

# SCALING PLAYBOOK

For Achieving Durable Impact at Scale

care®




# SCALING PLAYBOOK

**For Achieving Durable Impact at Scale**

Developed in Partnership with:

IDinsight

Contents


**Glossary**

**Introduction**

09. About the Scaling Playbook
10. What the Playbook Covers
11. How to Use This Playbook
12. Why is Scale Essential to Achieving CARE's Impact Goals?
13. Underpinnings of This Playbook
16. An Evolved View of Scale

**Scaling Phases**

27. Introduction to Scaling Phases
28. Scaling Journey: Key Phases and Tools

**32. Phase 1: Review**

33. Phase 1 Review Activity: What to Expect
35. Evidence Review
39. Comprehensive Theory of Change Review
41. Competitive Landscape Assessment
45. Cost Review
47. Scaled Theory of Change Design
48. Enabling Environment Review

**54. Phase 2: Test**

55. Phase 2 Test Activity: What to Expect
56. Identify Testable Variants and Field Trials
57. Test Lean Variants through Field Trials
59. Final Design Testing and Multi-Site Roll Out

**63. Phase 3: Scale**

64. Phase 3 Scale Activity: What to Expect
65. Systems Orchestration Process

**69. Best Practices for Designing for Scale from the Start**

**71. Conclusion**

Appendix

## Appendix


**72. Phase 1: Review Activity**

73. Worksheet 1.1: Evidence of Impact Review
75. Worksheet 1.2: Theory of Change Review
79. Worksheet 1.3: Competitive Landscape Assessment – Developing a Longlist and Shortlist
82. Worksheet 1.4: Cost Review
84. Worksheet 1.5: Scaled Theory of Change Design
86. Worksheet 1.6: Enabling Environment Review

**88. Phase 2: Test**

89. Workshop Session 2.1: Identify Testable Variants of the Scaled Theory of Change
90. Worksheet 2.2: Test Lean Variants through Field Trials
93. Worksheet 2.3: Final Design Testing and Multi-Site Roll Out

**96. Acknowledgments**

# Glossary

| Term | Definition |
|---|---|
| **Capability** | The resources, procedures, and circumstances required to turn capacity into real action and outcomes. |
| **Capacity** | The technical knowledge and skills to do something. |
| **Catalytic impact** | Impact that scales exponentially when host-system actors sustainably scale solutions through local systems, utilizing their own resources, without direct intervention by CARE. |
| **Competitive landscape assessment** | Scan the field for similar solutions to understand if they are more impactful, simple, affordable, cost-effective, or scalable than the CARE solution. Use this assessment to prioritize the proven solution with the greatest impact, which the system will be best able to own and sustain in the long run instead of preserving any one organization's model." |
| **Comprehensive theory of change review** | Pressure test your existing theory of change with scale in mind. Explore how your intervention achieves outcomes, challenge key assumptions, and determine which elements must be preserved, adapted, or dropped to remain effective at scale. |
| **Core model** | The fundamental parts of an intervention—elements essential to achieving impact. This differs from the full model, which may include optional or flexible components that can be trimmed with minimal impact. |
| **Direct delivery** | A traditional grant-funded approach, where CARE designs and implements a program in a specific location over a defined project cycle, impacting a limited number of people with no clear path to sustain or expand reach beyond the organization's footprint project cycle. |
| **Doer** | The actor or actors  who will implement the solution. |
| **Doers at scale** | Stakeholders and institutions responsible for sustaining and scaling the implementation after CARE's (and other external partners') direct role ends. They may include government, civil society, and private-sector actors at central, regional, and local levels. |
| **Enabling environment** | The combination of political, economic, social, cultural, and environmental conditions that support an intervention's success, shaped by local actors and institutions. |
| **Enabling environment review** | Analyze the system in which scaling will occur—its policies, stakeholders, incentives, and barriers—to understand whether and how systemic realities will impact scaling efforts. |
| **Evidence review** | Evaluate whether your intervention is supported by substantial evidence of impact and if that impact is likely to hold at scale. Disaggregate data, assess external research, and identify gaps that must be addressed to build confidence in scale readiness. |
| **Host system or host country system** | The overarching governmental and non-governmental structures and institutions—at local, regional, and national levels—that comprise the system through which scaling will happen. |
| **"Hothouse flower"** | Interventions that are designed with intensive attention and resources typical of grant-funded pilot settings but not available to local systems that must sustain and scale them post-funding. |
| **Impact at scale** | Sustainable scale that leads to exponential impact, reaching the majority of a target population, beyond the limited reach of direct implementation. Achieving this requires host country systems to act as "doers at scale" and "payers at scale." |

| Term | Definition |
|---|---|
| **Intervention** | Any project, program, approach, solution, or initiative designed to drive meaningful, measurable impact. |
| **Local system stakeholders or actors** | Institutions and individuals embedded within and comprising the local system. Distinguishing between stakeholder actors that are institutions vs. individuals is crucial because though institutions change infrequently, actors within them can change frequently, and influencing each requires different strategies. |
| **Optimal fidelity model** | A streamlined version of an intervention that preserves core components essential for impact while simplifying the solution and reducing costs to enhance scalability by local actors. |
| **Payer** | The actor or actors who will fund delivery of the solution. |
| **Payers at scale** | Institutional entities that will fund an intervention as system integration and scaling occur. There can be multiple payers (government, civil society, private sector, international) at all administrative levels. |
| **Replication** | When CARE and its partners implement an intervention that has been successful in one context, adapting it for new settings and directly delivering it using grant funding, thereby increasing its impact linearly, through the organization's own footprint. |
| **Scaled theory of change design** | Design a revised theory of change that reflects how the intervention would function at scale, anticipating trade-offs and forecasting likely outcomes based on a leaner or adapted model. |
| **Scaling pathways** | Scaling pathways are defined by who will have ownership and/or responsibility for the solution over time (the "doer") and who will take financial responsibility for covering program costs (the "payer"), enabling system integration and sustainable scale. |
| **Sustainable scale** | Sustainable scale is achieved when other actors—particularly governments, the private sector, and civil society organizations—adopt, implement, and pay for a proven solution, impacting a majority of the target population facing the problem. Achieving sustainable scale requires deliberate steps to equip local systems as both "payers" and "doers" at scale. |
| **System integration** | Achieved when the intervention is fully embedded in the host country system, and when that system has committed to (1) allocating the resources, planning effort, and political capital for effective, continued delivery of the intervention; and (2) training, supervision, and ongoing support to sustain, refresh, and evolve the intervention over time. |
| **Systems Orchestration** | A strategic process that transfers the "doing" and "paying" of an intervention from external actors (e.g., CARE) to permanent local institutions—government, civil society, and the private sector. This multi-year process builds local ownership and capacity. |
| **Systems Orchestrator** | As a Systems Orchestrator, CARE delivers networked impact by strengthening the conditions and relationships that enable sustained delivery and financing of proven models—playing a supportive role in identifying, aligning, and strengthening the people, institutions, and conditions necessary for local stakeholders to adopt, fund, and deliver solutions at scale. |
| **Systems-level impact** | The result of efforts to change entrenched systems so they work better for people—by strengthening institutions, shifting social norms, influencing policy, and supporting social movements—advancing structural changes that improve lives beyond CARE project participants. |



# About the Scaling Playbook

## Aim of the Playbook

The 2030 Sustainable Development Goals (SDGs) are unlikely to be met unless the development sector significantly changes how it generates impact with and on behalf of communities living in poverty and in conflict-affected regions. CARE's ambition is to support 200 million people from the most at-risk and excluded communities to overcome poverty and social inequalities by 2030. Scaling CARE's reach, impact, and influence through governments, the private sector, civil society partners, and local/global networks is crucial to solving the intensifying crises faced by those we serve. To do this well, we must rapidly improve both what we do and how we do it.

Over time, CARE has implemented hundreds of donor-funded programs to reach millions of people in over 100 countries every year. But no organization alone can address the scale of challenges facing those living in poverty globally. Grant funding, which is highly variable and project-driven, is also inadequate given the magnitude of need. With these realities in mind, the CARE of the future must work beyond its own footprint—supporting other doers and payers to adopt and implement proven models sustainably.

This requires asking hard questions about the interventions we develop, test, and aim to scale:

- What is the most simple, affordable, and cost-effective solution to address a critical development challenge for a majority of the population experiencing it?
- Once rigorously tested and proven, who will implement (doer) and fund (payer) continued delivery of the solution beyond project cycles?
- What resources and capabilities are needed for doers and payers to successfully adopt and sustain the solution?

This Playbook is the first in a series designed to support CARE's evolving role—from direct implementer to enabler of impact at scale. This evolution requires us to not only develop scalable solutions but also, critically, to support networks of local stakeholders to own and operate them. In this new role, CARE increasingly acts as a Systems Orchestrator—a term used to describe how CARE facilitates others to lead. As a Systems Orchestrator, CARE delivers networked impact by strengthening the conditions and relationships that enable sustained delivery and financing of proven models. This requires CARE to play a supportive role in identifying, aligning, and strengthening the people, institutions, and conditions necessary for local stakeholders to adopt, fund, and deliver solutions at scale. This Playbook doesn't fully define or operationalize the Systems Orchestration role yet—that is the focus of a subsequent Playbook—but it's designed to prepare solutions for that transition.

This Playbook will help you:

- Assess your intervention's readiness to scale.
- Apply evidence-based tools and decision criteria in determining readiness.
- Inform the transition from pilots and programs to scalable solutions across diverse systems.

***Note:*** *This Playbook incorporates tools developed in partnership with IDinsight. Some visuals and worksheets are used with permission and cited where applicable.*

# What the Playbook Covers

The Scaling Playbook provides practical guidance and tools rooted in CARE's decades-long experience in taking proven models to scale. In this Playbook, you will find:



**An Introduction to Sustainable Scale**
We begin by defining what "sustainable scale" means for CARE—how it differs from direct delivery or replication—and how sustainable scale and systems-level impact are critical to CARE's 2030 impact ambition.



**Practical Processes and Tools**
We unpack essential concepts—scaling pathways, readiness assessments, cross-phase considerations—and the core methods you'll use to evaluate and prepare your intervention for scale.




**A Step-by-Step Framework**
Our three-phase approach (Review, Test and Scale) breaks the scaling journey into clear stages, each with decision points, activities, and diagnostics to move from proof-of-concept to full Systems Orchestration.



**Interactive Materials**
To link theory to your context, you'll find worksheets, assessments, and guided reflection questions designed to spark creativity, align stakeholders, and plan actionable next steps at every phase.




**Examples from CARE Programs**
We intersperse real-world case studies—from VSLA government partnerships to Farmer ID Card adoption in Nepal—to illustrate how teams have applied these tools to achieve catalytic impact.

The guidance, activities, and tools in this Playbook are designed not only to stimulate dialogue, but to guide your efforts to adapt and ultimately design scalable interventions that local actors—the eventual "doers and payers at scale"—will be able to integrate into their own systems and structures delivering enduring impact at scale.

# How to Use This Playbook

This Playbook is designed as an active facilitation tool not just a reference. Use it to guide your team through CARE's three-phase scaling journey (Review, Test, Scale) with structured steps, co-creation activities, and reflection prompts. It will help you navigate the complexity of adapting proven models for sustainable scale, while keeping your process collaborative, context-driven, and grounded in the core vision of local ownership.

Each section builds on the last, taking you from assessing readiness to piloting a lean "optimal fidelity model", and finally mobilizing host-system actors to own and sustain scaled impact. Whether you're launching a new scaling initiative or deepening an existing one, this Playbook will steer your team through collaborative planning, partner alignment, and system-level decision-making.

