‣ Comparing teams that chose to use a coaching program with teams that did not

Option 3: A deliberately assigned comparison group

Use this when you have another group to compare with, and the groups were assigned on purpose using a structured process to create a fairer comparison.

This option is best when staff were assigned by chance to receive the initiative or not, units or clinics were assigned to start earlier versus later, or rollout order was intentionally structured to support evaluation. This kind of design usually gives the strongest basis for testing whether the initiative likely made a difference because the comparison was set up in a more controlled way. The goal is to reduce the chance that the groups differ for important reasons other than the initiative itself.

This does not mean the evaluation is automatically perfect. Good measurement, good timing, and good implementation still matter. But in general, deliberately structured comparisons provide a stronger basis for asking whether observed differences were likely due to the initiative.

This is best for questions like:	‣ Did the initiative lead to better outcomes than not receiving it? ‣ Did groups assigned to receive the initiative improve more than groups assigned later or not at all? ‣ Is there strong evidence that the initiative likely contributed to change? ‣ Does this initiative have enough evidence of impact to justify scale-up or investment?
This type of comparison is good for:	‣ Providing the strongest basis for evaluating likely impact ‣ Supporting funding, scale-up, or continuation decisions ‣ Reducing bias in how groups are compared ‣ Testing whether outcome differences are more likely to be related to the initiative ‣ Strengthening confidence in evaluation findings
This type of comparison does not automatically allow you to:	‣ Ignore issues with poor data quality or missing data ‣ Assume implementation was consistent across all sites or groups ‣ Answer every question about experience, feasibility, or acceptability on its own ‣ Guarantee perfect conclusions if the design was not carried out as planned
Examples:	‣ Staff were randomly assigned to receive a coaching offer or usual practice ‣ Clinics were intentionally assigned to begin the initiative in different phases ‣ A stepped rollout was structured in advance to allow fair comparison between earlier and later groups ‣ Units were assigned by design, rather than by self-selection, to different implementation conditions

WHAT TO DO WITH THE RESULTS

Now what? Using results to decide what happens next

Once you have results, the next question is not simply whether the initiative "worked." The more useful question is: **What should we do next?**

> Ultimately, evaluation is only useful if it leads to action.

In practice, organizations described a range of next steps based on what they learned:

- Expanding programs that showed strong reach and measurable impact
- Refining initiatives that were valued but lacked clear outcomes
- Redirecting resources away from low-usage efforts

In some cases, initiatives were not stopped, but redesigned to better align with measurable outcomes. In others, evaluation revealed that an initiative was solving a different problem than originally intended, leading to a shift in how success was defined.

> "Programs that demonstrated both reach and impact received more support. The budgets assigned to those that were used very infrequently were reprioritized." – Arpan Waghray, M.D., CEO of Providence's Well Being Trust

Evaluation is most valuable when it helps an organization make a decision. In some cases, the results may suggest that an initiative is worth continuing as it is. In others, they may suggest that the initiative shows promise but needs refinement. Sometimes the findings point toward scaling the effort to additional teams or settings. And sometimes the most responsible decision is to stop investing in an approach that is not delivering enough value.

That is why evaluation should not end with a report or a slide deck. It should end with a decision.

The purpose of analysis is not just to describe what happened. It is to help leaders, teams, and stakeholders make better choices about where to focus time, energy, and resources. Even a modest evaluation can be useful if it helps answer practical questions such as:

- Should we continue this initiative?
- Should we change it?
- Should we expand it?
- Should we target it differently?
- Should we stop doing it?

These decisions are especially important in healthcare settings, where initiatives compete for limited funding, staff attention, implementation effort, and organizational trust. A thoughtful evaluation helps ensure that future decisions are guided by more than enthusiasm, anecdotes, or assumptions alone.

Echo back to the original goal

When reviewing results, it helps to return to the questions asked at the beginning of the process.

- What problem was this initiative intended to address?
- What outcomes were we hoping to improve?
- Whose needs was this meant to serve?
- What would have made this worth continuing?