## What You'll Find Inside

- **Phase Overviews:**
  Set out the purpose, core questions, and decision checkpoints for each phase of the scaling process.

- **Activity Pages:**
  Show you how to run workshops and exercises—like Evidence Reviews, ToC deep dives, and micro trials—with clear instructions, materials lists, and facilitation tips.

- **Core Content:**
  Provides concise, plain-language explanations of key scaling concepts (e.g., "simple, affordable, cost-effective," scaling pathways, enabling environments). Always read this first to ground your team's mindset and terminology.

- **Worksheets and Reflection Questions:**
  Offer structured templates and prompts to surface critical insights, align stakeholder perspectives, and track progress against your scaling hypotheses.

- **Appendix:**
  Houses all editable tools—detailed worksheets, sample templates, and background materials—organized by phase and labeled clearly (e.g., Part 1, Part 2) when multi-page resources are included.

## Keep This in Mind

Scaling is not about checking boxes. It's about helping build interventions that can embed and thrive within real systems—with real constraints, actors, and opportunities. Use this Playbook to navigate that complexity, build with the end goal in mind, and design for long-term resilience.

# Why is Scale Essential to Achieving CARE's Impact Goals?

As noted, CARE's impact ambition—to support 200 million people from the most at-risk and excluded communities to overcome poverty and social inequality by 2030—cannot be achieved through direct implementation alone. While CARE continues to deliver hundreds of programs across more than 100 countries, this Playbook is part of a broader shift: one that emphasizes scaling through networks and partnerships, not just projects.

Durable impact at scale is only possible when proven interventions are fully embedded within a country's existing systems—led and implemented by networks of local actors, using local resources. This makes localization inseparable from scale. For an intervention to scale beyond CARE's footprint, it must fit within the real-world constraints of government agencies, private sector actors, and civil society organizations—simple enough to deliver, affordable enough to fund, and cost-effective enough to justify at scale.

The first step in supporting local actors to adopt, adapt, and own proven solutions is to ensure that solutions are built to enable this kind of handoff and sustained delivery.

A solution that is ready for scale, and ready to be delivered by others must be:

- Simple, affordable, and cost-effective
- Backed by evidence of impact and of scalability
- Aligned with the needs and capacities of potential doers and payers at scale

When interventions integrate scale and localization from the outset, they enable solutions that are not only impactful, but built to last. By supporting local host system actors to become the long-term doers and payers at scale, and then stepping back as they take full ownership of the intervention, CARE can achieve exponential impact that matches the scale of the challenges our target populations face.

# Underpinnings of This Playbook

CARE-USA's Impact at Scale (I@S) team was established in late 2020, with the specific directive to focus on "Scaling and Adapting Proven Models". The team's mandate was to design a process that CARE could use to identify solutions with a strong evidence base that had the potential for sustainable scale. This would enable solutions to be prioritized, and resources to be allocated accordingly.

To guide this work, CARE adopted the "Six Stages of Scaling Innovations" framework from the International Development Innovation Alliance (IDIA). This framework offers a clear path: from idea to proof of concept, to small-scale replication, and ultimately to wide-scale adoption by system actors like governments and civil society. While this six-stages framework is well-suited to taking new innovations to scale, it is less useful for assessing the scalability of existing solutions.

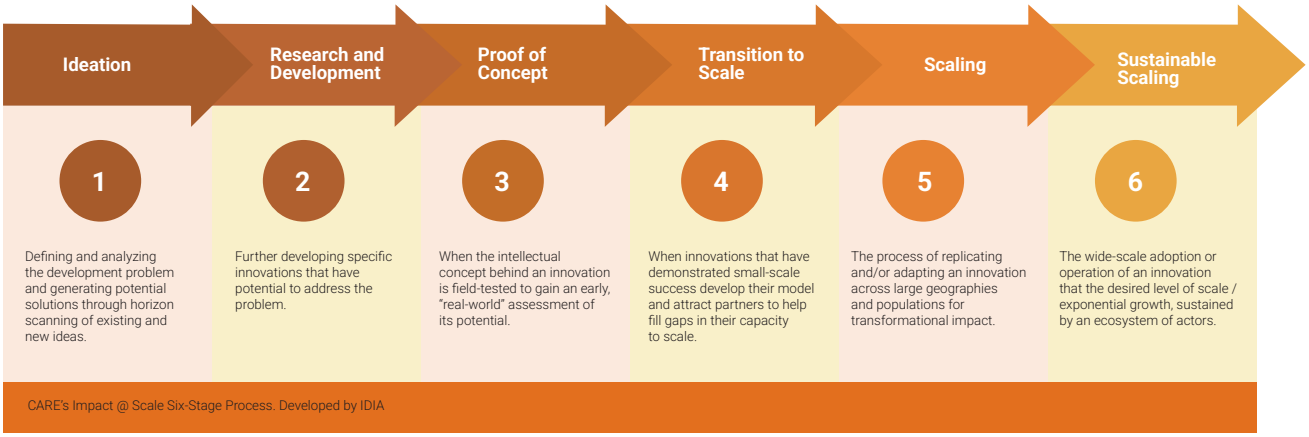
Yet, most development organizations—including CARE—have portfolios of existing legacy solutions, which are understandably their first entry point in the search for scalable solutions. Early on, CARE expected that many of its successful models—grant-funded interventions that had demonstrable impact, donor support, and multiple implementations—were ready to transition to scale. The assumptions underpinning this expectation were:

1. CARE's programs are proven and impactful.
2. They are simple, affordable, and cost-effective enough for others to adopt.
3. They outperform comparable solutions in the sector.

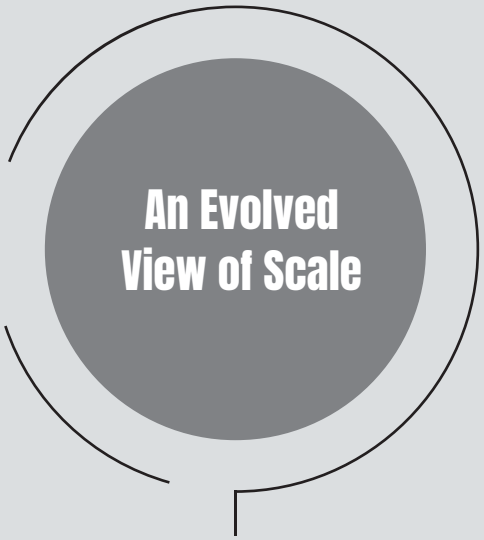
Partnering with reputed evidence and decision-support organization IDinsight, the Impact at Scale team conducted rigorous internal assessments of 17 interventions to test these assumptions from 2021-2024. This work revealed that many solutions lacked robust evidence or were too complex or expensive to be adopted by external actors.

This prompted a shift in how CARE approaches scale—highlighting the importance of moving from assumption to evidence-based assessment, and from ownership to orchestration. This Playbook compiles insights from that journey, offering a practical guide to help teams across CARE assess the scalability of ongoing programs and identify paths forward.

Figure 1: CARE's Impact @ Scale Six-Stage Process.




care


An Evolved
View of Scale

# An Evolved View of Scale

## 1. Defining Scale

Scale is a very loosely defined term. For some it means increasing footprint or expanding to new areas; for others it is about achieving greater saturation, depth, and/or breadth of impact at the same reach.

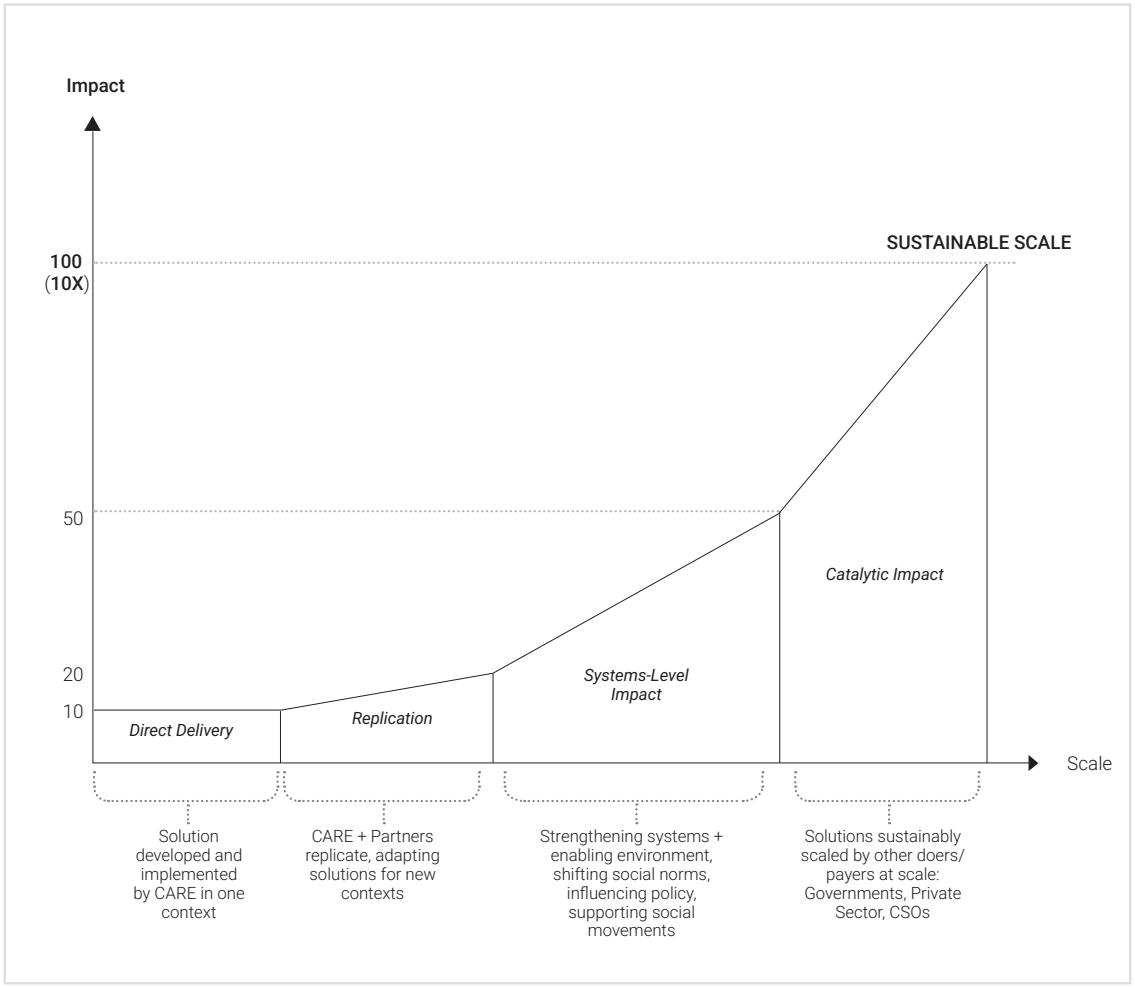
CARE and most other development organizations have broadly defined scale as reaching more people with meaningful impact. But as the scope and urgency of the world's problems intensify, CARE has sharpened that definition, honing in on the *sustainable scale* that is achieved when we work beyond CARE's projects, footprint, and donor funding. When other actors—particularly government, the private sector, and civil society organizations—adopt, implement and pay for a proven solution to be scaled, impacting a majority of the target population facing the problem, we have achieved sustainable scale.

In CARE's vision for sustainable scale, our role shifts from direct implementation to Systems Orchestration—a strategic process that supports the transfer of intervention ownership to host system actors, including government, civil society, and the private sector. This transition requires these local actors to fully plan, fund, implement, monitor, and adapt solutions within their ecosystems. Crucially, this process is inseparable from localization: scale cannot be achieved unless solutions are integrated into and led by host systems. True localization and scale happen together when CARE supports these networked actors as doers and payers, while stepping back to enable their long-term leadership and accountability.



The graph below conceptually captures why a focus on sustainable scale is essential to CARE's ability to achieve its ambitious impact goals, illustrating the connection between how we work and the impact we achieve.

Figure 2: Sustainable scale



## Direct delivery

Direct delivery is the traditional way that development organizations, including CARE, have operated. It involves designing and implementing a grant-funded program in a specific place, usually over a 3-5 year project cycle. This way of working generally impacts limited numbers of people with no clear path to sustain impact or increase reach beyond project cycles. With direct delivery, impact is flat, limited by CARE's footprint in the implementation locale, and the grant funding available.

### Direct Delivery Example: Women's Empowerment and Agricultural Production in Burundi

Over 2015-2020, CARE and partners implemented the Win-Win project in Burundi, with a grant of $2.6M from the Gates Foundation. CARE engaged with savings and community groups to deliver a combination of agricultural training and support through the Farmer Field and Business School (FFBS) approach, and the EKATA approach (Empowerment Through Knowledge and Transformative Action). Aiming to enhance women's empowerment through household reflection, dialogue and actions toward equality, while also promoting food security, nutrition, and economic well-being, the project engaged 9,911 participants directly and 37,000 indirectly.

Win-Win achieved impressive impact while also demonstrating the limitations of a direct delivery model—revealing a distinct difference in impact between groups engaged in the most intensive version of the intervention vs. those engaged in the lighter-touch version. Farmers in the intensive group, which received the highest level of direct engagement, showed a 74% increase in rice production and a 5:1 ROI, versus the 2:1 and 3:1 returns for lighter-engagement groups. In this intervention, direct delivery proved to be highly effective but also relatively expensive.

## Replication

Replication occurs when CARE and partners take interventions that worked well in one place, adapt them for new contexts, and then directly implement them in those contexts using grant funding. This way of working increases the scale of impact in a linear way, limited by CARE and partners' footprints in the implementation locales, and by the grant funding available.

### Replication Example: Accelerated Education for Out-of-School Girls

Strengthening Opportunities for Adolescent Resilience (SOAR) is a multi-country initiative working with Ministries of Education to provide accelerated education for out-of-school adolescents, particularly girls, enabling them to acquire key academic and life skills. In addition to building their capacity to read and do math, the SOAR curriculum includes leadership skills development, financial literacy, and adolescent sexual and reproductive health information. Adolescents who complete the program are able to transition into formal school or use their new knowledge and skills to enter a vocational training program or seek more meaningful work. SOAR was first developed and implemented in India in 1999 before being replicated in seven additional countries: Afghanistan, Malawi, Nepal, Pakistan, Somalia, Zambia, and Zimbabwe.

While the basic SOAR approach remains the same across geographies, its curricula are adapted for each country-specific context and designed in partnership with Ministries of Education to match local requirements for certification and transition to formal schools. Among other positive outcomes, the project has seen a significant decrease in likelihood of program participants marrying young; high rates of success for program participants in transitioning to formal school, including a 95% pass rate on state board exams in India; and increases in income for program participants across geographies. The replication method of scaling has allowed SOAR to reach 4.1M participants in eight countries.

## Systems-Level Impact

Systems-Level Impact is the result of work CARE does to change entrenched systems so that they work better for people, by addressing the root causes of structural inequalities for women, girls, and underrepresented communities. This includes efforts to strengthen local institutions and enabling environments, shift social norms, influence policy, and support social movements. Because this foundational work advances structural changes that improve the lives of people beyond CARE project participants, it accelerates the scale of impact.

## Systems-Level Impact Example: Indigenous Participation in Peru

CARE Peru is working through projects such as "El Agua Nos Une" (Water Unites Us) and "Paisajes Amazónicos Vivos" (Living Amazonian Landscapes) to ensure indigenous peoples' voices are heard in the governance of natural resources in the Andes-Amazon region. The aim of these projects is to build capacity at the institutional level and at the population level to improve communities' technical, administrative, commercial, financial and organizational management capabilities. El Agua Nos Une applies a systemic approach to help local government officials, indigenous leaders, and private sector water actors recognize and acknowledge each other and support evidence-based, collaborative decision-making.

Ultimately, strengthening the participation of 200 indigenous communities in forest and water governance will improve incomes by increasing their access to non-timber value chains, support conservation efforts of nearly 5M hectares of Amazonian rain forest, and ensure key inter-cultural dialogue and insights from communities historically left out of these decisions shape the management of their ancestral environment.

## Catalytic Impact

Catalytic Impact is the result of other doers and payers at scale—governments, private sector, CSOs—sustainably scaling solutions through local systems, using local system resources. The sustainably scaling intervention may have been initially introduced or identified by CARE, and piloted by CARE to demonstrate impact, but once the host system takes ownership of it, impact scales exponentially without direct action by CARE.

## Catalytic Impact Example: Scaling VSLA Through Government

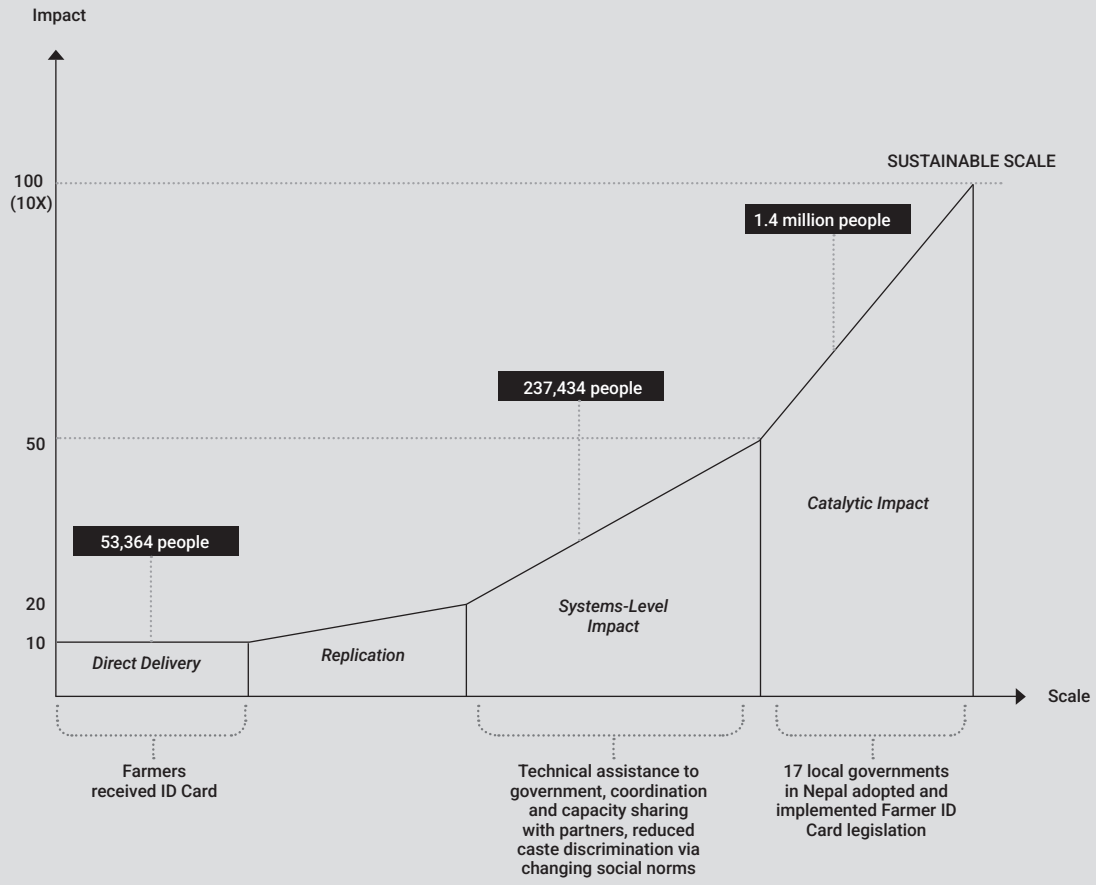
After heavily investing in direct formation of VSLAs for many years, CARE's VSLA team shifted focus to catalyzing localized uptake and ownership of VSLA programming by national governments and stakeholder coalitions. CARE has partnered with governments in Uganda, Cote d'Ivoire, Nigeria and Rwanda to integrate VSLAs into government-funded social protection programming, and to identify government budgets to fund VSLA scaling. This targeted government engagement has resulted in government-funded programs in Uganda and Rwanda, with similar programs and funding streams in design for Côte d'Ivoire and Nigeria.

This expansion of the VSLA model through national governments is expected to impact approximately 1.3 million women in Cote d'Ivoire, 440,000 women in Uganda, 700,000 women in Rwanda, and 20 million women in Nigeria.

# CASE STUDY: Achieving Impact at Scale in Nepal

The case of CARE-Nepal's work on Farmer ID (FID) Cards clearly illustrates the relationship between how we work and the impact we achieve. In Nepal, landless, tenant and small farmers were not receiving government services and subsidies designed to help them, often because they lacked identification. Without proper identification, these farmers were denied a voice in policymaking and access to vital resources. Farmer ID Cards would enhance access to timely and high-quality services, opportunities, and technology such as pesticides, fertilizers, and seeds.

Figure 3: Sustainable Scale of CARE-Nepal's Farmer ID Program




- **Direct Delivery:** In 2020, CARE-Nepal piloted and validated an FID initiative, leveraging internal resources and funding from donors and local government. Through this effort 53,364 smallholders in 17 districts of Nepal secured the IDs needed to access the government support they were entitled to. During this pilot period, CARE also delved into root causes, identifying the systemic challenges that caused smallholder farmers to not have the IDs and support they needed.

- **Systems-Level Impact:** To address those systemic failures, CARE Nepal supported long-term local partners including the National Farmers Groups' Federation (NFGF)—which represents 5,550 farmers groups nationwide—forming a coalition that actively advocated for policy changes to increase farmers' identity, access to services, and social security, while emphasizing agriculture's dignity. These systems change efforts have contributed to improved livelihoods, health and social equality for 237,434 people. Dalit and landless women in particular have benefitted from access to agricultural resources and services, improving efficiency and accountability. By shifting social norms about caste and landless farmers, discrimination has also decreased.

- **Catalytic Impact:** As a result of this locally-led work, 17 local governments in Nepal adopted the FID model into their agricultural policies and budgets, providing access to agricultural services, subsidies and training to farmers. A significant number of landless and underrepresented farmers are anticipated to benefit from municipalities implementing the FID card system. Modeling estimates the catalytic impact of FID to be 1.4 million people.

## 2. Not Everything Will Scale

It is important to recognize that though relatively few interventions designed and implemented using grant funding have the potential for sustainable scale, this does not mean that anything that cannot scale is a bad intervention. Not everything can, should or needs to scale. Many solutions—especially those related to changing social norms or advocating for the rights of underrepresented communities—may always be too expensive or low priority for governments and private sector actors to take on. Yet, solving these foundational challenges is critical to dismantling the power structures that perpetuate inequality and poverty. For challenges such as those, there will always be a need for NGOs and philanthropic dollars to fill the gaps. This being the case, it is imperative that organizations like CARE continue to replicate impactful interventions in specific thematic areas using grant funding to lift up at-risk communities, collaborating as equals with local partners throughout.