Returning to those original goals helps keep interpretation grounded. It also helps avoid two common mistakes: continuing an initiative simply because people are attached to it, or abandoning an initiative too quickly without understanding what it did achieve.

In some cases, the results may show that the initiative did not achieve its primary aim, but did create other meaningful benefits. In other cases, the initiative may have been well liked but not delivered enough meaningful improvement to justify ongoing investment. Both are important findings. Evaluation is useful not because it guarantees a positive result, but because it makes the next decision clearer.

Report back to key stakeholders

Once findings are available, they should be shared back in a way that is useful to the people involved.

Different audiences usually need different kinds of reporting:

Senior leaders may need:	- A concise summary of what happened - Whether the initiative appears worth continuing - The likely operational, strategic, or financial implications
Frontline teams and participants may need:	- A clear summary of what was learned - How their feedback contributed - What the organization plans to do in response
Sponsors, partners, or funders may need:	- Whether the investment appears worthwhile - Whether the initiative should continue, shift, expand, or end - What the findings suggest about next steps

The goal is not only to present findings. It is also to **close the loop**. Reporting back helps organizations:

- Show that the initiative was taken seriously
- Demonstrate that feedback and data were used
- Strengthen trust in the evaluation process
- Improve engagement in future implementation and evaluation efforts

Maximizing return on investment through a more targeted approach

In many healthcare settings, the same intervention is rolled out broadly across an entire workforce, even though:

- Need may vary across groups
- Fit may differ by role, team, or context
- Likely benefit may not be the same for everyone

An important future direction for initiative design and evaluation is a shift away from blanket, one-size-fits-all implementation and toward a more targeted approach.

A more targeted approach starts with two questions:

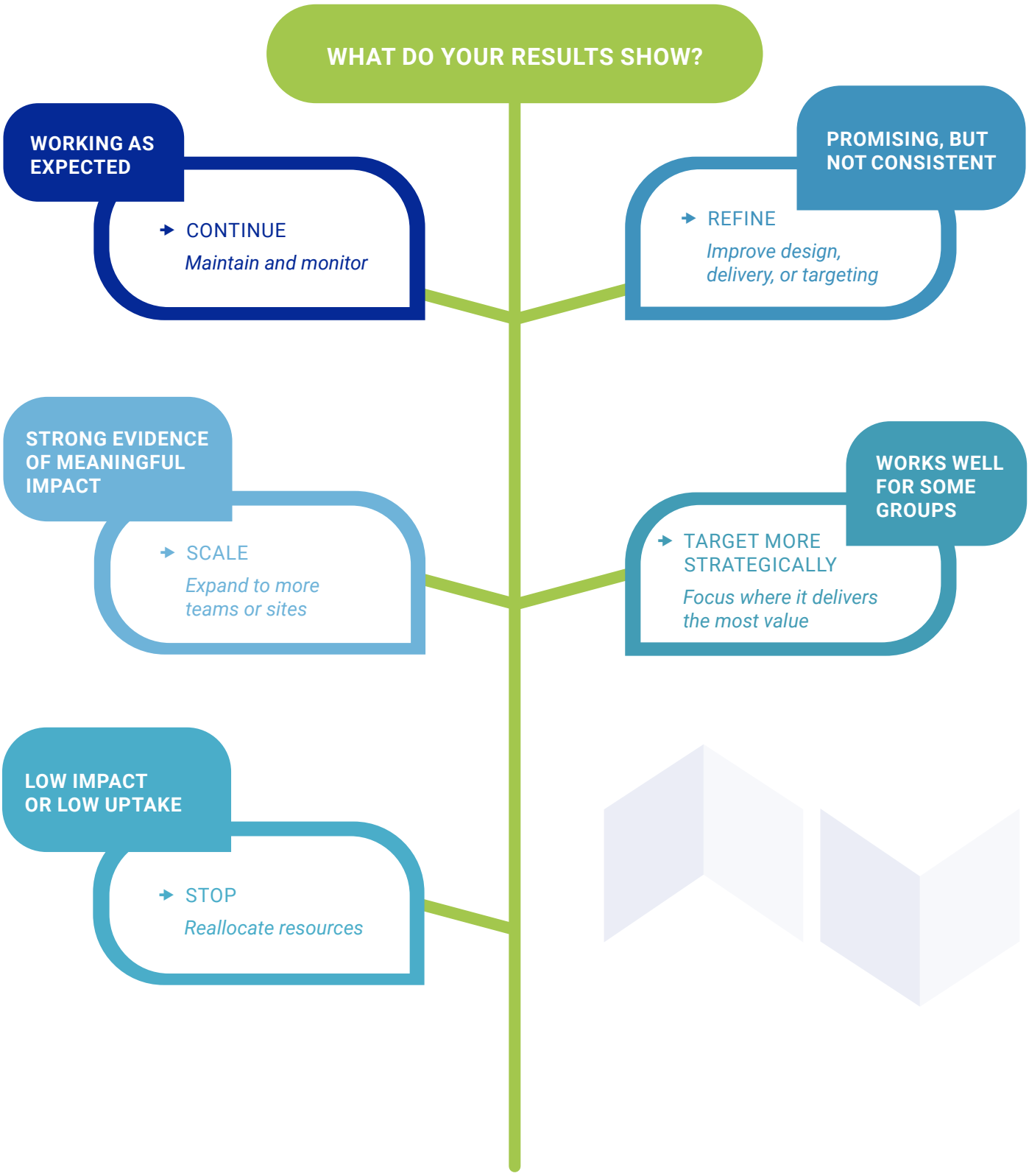
1. Who is most in need of support?
2. Who is most likely to benefit from this particular intervention?