## 3. Scaling pathways

The global community of scaling practitioners has long recognized that there is more than one road to scale. As such, identifying and planning for feasible scaling pathways is a core element of many existing scaling methodologies. However, while scaling methodologies do share some similarities in describing the typology and terminology of scaling pathways, they are also full of contradictions which can create confusion for intervention teams.

For the purposes of this Playbook, we define scaling pathways in terms of who will have ownership and/or responsibility for the solution over time (i.e. become the main "doer") AND who will take the financial responsibility for covering program costs (i.e. become the main "payer"). Through these pathways, networks of host system partners adapt and embed CARE interventions into their own implementation approaches, services, and budgets, to achieve "system integration" and deliver sustainable scale.

The doer and payer can be the same or different actors. At the simplest level, the scaling pathways we generally think of include the following host system actors, or collaborative networks that include all of them:

1. **Government systems:** in this case, the government is the "doer" and government budgets are the "payer"
2. **Private sector:** in this case, a private sector actor is the "doer" and the customer (through market mechanisms) is the "payer"
3. **Civil society organizations:** in this case, community-based organizations or other local non-government partners are the "doer" and donors are the "payer"

Scaling through these pathways is an objective pursued by many but successfully achieved by relatively few. This is often because of the presumption that any intervention with measurable impact, which has expanded to multiple sites, and which donors have been willing to fund, is scalable. While such interventions are good place to start the search for scalable models, our work indicates that the majority of legacy interventions designed and tested through multi-year grant funding do not have the potential for sustainable scale, and either cannot or need not be retrofitted for sustainable scale.

## 4. Simple, Affordable, Cost-Effective

Designing and scaling proven solutions is a long, complex and dynamic process. The first instinct of large, established organizations like CARE is to scale existing solutions. As noted above, the presumption is that because these solutions have achieved positive outcomes and donors are willing to fund them, they are ready to be sustainably scaled by other actors.

However, two realities run counter to this perspective:

- The challenges that the development and humanitarian sector seeks to address are almost always complex, the result of various economic, social, cultural, and environmental factors. To solve complex problems, CARE and organizations like it have traditionally developed and deployed holistic interventions that address these many underlying and interwoven factors. While the complex solutions that result may be impactful, and even replicable if CARE remains the doer and grant funding remains the payer, complex solutions do not lend themselves to sustainable scale.

- The interventions of NGOs like CARE have traditionally been supported by large grants from institutional funders, which allow CARE to bring a very high level of capacity, capability and resources to direct implementation. But these same levels of capacity,




capability and resources are usually not available to the governments and market actors in low-income contexts who must adopt, implement, and pay for the solution for it to sustainably scale. Consequently, these interventions may not be scalable.

Yet, the instinct to turn first to CARE's own "proven" interventions in looking for scalable solutions is understandable. It is easy to fall into the trap of believing that every solution is scalable, that all solutions should scale, that scaling solutions is always the goal, or that solutions that can't scale are bad. This is not true, because there are many important development challenges that NGOs will always be best positioned to solve via replication, using philanthropic funding.

This is especially likely to be true, for instance, for interventions that target social norms, that benefit excluded or at-risk groups, or that generate a critical public good – which governments may not have sufficient funding to solve, and which markets may not have sufficient incentive to solve. There may also be instances in which an intervention is proven to be best-in-class by rigorous evidence, but is too costly for local systems to own and finance—cash transfer programs are a good example of this. In all of these cases grant-funded direct delivery or replication through an NGO's own footprint is not only acceptable, but potentially optimal and absolutely essential to ensuring that critical factors underpinning poverty and inequality are addressed.

## 5. Trade-Offs

As teams explore how to bring proven interventions to scale, they must often confront difficult design choices. At the heart of these choices lies a core tension: how to preserve the effectiveness of an intervention while adapting it to the realities of local delivery systems. Given that most of the legacy grant-funded programs being assessed for scalability are complex and costly, maintaining their quality and fidelity at scale can be difficult. Interventions developed under ideal grant-funded conditions may not automatically translate to sustainable scale—particularly when handed over to government, market, or civil society actors operating with different levels of capacity, incentives, and funding structures.

Making a solution "scalable" often requires simplifying delivery models, reducing cost drivers, or streamlining implementation roles. However, the prospect of simplifying an intervention to align it with local capabilities, or changing it to make it more affordable, leads to the reasonable fear that the loss of quality and fidelity may be so significant that the adapted intervention yields little to no impact, or in the very worst cases, actually causes harm.

Yet, if these adaptations are grounded in clear design principles, evidence-informed decision-making, and continuous testing, they can increase the chances of long-term adoption and ownership by local actors. As long as meaningful standards for directionality and magnitude of an intervention's impact are established and closely monitored as it is adapted for sustainable scale, streamlining it to improve scalability may still be the best course of action. Because even if the adapted or redesigned model results in lower impact or smaller effect sizes than the original model, the more scalable model is likely to have durabilty greater reach. Ideally, we can identify a model that optimizes for simplicity, affordability, cost-effectiveness and impact – the optimal fidelity model.

Teams must be intentional about which aspects of an intervention are core to impact and must be preserved, and which can be adjusted without compromising outcomes. This demands a disciplined focus on what's truly essential, and a willingness to revisit design assumptions as the context evolves. Rather than viewing these trade-offs as losses, they can be framed as strategic decisions to increase relevance, durability, and reach. Scaling, then, is not about replication—it is about evolution. And navigating trade-offs well is what allows teams to move from proven pilots to interventions that can truly shift outcomes.

It's important to remember that as an intervention is adapted to fit with the funding and capacity limitations of host system actors, existing rigorous evidence from the more complex model becomes less relevant. Therefore, the new optimal fidelity model will most likely require further evidence generation to test hypotheses and assumptions, and to re-estimate effect sizes or impact. Having this data is critical to generating confidence in a program's scaled impact.

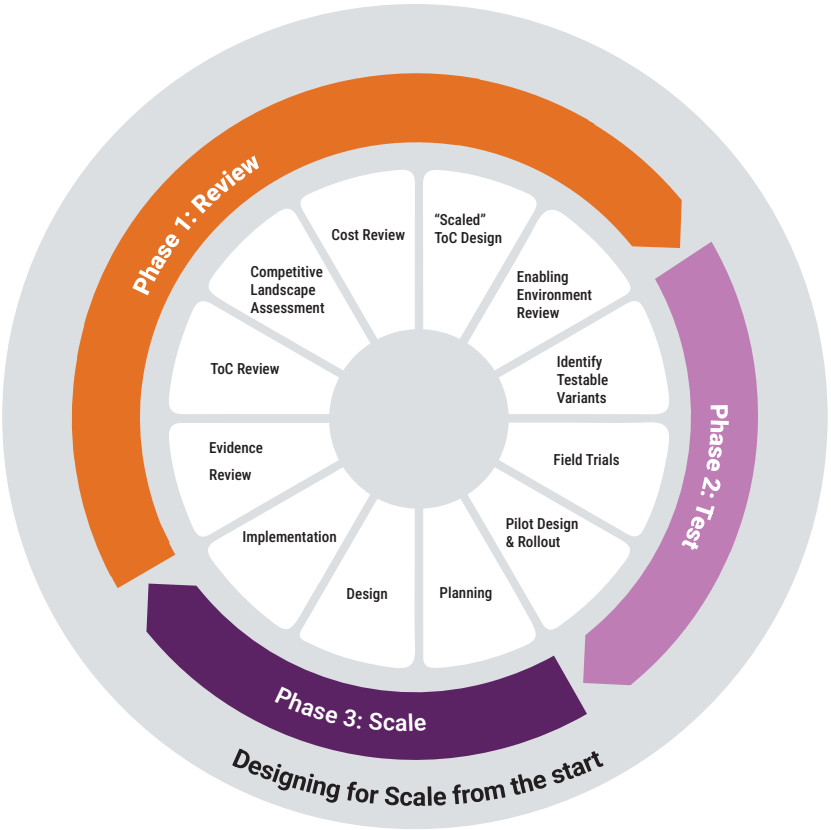

আমার
বাংলা বই

# Scaling Phases

The process of scaling is generally understood to have distinct phases, each with distinct challenges, targets, and validation opportunities. Understanding each of these phases—along with their definitions, activities and associated tools—helps teams structure and manage their process, develop a shared understanding of where different interventions stand in terms of scalability, and choose the right tools and questions at each stage. While scaling frameworks often present a linear process of phases and accompanying tools for conceptual simplicity, they almost always include a reminder of the fact that the scaling process is not linear, and that it is iterative. Activities and tools can and should be used non-sequentially and revisited over time. Furthermore, an intervention may have to return to an earlier phase in the scaling process—or exit the process altogether—if learnings from a later phase do not validate key hypotheses about the intervention at scale.

As noted previously, the six-stage process CARE originally adopted as its scaling framework was not designed to assess existing programs for scale, but rather to design and scale new innovations. The insights gained from efforts to scale existing CARE solutions led us to adapt that framework to something better suited to assessing the scalability of legacy interventions (Review, Test and Scale).

# Scaling Journey: Key Phases and Tools




MINISTRY OF HEALTH



# Phase 1: Review

# Phase 1: Review

This phase helps your team assess whether your intervention is truly ready for scale. It includes six critical tools that interrogate evidence, clarify logic models, explore the landscape of scale, and quantify financial feasibility. By completing these activities, teams can surface gaps, strengthen their solution, or pivot altogether. These tools do not assume readiness—they help test it.

## Overview

01 Evidence Review: Evaluate whether your intervention is supported by strong evidence of impact and if that impact is likely to hold at scale. Disaggregate data, assess external research, and identify gaps that must be addressed to build confidence in scale readiness.

02 Comprehensive ToC Review: Pressure-test your existing ToC with scale in mind. Explore how your intervention achieves outcomes, challenge key assumptions, and determine which elements must be preserved, adapted, or dropped to remain effective at scale.

03 Competitive Landscape Assessment: Scan the field for similar solutions. Identify what already exists, where your solution fits, and whether you should scale your model, strengthen it, or support a more scalable competitor instead. Based on these insights, decide whether to pivot your approach (make major changes), partner with an existing solution, or proceed with scaling as designed.

04 Cost Review: Compile detailed cost data (one-off vs. ongoing), conduct cost-effectiveness and cost-efficiency analyses, model projections under different scale scenarios, apply reverse benchmarking to set competitive cost thresholds, and incorporate qualitative willingness-to-pay signals to assess overall affordability.

05 Scaled ToC Design: Design a revised ToC reflecting how the intervention would function at scale. Identify which actors will own and implement each component, anticipate trade-offs, and forecast likely outcomes based on a leaner or adapted model.

06 Enabling Environment Review: Analyze the system in which scaling will occur—its policies, stakeholders, incentives, and barriers. Identify champions, funding pathways, and system-readiness indicators to inform your Systems Orchestration strategy in Phase 3.

## Phase 1 Review Activity: What to Expect

This phase guides your team through a structured assessment to determine whether your intervention is ready for sustainable scale. Phase 1 focuses on six foundational areas: reviewing the evidence base, pressure-testing your theory of change (ToC), understanding the competitive landscape, designing for scale, analyzing cost feasibility, and analyzing the enabling environment. These activities are designed to help your team make confident, evidence-based decisions—whether to proceed, pause, adapt, or redesign your solution.

The worksheets in this phase prompt rigorous reflection and structured analysis. They surface the assumptions, gaps, and opportunities that may affect your solution's ability to scale—and set the stage for building a rough draft of the optimal fidelity model to test in the subsequent phase.

### 1. Worksheet 1.1: Evidence of Impact Review

**Purpose**

Assess whether the solution is supported by robust, disaggregated evidence of impact and early signs of scalability. Review evaluation data, monitoring results, and implementation reports to determine whether your model consistently works—and if that success is replicable across different contexts.

### 2. Worksheet 1.2: Theory of Change Review

**Purpose**

Revisit and adapt your ToC with scale in mind. Pressure-test assumptions, identify logic gaps, and map causal pathways from intervention to systems-level change. Clarify which components of the model are essential to maintain fidelity and which can flex to local conditions.

### 3. Worksheet 1.3: Competitive Landscape Assessment

**Purpose**

Explore comparable solutions already in the field. Start with a broad list of peer interventions tackling the same problem, then narrow to a focused set for in-depth comparison. Analyze their effectiveness, adoption rates, cost structures, and institutional integration to understand how your model stacks up—and where pivoting, partnership, or proceeding might be needed.

## 4. Worksheet 1.4: Cost Review

**Purpose**

Assess the financial feasibility of scaling by: compiling and differentiating one-off versus ongoing operational costs; conducting cost-effectiveness analyses to calculate cost per unit of outcome and cost-efficiency analyses where impact data are limited; modeling how economies of scale, government integration, or technological innovations alter cost structures and comparing projections to sector budgets; applying reverse benchmarking to set competitive cost thresholds; and incorporating qualitative signals—such as past government procurement—to gauge overall affordability and priority.

**DIFFICULTY**
Medium

**WHAT YOU'LL NEED**
Printed worksheets,

Pens/markers,

Sticky notes

**PARTICIPANTS**
Project lead,

Technical advisor,

Meal lead

## 5. Worksheet 1.5: Scaled Theory of Change Design

**Purpose**

Redesign your ToC with a scaling lens. This version should reflect how impact will be achieved when implemented by local actors—not your organization. Incorporate what you've learned about enabling conditions, delivery constraints, and peer models. Clarify roles, assumptions, and pathways that support real-world scale.

## 6. Worksheet 1.6: Enabling Environment Review

**Purpose**

Assess the broader ecosystem in which your model would need to operate at scale. Consider policy alignment, availability of funding, capacity of local institutions, and other systemic enablers or barriers. Use this review to identify high-risk assumptions and shape your strategy for testing the model in the places that are most appropriate.

# Evidence Review

Before beginning to sustainably scale a solution, it is important to assess if it has the potential to scale in the first place. A solution's potential for scale is determined by a) evidence of impact and b) evidence of scalability.

## Step 1: Evidence of Impact

Often, the instinct of large non-profits running multi-million dollar programs is that our most celebrated interventions, that have reached large numbers of people, are the ones that have the greatest potential for scale. We assume these programs are impactful based on internal program evaluations that are often limited to baseline, midline and endline studies based on key performance indicators. But neither an intervention's reach nor its financial backing are sufficient evidence of its potential for scale. Internal program evaluations are also often not rigorous enough to confidently determine whether the intervention is responsible for any of the program outcomes.

The most compelling evidence of impact will always be a well-designed, externally valid randomized controlled trial (RCT) of the solution proposed for scaling. While RCTs are expensive, they are critical for definitively establishing proof-of-concept. The high-value data generated is worth its cost given the potential financial and reputational risk of scaling interventions that have limited impact or unintended negative consequences. RCTs should ideally be as generalizable as possible to be externally valid. For instance, an RCT that works with a cherry-picked population in a strong enabling environment may provide evidence of impact under optimal conditions, but will do little to prove scalability of impact in less favorable contexts.

RCTs are also valuable because they generally offer the least biased estimate of impact. While in a non-scaling context we might accept "directional evidence" from other methods, precise measurement is especially important when scaling is the goal. Without reliable impact estimates, for instance, it's impossible to reliably calculate cost-effectiveness, which is a critical component of scalability.

Where RCT evidence for the specific solution is unavailable, rigorous evidence for similar or analogous solutions may be instructive. However, the project team should carefully consider similarities and differences between their program and the evidenced interventions. In such cases, parallels between the theories of change, generalizability to context, and the amount of external evidence will be key.

The weakest evidence of impact will come from descriptive studies, simple pre/post analyses, or monitoring data that doesn't provide a representative sample of the communities and populations that are expected to benefit from the solution. And while qualitative evidence is important and valuable for giving full color and expression to the findings of a quantitative impact evaluation, and is often critical to understanding causal mechanisms and "why" an approach works, it is not sufficient on its own to establish evidence of impact.

### Evidence Insight: When RCTs Don't Fit Systems Change Work

*Traditional randomized controlled trials (RCTs) often don't apply to systems change and strengthening efforts due to their scale, complexity, and long-term nature. These interventions aim to shift institutions and infrastructure—making it difficult to isolate impact through standard treatment/control comparisons.*

*When high-quality, time-stamped administrative data is available, teams can use time-series, difference-in-difference, or synthetic control methods to evaluate population-level changes. Where such data is absent, population-representative surveys can help produce reasonable pre/post outcome estimates.*

*In many cases, a robust process tracing framework is the best fit. This combines:*

1. ***Implementation fidelity checks**—was the systems approach delivered as designed?*
2. ***Causal pathway analysis**—does the evidence plausibly link activities to observed change?*

*When well executed, these methods can offer strong circumstantial evidence of impact, especially where traditional RCTs are impractical.*

## Step 2: Evidence of Scalability

Ultimately, scalability is about whether the intervention is simple, affordable, and cost-effective enough for doers at scale to implement it with fidelity, and for payers at scale to be able to afford the at-scale implementation cost. Evidence of an intervention having been scaled already—adopted and implemented by a doer that is not the original implementing organization, and paid for by a payer that is not grant funding—can be a good initial indicator of scalability, but to be certain we must delve more deeply.

Subsequent activities and tools will allow teams to more deeply interrogate the scalability of their interventions. However, at this early stage of evidence review, there are a few aspects teams should consider:

1. ***Scaling pathway:*** It is helpful to identify the (hypothesized) scaling pathway early on. Some key questions that can help guide this exploration include:

- Could this solution or specific components of it be integrated into government systems? If so, which components, which departments, which budgets?
- Are there plausible incentive structures for private sector actors to offer the product or services that comprise the solution, or its components? If so, which private sector actors, and which "customers"?
- Could other civil society organizations be interested in replicating the intervention or specific components of it? If so, which components, which organizations, and what capacities would they need?

### Example: Cash Transfer Programs as a Dual Path Scaling Success

*While governments are often the preferred pathway for scaling due to their reach and legitimacy, they can also be the most challenging—success stories are few. Similarly, civil society offers promise but often lacks the sustained funding or infrastructure for true scale.*

*A notable exception is cash transfer programs, which have successfully scaled through both government systems and expansive civil society networks. Many governments now operate national social protection programs using cash transfers, while international NGOs have adapted the model for diverse contexts—reaching millions globally. This dual-track scaling illustrates how alignment with national systems and distributed networks can both play a role in achieving impact at scale.*

2. ***Sensitivity to changes in context and implementation quality:*** If there are any existing replications of the intervention (within CARE, or by others), review those to understand how sensitive the intervention is to changes in context or implementation quality. This may tie back to the above evidence review. For example, teams can consider the following scenarios and questions:

- If there is rigorous evidence of impact from multiple different contexts, are the impact estimates consistent across contexts or do they vary notably? Perhaps you are even looking at situations where in one place the evaluation found positive results, and in another it didn't. Such cases may indicate challenges to scaling the intervention across contexts.
- In addition to any rigorous evidence, one can also consider qualitative or descriptive evidence. Is there any qualitative evidence concerning the ease or difficulty of maintaining fidelity and implementation quality? Or do monitoring data suggest that the intervention has been replicated with fidelity? For example, replications might find that it is difficult to maintain the same level of participation in the program as in the original implementation. Or qualitative research may reveal that some of the key assumptions do not hold in a new context.
- Overall, if there is any existing evidence that an intervention's ToC and impact is highly sensitive to changes in context and differences in implementation quality, this may indicate essential limitations to its scalability.

### Example: Cultural Anchors and Replicability – United for a Better Life

*Some intervention designs signal their potential scalability—or limitations—based on how tightly they're woven into local context. For example, the United for a Better Life (UBL) GBV prevention program in Ethiopia was built around the cultural practice of the coffee ceremony, a core element in its ToC.*

*While this design rooted the intervention deeply in community life and norms, it also raises scalability questions. The Ethiopian coffee ceremony does not exist in many other cultural settings, and true parallels may be hard to find. Adaptations of UBL must identify contextually relevant communal rituals in new geographies—and test whether they can host the same transformative dialogue and outcomes.*

3. ***Cost-effectiveness:*** Cost-effectiveness describes the relationship between impact and cost. Cost-effectiveness improves as impact increases, and/or as costs decrease. As part of the evidence review, teams should ideally look for existing cost-effectiveness data. Formal cost-effectiveness assessments require rigorous impact data and granular cost data - often both, or at least one of these required inputs is missing. However, should a formal cost-effectiveness assessment be available, this can provide compelling evidence that an intervention may be scalable, if compared to other solutions targeting the same outcomes (see competitive landscape assessment and cost review discussions).

   Cost-effectiveness is necessary, but not sufficient evidence of scalability. For instance, when cost-effectiveness is achieved by reducing costs without maintaining minimum quality standards, the impact of an intervention that is highly sensitive to implementation quality may suffer. Alternatively, an intervention may be impactful and cost-effective but still unaffordable for the target payer at scale.

**Example: Affordability Constraints—Indashyikirwa in Rwanda**

*The GBV prevention intervention Indashyikirwa had strong positive impacts on direct program participants in its original, high-fidelity implementation. However, a cost-effectiveness study estimated that it would take up to $70 million to implement it at scale in Rwanda, compared to the responsible line ministry's annual budget of ~$2.5 million.*