Rather than assuming the same solution is equally valuable for everyone, this approach uses measurement to identify where an intervention is most likely to make a meaningful difference. A more targeted approach can help organizations make smarter decisions about:

- **Where to focus effort**
  (e.g., offering an intervention first to the staff most likely to benefit)
- **Where to invest funding**
  (e.g., identifying subgroups with different needs or response patterns)
- **Which groups to prioritize**
  (e.g., recognizing that a resource may be highly valuable for one group but less useful for another)
- **How to support adoption and sustainability over time**
  (e.g., avoiding broad rollout of interventions that are unlikely to deliver value across all settings)

This may ultimately be one of the most cost-efficient ways to implement interventions in complex healthcare environments.

What Should You Do Next? Use your evaluation findings to guide action, not just interpretation



DECIDE WHAT HAPPENS NEXT

The purpose of evaluation is not only to understand what happened. It is to support a decision about what should happen next.

By the end of an evaluation, an organization should be in a better position to answer a practical question: **What are we going to do now?** That question matters because evaluation only creates value when it informs action. A report, dashboard, or presentation is not the endpoint. The endpoint is a clearer, better-informed decision.

In many cases, the next step will fall into one of a few broad categories.

Continue

In some cases, the evidence suggests that the initiative is delivering enough value to continue in its current form.

This may be appropriate when:

- The initiative appears to be helping
- Implementation is working reasonably well
- The benefits appear meaningful enough to justify ongoing effort or investment
- There are no major barriers that require immediate redesign

A decision to continue does not mean the work is finished. It means the organization has enough reason to keep going while continuing to monitor performance over time.

Refine

In other cases, the initiative may show promise, but need adjustment.

This may be appropriate when:

- The initiative appears useful, but uptake is lower than expected
- Certain parts of the experience are creating frustration or barriers
- The initiative seems to work better in some contexts than others
- The core idea is sound, but fit, usability, communication, or delivery need improvement

In these situations, the right decision is often not to abandon the initiative, but to improve it. Evaluation can help identify where the friction is and what kinds of changes are most likely to strengthen the next version.