## Comprehensive Theory of Change Review

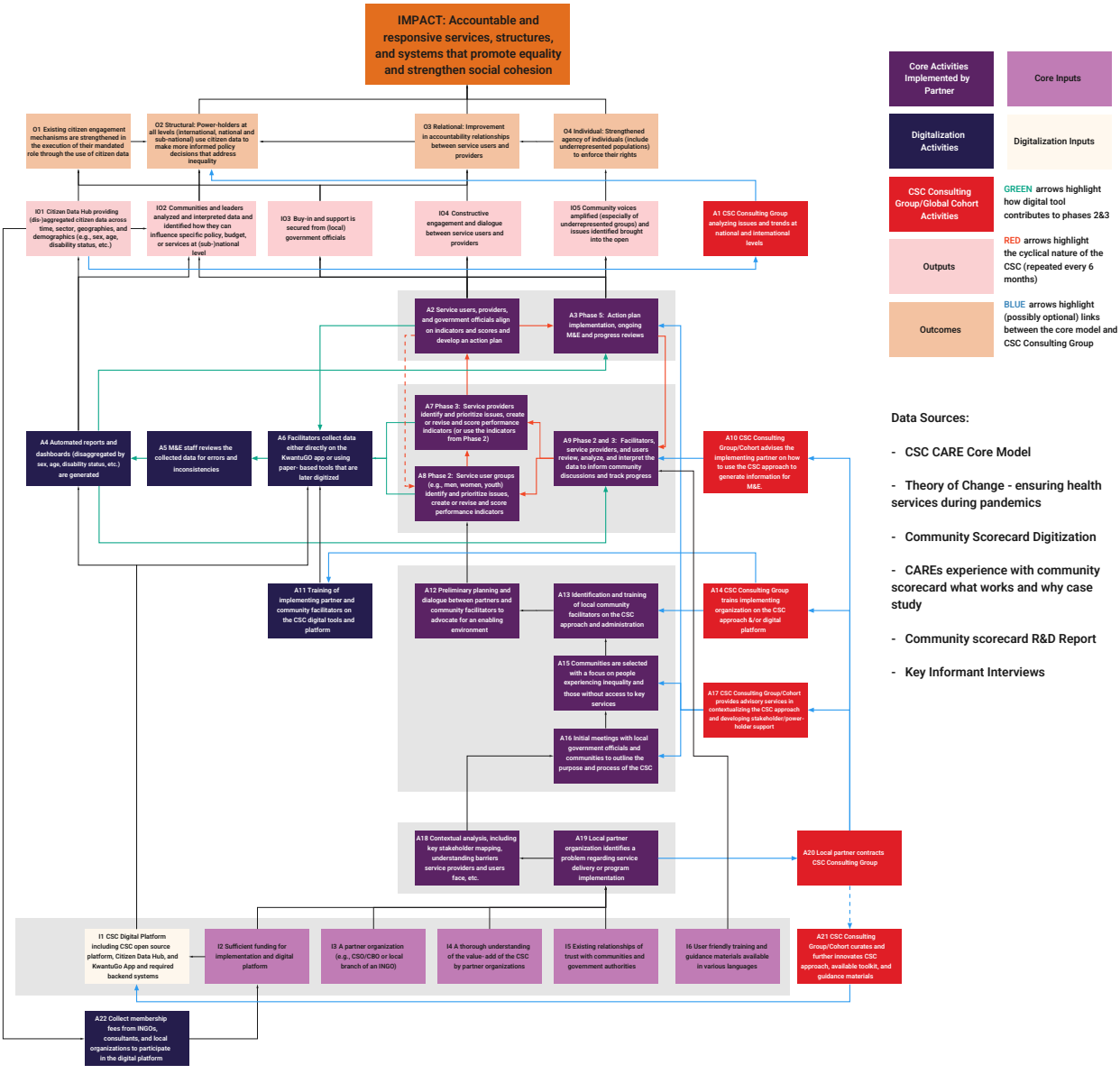
When preparing for scale, teams must move beyond a high-level or outdated ToC for the status quo intervention and instead articulate a detailed, current version that clearly maps how the intervention leads to its intended impact. This ToC should outline all essential inputs, activities, outputs, intermediate and final outcomes—along with their interdependencies and underlying assumptions.

Developing a comprehensive ToC involves several key steps:

1. **Identify the Desired Change:** Begin by clearly defining the long-term goals or final outcomes that the program aims to achieve. This involves understanding the ultimate impact you want to create.
2. **Conduct Backwards Mapping:** Work backwards from the desired change to identify the intermediate outcomes, outputs, activities, and finally inputs required to reach the final goal. This step helps in understanding the sequence of changes needed.
3. **Identify Impact Pathways and Causal Links:** Define the pathways to impact, using arrows to show the connective causal links between inputs, activities, outputs, outcomes, and impact. Causal links represent a clear direction (though they may be circular), and allow you to identify new causal pathways in the ToC iteratively as you learn more.
4. **Document Assumptions:** Clearly document the assumptions underlying the causal links in the ToC. These are the conditions that need to hold true for the program activities to lead to the desired outcomes.
5. **Create a Visual Representation:** Develop a visual model of the ToC that maps out the inputs, activities, outputs, outcomes, and impacts. This helps in communicating the ToC effectively to stakeholders.
6. **Iterate and Refine:** The development of a ToC is an iterative process. It should be revisited and refined as new evidence and insights are gained. This allows for adjustments in the pathways and assumptions as needed.

This comprehensive ToC of the intervention as currently being implemented will form the basis for later discussions and reflections on whether the intervention is lean and affordable enough to scale; and if not, whether there are opportunities to simplify the model by identifying the components of the ToC that are core or essential, and those that are not absolutely essential in driving impact.

Figure 4: Detailed Theory of Change for Community Scorecard




**Data Sources:**

- CSC CARE Core Model
- Theory of Change - ensuring health services during pandemics
- Community Scorecard Digitization
- CAREs experience with community scorecard what works and why case study
- Community scorecard R&D Report
- Key Informant Interviews

## Competitive Landscape Assessment

Alongside analyzing the readiness for scale of CARE's existing models, it is important to conduct a competitive landscape assessment to understand if there are other solutions outside of CARE that are addressing the same challenge in a more impactful, simpler, more affordable, or more cost-effective way and whether these solutions are already on a pathway to scale.

### Step 1: Identify a long list of comparable solutions

A competitive landscape assessment begins with conducting an online scan to pull together a long list of comparable solutions developed by other organizations that have already been 1) implemented, 2) evaluated, and 3) resulted in a positive impact. Ideally, the work of compiling the long list should be conducted by an independent evaluator within or outside of CARE, to ensure that the process is free of bias.

The following are suggested criteria for the long list of comparable solutions, with illustrative examples for a solution to reduce violence against women and girls:

1. **Solutions that are highly similar to CARE's model:** same objective (reduce violence against women and girls), same participant group (couples living in rural communities), and also using a transformative approach designed to serve women and girls facing violence OR
2. **Solutions that are somewhat similar to CARE's model:** same objective (reduce violence against women and girls), varied participant group (e.g., men living in rural communities). OR
3. **Solutions that are dissimilar to CARE program:** primary objective is different but secondary objective is the same and resulted in positive outcomes (-i.e. direct cash transfer with secondary objective of reducing violence against women and girls).

### Step 2: Arrive at the short list of comparable solutions

Once a long list of candidate interventions has been compiled, CARE technical experts should refine it to a focused shortlist for further investigation. Again, this work should ideally be led by an independent evaluator within or outside of CARE.

**Suggested Criteria for Shortlisting:**

1. **Evidence of Impact**
- Strong, rigorous evidence of positive outcomes in at least one implementation context.

2. **Evidence of Scalability**

Indicated by at least one (preferably more) of the following:
- Model Simplicity: Lean, affordable, and adaptable design.
- Adoption by Others: Uptake by governments or other INGOs/CSOs across multiple settings.
- Cost-Effectiveness: Formal analyses showing good value for outcome delivery.
- Flagship Model Status: Widespread implementation across an INGO's footprint (e.g., CARE's VSLA model).

3. **Promising "Maverick" Solutions**
- Interventions with strong anecdotal potential—even without rigorous evaluation—can be considered if they address emerging issues or fill critical gaps.

### Step 3. Conduct a Deep-Dive of Shortlisted Solutions

Once you have arrived at the shortlist of interventions, you will review each of them more deeply.

- Develop a deep understanding of the program logic for each competitor intervention to understand how they achieve impact, and then assess their evidence of impact.

- Conduct a S.W.O.T. analysis to ensure a structured comparison between the CARE intervention and competitor interventions.

- Use insights from that process to draw informed conclusions about whether the competitor interventions are a) more impactful than your intervention, b) further ahead in their scaling journey, or c) more simple, affordable, and cost-effective and therefore easier to scale.

These findings will help determine the appropriate strategic posture going forward. You may find that another organization's solution is more scalable, and CARE could partner to support it with your networks and expertise. Alternatively, you may decide to pivot by adapting your solution based on learnings from others to make it simpler or more cost-effective. Or, you may validate that your current intervention is indeed the most impactful and scalable option—and should proceed toward scale. This analysis sets the stage for the decision

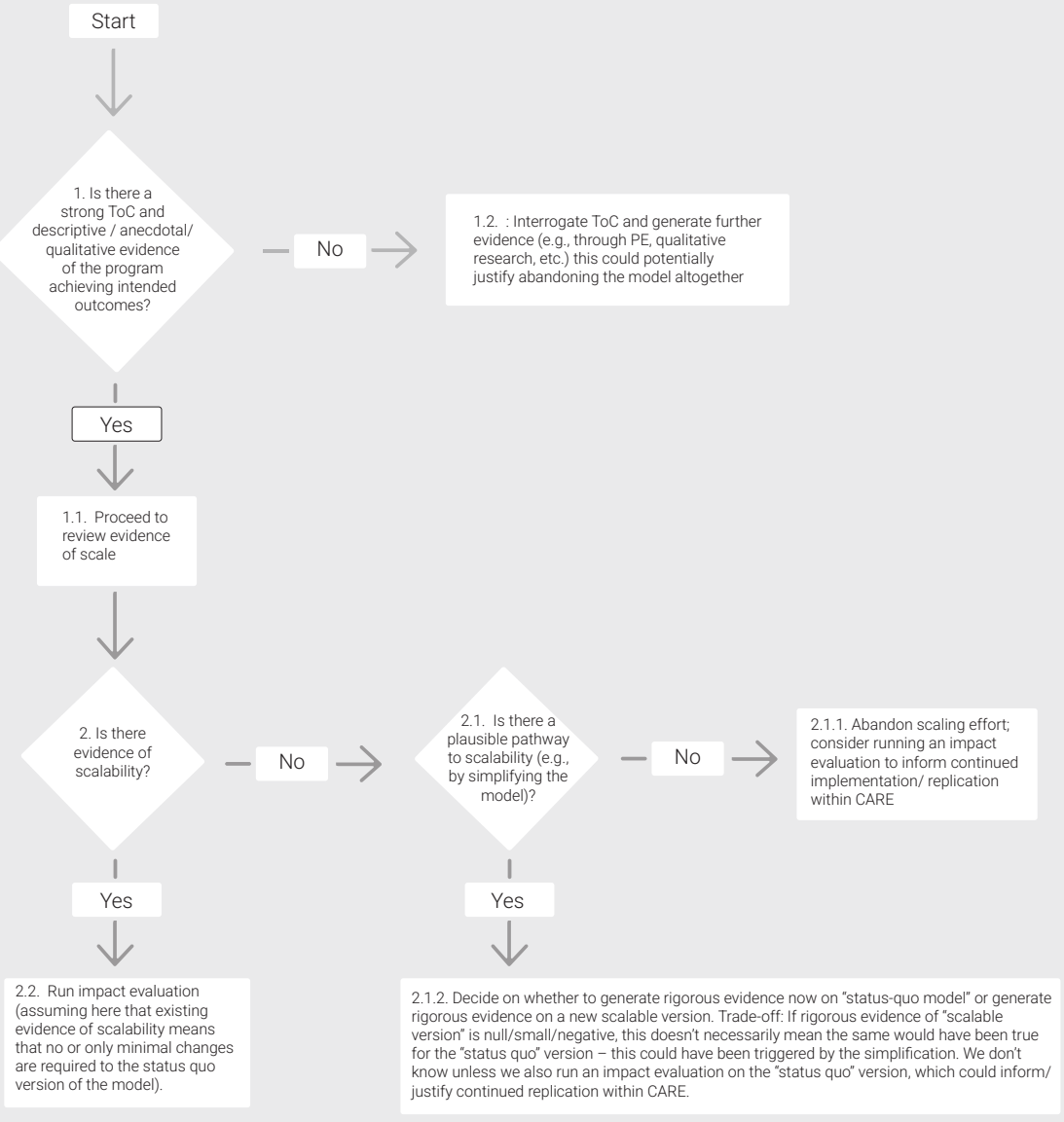
framework on the next page: Partner, Pivot, or Proceed.

## Pivot, partner, or proceed

At the completion of the evidence and competitive landscape reviews, the independent evaluator should be able to advise on the gaps and limitations of the CARE solution from an impact and scale perspective, whether other solutions in the sector that address those gaps have a design that is more scalable, and whether CARE's best course of action is to pivot, partner, or proceed. That makes this an important decision point.

If an intervention has limited or no rigorous evidence of impact, the flow chart below can help navigate decisions

Figure 5: Decision process for no existing evidence of impact

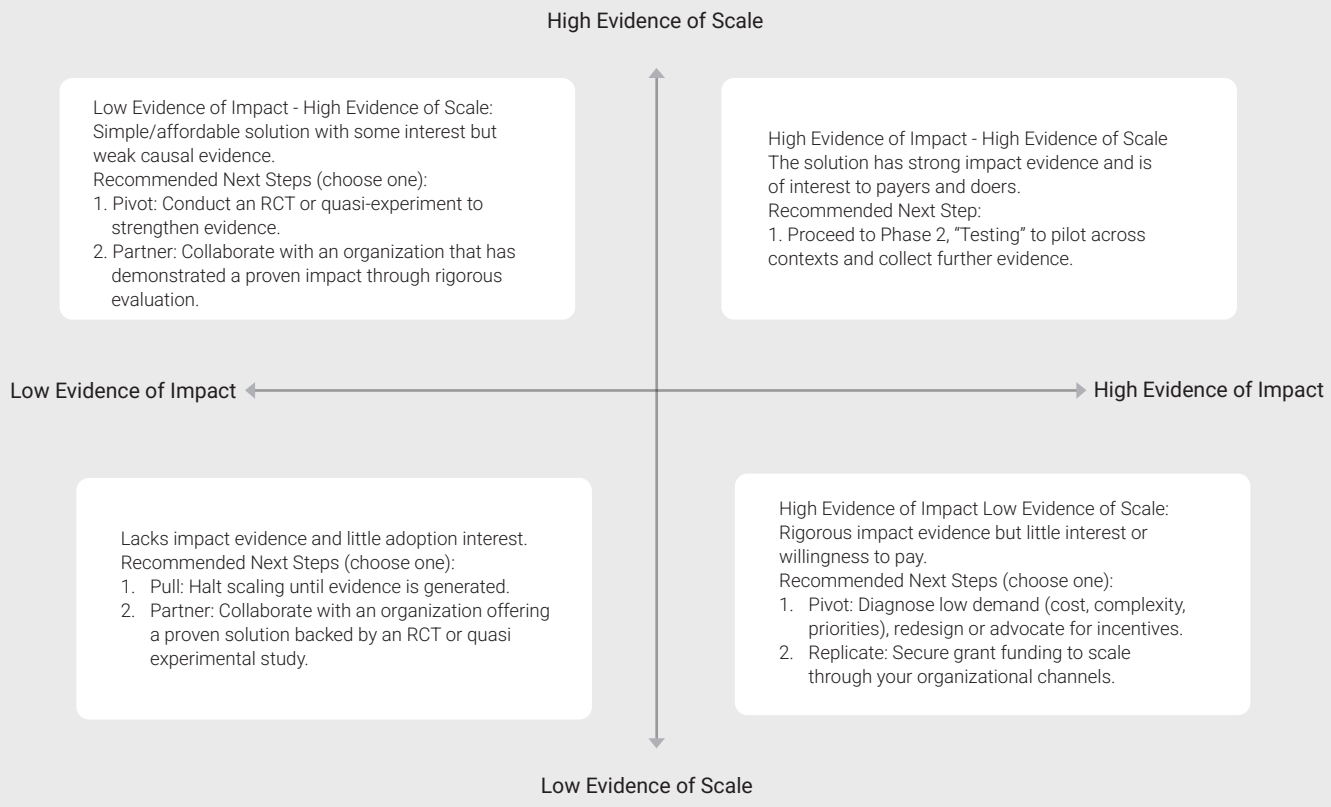


about whether it makes sense to generate such evidence and/or continue scaling efforts.
When existing rigorous evidence suggests that an intervention's impact is null, small, or negative, the choices for how to proceed include:

- Abandon the intervention
- Re-design the intervention—with scale in mind, if scale is the goal—and generate new rigorous evidence
- Partner with others that have an evidence-based model that could be scalable

These and other possible recommended next steps based on evidence of impact and evidence of scale are described in

Figure 6: Impact-Scale Decision Matrix



## Cost Review

To make informed decisions about scalability, we need to understand not just whether an intervention can be simplified, but also the financial feasibility of scaling. A holistic cost review aims to quantify cost-effectiveness, efficiency, and affordability, providing concrete financial information to guide scaling decisions.

Multi-component interventions with long implementation timelines and intensive human capital requirements are expensive, sometimes costing millions of dollars to implement while reaching relatively few people. Though this level of resourcing may be possible for an organization implementing a grant-funded project, for potential doers and payers at scale like governments, implementing such interventions nationally may be financially impossible, even if they recognize the intervention's value. While economies of scale and integration into existing systems may reduce per capita costs and improve cost-effectiveness, the absolute amount required might still far exceed available sectoral budgets.

A thorough cost review includes several analytical approaches:

### Step 1: Compile cost data

**We begin by compiling detailed data on actual program implementation costs.** We separate one-off costs that won't recur as the intervention scales (such as program design, contextualization, or initial training) from ongoing operational costs that will continue throughout implementation. This differentiation helps us understand the potential cost structure of taking an intervention to scale.

### Step 2: Conduct cost analyses

**With cost data in hand, we conduct various analyses depending on the available evidence.** When rigorous impact data exists, we may do a formal cost-effectiveness analysis that calculates the cost per unit of outcome achieved (e.g., cost per percentage point reduction in malnutrition). Common metrics, such as disease incidence (health), child test scores (education), and crop yields (agriculture) allow comparison across different interventions within a sector targeting the same outcomes. Conversion of outcomes to a common unit, such as utility or monetary terms, can build on a cost-effectiveness analysis framework to allow for comparisons across sectors.

In cases where impact data is limited, we might analyze per-unit-costs based on outputs or intermediate outcomes available through monitoring data. Such a cost-efficiency analysis can be used to compare interventions in a sector that have similar theories of change with comparable outputs. For instance, we might calculate the intervention cost per latrine sold or per savings group established. Further, if there is strong causal evidence for linking inputs-to-outputs and outputs-to-impact, cost-efficiency analysis can help to find the lowest-cost configuration of inputs that supports impact.



### Step 3: Develop Cost Projections

We may also develop projections for how costs would scale under different scenarios. This includes modeling how economies of scale might reduce certain costs, how utilizing existing government systems could change the cost structure, and how technological innovations might improve cost-efficiency. Predictive analytics and machine learning can also be useful tools in this effort. Projections are then compared to relevant government sectoral budgets to assess affordability in absolute terms.

In some cases, we use a reverse approach, calculating the maximum cost per participant or output that would make our intervention competitive with alternatives. This provides a helpful reference point: keeping one of the two parameters (financial inputs or desired outputs) constant tells us by how much the second parameter would need to change to be competitive with other evidence-backed interventions.

**Throughout the cost review, we consider both quantitative financial data and qualitative information about willingness to pay.** For instance, examples of governments having previously paid for similar interventions can provide valuable signals about affordability and priority.

## "Scaled" Theory of Change Design

CARE's programs are often built around CARE and its partners as the doers with CARE's grant funding as the payer. When aiming for scale, however, the ToC must likely be shifted from the status quo version—it must be distilled into a version that other doers and payers can adopt, implement, and pay for. The goal of "scaled" ToC design is to review an existing intervention's logic to:

1. **Define Core vs. Flexible Components**
   - Core components are key impact drivers and essential to achieving impact.
   - Modular/flexible components can be adapted to local contexts or dropped if not critical.

2. **Simplify and Integrate**
   - Identify opportunities to streamline activities and reduce cost.
   - Leverage existing systems (e.g., government infrastructure, community networks) to deliver elements of the intervention.

3. **Interrogate Assumptions**
   - Surface and challenge underlying assumptions that may not hold at scale.

4. **Incorporate Competitive Landscape Insights**
   - Review external solutions addressing the same challenge to learn how they've simplified or cost-reduced their models.

The output of this process is a ToC that includes only the components necessary for impact, with clear rationale for whether or when to include any modular elements. It should be reduced to the essential program logic, and should be practical to implement across a range of settings. This pared-down model—which can be thought of as a 'rough draft' of the optimal fidelity model—is then ready to test for impact in the proof-of-concept step of Phase 2.

### Key Considerations for Scaled ToC Design

1. ***Envisioning Scale Is Hard***
   *Because it can be challenging for teams focused on direct delivery or replication to think in terms of sustainable scale, it is important to bring in "critical voices" to challenge your assumptions.*

2. ***Beware Over-Complexity***
   *Large, expensive interventions are difficult to trim, and scaling them is unlikely to be successful. Program teams should consider avoiding this time-consuming and often ineffective process.*

3. ***Focus Your Scope***
   *For complex interventions targeting many disparate outcomes, it may make sense to focus on one key outcome of interest and then identify one program component (from the internal or an external solution) to scale, or to design a new scale-centric model from scratch.*

# Enabling Environment Review

The Enabling Environment (EE) Review helps teams assess the systemic conditions—such as policies, institutions, infrastructure, norms, and incentives—that influence whether an intervention can take root and scale sustainably. While strong program design is critical, even the most promising solutions can stall without the right enabling conditions. This review ensures we are choosing suitable implementation environments and appropriately adapting the intervention to different contexts.

There are two types of EE review, depending on where your team is in the program design or scaling journey.

## Type 1: Rapid EE Review – For Selecting Where to Scale

This is a lighter-touch review intended to help teams identify promising new contexts for testing or piloting their proposed optimal fidelity model. It aims to ensure there is a minimum level of systemic readiness that supports implementation and learning without requiring major system reforms. Usually this review relies on a pre-selection of potential locations for testing or piloting.

### Step 1: Define Minimum Implementation Criteria

Identify the foundational enablers that must exist for viability (not guarantees of success):

- **Policy and Fiscal Alignment:** National/subnational strategies and budgetary space to sustain the program.
- **Supportive Delivery Structures:** Existing formal or informal channels (e.g., community networks, VSLAs).
- **Population Need and Fit:** A clear, large-scale problem that your intervention can address.
- **Cultural Acceptability:** No major social-norms barriers or backlash risks (e.g., evidence of prior similar programs, policy discourse support).

### Step 2: Conduct a Light-Touch Scan

Use desk research, internal knowledge, and a few key informant interviews to assess where those criteria hold, adhering to these key data collection principles:

- **Standardized Data:** Leverage global data sets like demographic and health surveys, World Development Indicators, etc.
- **Recent and Comparable:** Prioritize up-to-date data (≤10 years) and note any temporal mismatches in comparisons across countries.
- **Triangulation:** Use internal reports and targeted staff interviews to validate findings and fill gaps.
- **Transparency:** Document any data limitations or proxies and flag uncertainties.

### Step 3: Prioritize Candidate Contexts

Rank pre-selected geographies by their readiness, using a simple matrix of key criteria (policy, capacity, need):

- **Moderately Enabling Environments:** Avoid extremes—neither too easy (skewed success) nor too challenging (predetermined failure).
- **Actionable Output:** A ranked shortlist of 3–5 locations that meet minimum EE conditions for early testing.