Scale

Sometimes the evidence suggests that the initiative is strong enough to expand.

This may be appropriate when:

- The initiative appears to produce meaningful benefit
- Implementation is feasible and sustainable
- There is enough confidence that the approach could work in other teams, sites, or settings
- Leadership needs evidence to justify broader investment

A decision to scale should usually be made carefully. An initiative that works well in one setting may still require adaptation elsewhere. Even so, evaluation can provide the foundation for moving from local success to broader implementation

Target more strategically

Sometimes the most useful finding is that the initiative is not equally valuable for everyone.

This may be appropriate when:

- Some groups appear to benefit more than others
- Need differs across roles, teams, units, or settings
- Uptake is strongest in particular subgroups
- The initiative is effective, but only under certain conditions

In these cases, the best next step may not be universal rollout. It may be a more focused strategy. An initiative may be worth continuing, but for a narrower set of users, contexts, or levels of need. This kind of targeting can improve efficiency, reduce waste, and increase return on investment.

Stop or wrap up

In some cases, the evidence suggests that the initiative is not delivering enough benefit to justify continued investment.

This may be appropriate when:

- The initiative is not producing meaningful improvement
- The burden of implementation outweighs the benefit
- Uptake remains low despite reasonable efforts to improve it
- The initiative no longer fits current priorities or needs
- Another approach appears more promising

A decision to stop should not be viewed as a failed evaluation. In many cases, it is the result of a successful one. Identifying that an initiative is not delivering enough value is important organizational learning. It prevents continued investment in approaches that are unlikely to pay off and creates room for better alternatives.

The goal is better decisions, not positive findings

Not every initiative should continue. Not every pilot should scale. Not every well-liked program should be sustained indefinitely. Likewise, not every initiative that falls short should be discarded immediately. Some deserve refinement, narrowing, or a more strategic use case.

That is why these outcomes should not be framed simply as success versus failure. A decision to stop, redesign, or target an initiative more carefully can be just as valuable as a decision to continue or expand it. In all cases, the evaluation has done its job if it helps the organization make a better-informed choice than it could have made otherwise.

A useful closing question

At the end of the process, organizations should be able to answer:

- ***What did we learn?***
- ***What does that mean for this initiative?***
- ***What are we going to do next?***

If the evaluation makes that answer clearer, it has been useful.

Final Takeaway

By working through this guide, you have built a practical framework for evaluating healthcare initiatives and making more informed decisions.

Rather than starting with measures or methods, you have focused on the questions that matter most:

- ***What decision are we trying to make?***
- ***What kind of evidence do we need?***
- ***What data are available?***
- ***What conclusions can we reasonably support?***

This shift is particularly important for workforce well-being, engagement, and retention initiatives, where meaningful outcomes are often more difficult to capture using traditional operational or financial metrics alone.

The goal of evaluation is not to prove that every initiative works. It is to understand where value is being created, for whom, and how efforts should evolve over time.

Done well, evaluation provides a stronger foundation for deciding whether to:

- Continue an initiative
- Improve it
- Scale it
- Target it more strategically
- Or stop investing in it

As healthcare organizations continue to invest in workforce programs, digital tools, analytics, and AI-enabled solutions, the need for thoughtful evaluation will only grow.

This guide does not explain how to approach personalized evaluation in detail. However, it points to an important next frontier in evaluation. Increasingly, organizations are looking beyond average results and asking more targeted questions: Who is benefiting? Who is not? What interventions work best for different individuals, teams, or settings?

As data sources become richer and healthcare organizations seek more targeted, actionable insights, there may be increasing value in partnering with organizations that specialize in data analytics, workforce intelligence, predictive modeling, and evaluation science.

These approaches can help organizations move beyond measuring whether an initiative worked overall to understanding where it created value and how to improve it.

The future of evaluation is not simply measuring more. It is measuring more intelligently so that healthcare organizations can make better decisions, invest resources more effectively, and improve outcomes for both the workforce and the patients they serve.