#### Example: Rapid EE Review for MBS Solution

***Step 1:*** *For a rapid EE review for where to test and scale a Market-Based Sanitation (MBS) solution with iDE, we started with a short-list of four potential East and Southern African countries: Ethiopia, Kenya, Mozambique and Zambia where both CARE and iDE had existing operations*

***Step 2:*** *For MBS, we focused on criteria linked to governance (e.g. presence of sanitation subsidy schemes, budget allocation to WASH interventions), the business environment (e.g. presence of cement in-country and tax burden for SMEs) and sanitation demand (e.g. rates of open defecation and coverage of improved sanitation in rural areas).*

## Type 2: In-Depth EE Review – For Adapting to a New Scaling Context

This version is more comprehensive and is used once a target geography or country has been selected. It helps teams adapt the program design and ToC to local realities, based on a nuanced understanding of the system in which the intervention will operate. This process is especially critical when adapting a program to a new country or scale level, helping to ensure the "scaled ToC" accounts for context in its inputs and activities.

### Step 1: Define scope based on your model's assumptions

Identify which enabling conditions matter most to your model (e.g. public financing, coordination mechanisms, norms affecting uptake, or private sector delivery capacity). These focus areas should shape the review.

### Step 2: Map key actors and systems

Once you have identified the enabling conditions most critical to your model, the next step is to map the ecosystem of actors who influence or are impacted by these conditions. This includes government bodies, private sector actors, civil society organizations, donors, and community-level stakeholders. Be sure to include underrepresented groups whose perspectives and barriers may differ, as their experiences can provide critical insight into how the system is functioning, and for whom.

### Step 3: Gather evidence through desk research and interviews

With your focus areas and key actors identified, the next step is to gather evidence to understand how the system functions in practice. Start with a combination of desk research and light-touch interviews—for example, brief calls with technical experts, implementing partners, or staff from country and regional offices. These help validate findings and fill gaps in documentation. Prioritize recent, context-specific information and triangulate insights using sources such as program evaluations, grey literature, academic studies, and government strategy documents. Where possible, draw from research conducted in similar geographic or implementation settings.

In contexts where CARE or partners already have a presence, consider complementing this with more in-depth qualitative fieldwork, such as focus group discussions (FGDs) or key informant interviews (KIIs) with community members, local government officials, or frontline service providers. While more resource-intensive, these methods can offer grounded, first-hand perspectives that help tailor solutions to local needs. This kind of research should be pursued where time, budget, and capacity allow.

### Step 4: Analyze gaps, barriers, and opportunities

Next, use the data gathered to assess what must be adapted for the program to succeed in this new context. Highlight areas where the model aligns well, and identify where systemic constraints may require workaround strategies or targeted support.

- Which assumptions in your ToC may not hold in this context?
- Are the required system shifts feasible within your timeline and budget?
- What systems strengthening activities are required to support the implementation of the program? Are they feasible?
- What further information do you need, if any, to support your decision making?

The output of this exercise should be a set of actionable recommendations for adapting the program design and ToC to the local context. This should also include suggestions for stakeholder engagement or systems strengthening required to support implementation and scale. A thorough review is participatory and ensures all voices are heard, incorporating feedback from underrepresented groups. A robust EE review not only diagnoses systemic challenges but also helps design targeted interventions that promote durable development that promotes equality.

## Key Considerations for EE Review

1. ***Be realistic***
   *No context will be perfect. The goal is to understand what can be changed and what must be worked around.*
2. ***Stay flexible***
   *You may uncover dynamics that challenge your original design assumptions. Let the review inform, not just validate, your strategy.*
3. ***Think long-term***
   *Even if short-term implementation is possible, will the system support scale in 5-10 years?*
4. ***Keep it useful***
   *Prioritize insights that directly inform your ToC and implementation plan. Avoid creating a lengthy report* that is hard to digest or not actionable.

## Example: Structuring the In-Depth EE Review for MBS

***Step 1: Frame Research Questions and Themes***

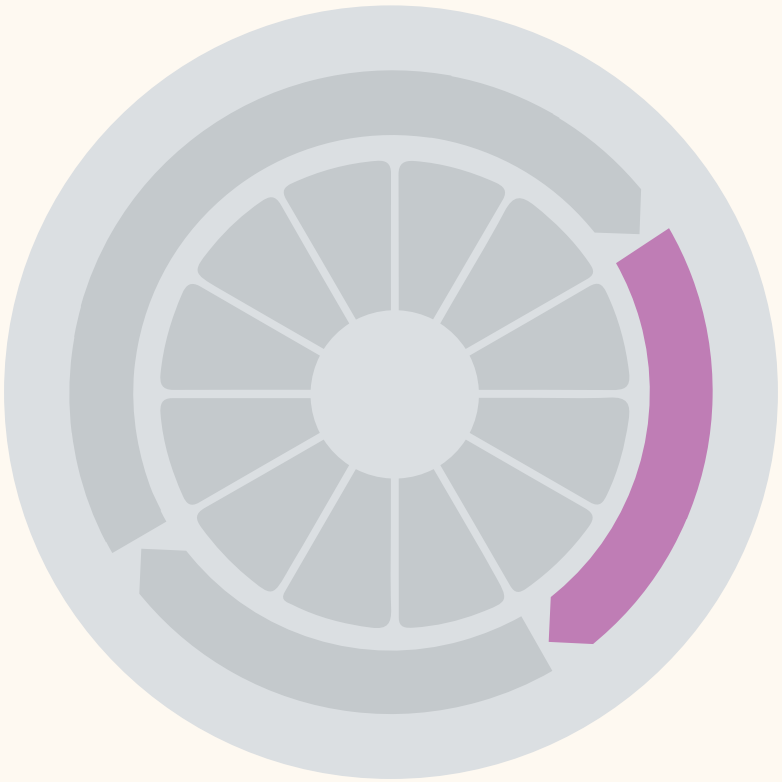
*For Market-Based Sanitation (MBS), we organized our review around three core themes—Demand Generation, Supply Development, and Enabling Environment Strengthening—each tied back to the model's hypotheses, with special emphasis on household and entrepreneur financing.*

***Step 2: Map Actors Against Key Hypotheses***

*A central hypothesis was that (women) entrepreneurs will secure financing. We therefore mapped all actors in the financial services ecosystem—microfinance institutions, SACCOs, mobile-money providers, government credit schemes—to assess:*

1. *Realism: Do these providers offer products women can access?*
2. *Equality Gaps: Are there barriers (e.g., collateral requirements, social bias) that need targeted interventions?*
3. *Support Needs: What partnerships or capacity building will strengthen financial channels that reach everyone?*


care®



# Phase 2: Test

# Phase 2: Test

Phase 1 developed the scaled ToC, which can be thought of as a rough draft of the optimal fidelity model. The Enabling Environment Review helped target suitable contexts for testing and piloting the proposed optimal fidelity model. The objectives of Phase 2 are to identify and test cost-efficient refinements of the scaled ToC, and then to develop the final optimal fidelity model, which will be validated in a single-site pilot and ultimately a multi-site trial.

## Overview

**01** Identify Testable Variants: Teams begin by translating the scaled ToC into a practical implementation roadmap for field trials. Through a series of structured workshop exercises, they unpack core components of the solution and surface areas for experimentation—such as streamlined delivery approaches, reduced dosage, or alternative implementation channels. Each variant is mapped to its hypothesized cost, fidelity, and impact outcomes. A monitoring framework is then drafted to track key learning indicators, including unit cost and implementation fidelity.

**02** Field Trials: Teams test 2–4 lean variants in a single site, running structured field trials over a 1–3 month period. These trials are designed for rapid learning using dashboards, short surveys, and quick-turnaround interviews. The objective is to compare the variants against the baseline model to understand what works, what's feasible, and where trade-offs lie between cost-efficiency, simplicity, and impact.

**03** Pilot Design and Rollout: Following the trials, teams consolidate insights to finalize the optimal fidelity model. This model is then tested through a full single-site pilot to validate implementation feasibility and performance. With successful validation, teams move toward multi-site rollout using standardized implementation guides, updated monitoring tools, and durability metrics to assess adaptability across varied delivery contexts.

# Phase 2 Test Activity: What to Expect

These worksheets guide your team through the testing phase to validate and refine your scaled ToC. This phase is about pressure-testing your model's assumptions, identifying viable delivery models, and generating real-world evidence to optimize for scale. By working through these activities, your team will gain insights into implementation fidelity, cost-effectiveness, and readiness for broader rollout.

**DIFFICULTY**
Medium

**WHAT YOU'LL NEED**
Printed worksheets,

Pens/markers,

Monitoring data,

Pilot partners

**PARTICIPANTS**
Project lead

Monitoring and evaluation lead

Technical advisor

Implementation manager

## 1. Workshop Session 2.1: Identify Testable Variants of the Scaled Theory of Change

**Purpose**

Interrogate your Scaled ToC to identify which components of the intervention are core to achieving impact and which can be simplified or removed. This worksheet helps your team document assumptions, articulate hypotheses, and define testable variants for further exploration in field trials.

## 2. Worksheet 2.2: Test Lean Variants Through Field Trials

**Purpose**

Design and execute field trials to test different lean versions of your intervention. Use this worksheet to track how each variant performs in terms of fidelity, feasibility, cost, and outcomes. This data will inform which elements should be retained, adapted, or eliminated in the optimal fidelity model.

## 3. Worksheet 2.3: Final Design Testing and Multi-Site Rollout

**Purpose**

Pilot your optimal fidelity model (OFM) in a single real-world setting to assess whether it can be implemented effectively by others with minimal support. Use this worksheet to monitor fidelity, adapt as needed, and prepare for a multi-site rollout to validate performance across diverse contexts.

## Identify testable variants of the scaled theory of change

Once a scaled ToC has been developed, the next step is to interrogate its contextualized structure to identify opportunities for streamlining. The goal at this stage is to assess whether the model can be made leaner and more cost-efficient without compromising impact—an essential consideration when preparing an intervention for scale.

This process can begin within the same workshop used to build the scaled ToC. Using familiar facilitation tools—such as dot voting, group commenting, and sticky-note exercises—participants are asked to go beyond validation and actively critique the model. Specifically, they are prompted to examine input-to-output pathways and surface hypotheses about where efficiencies might be possible. For example, participants might highlight which inputs or activities appear low-cost relative to the outputs they produce, or where redundancies or bottlenecks exist.

One structured exercise asks participants to identify strengths, weaknesses, opportunities, and threats (SWOT) across the ToC. Using color-coded dots and written annotations, they flag specific nodes and links that may warrant redesign. These crowdsourced insights help seed the search for lean implementation variants.

Following the workshop, the monitoring and evaluation and implementation teams work independently to refine and test these hypotheses. The monitoring and evaluation team revisits the enabling environment review with the pilot context in mind, checking whether the ToC aligns with real-world constraints and opportunities. Their focus is to match inputs and activities to local systems and identify any environmental enablers that could reduce implementation complexity or cost. In parallel, the monitoring and evaluation team develops a monitoring framework that identifies key intermediate outputs and links them to cost-tracking mechanisms.

This framework will later enable real-time data collection during pilot rollouts or short-cycle trials, supporting both implementation fidelity and cost-efficiency assessment.

Meanwhile, the implementation team begins translating the ToC into a detailed implementation plan for the core model. This includes finalizing the inputs, sequencing the activities, setting timelines, and defining roles through a RACI matrix. A key part of this step is interrogating the "how much" of each input—its intensity, duration, and frequency—to identify possible redesigns that reduce cost without eroding effectiveness. The implementation plan is then formalized through a process map or logframe, which ties each input and activity to intermediate outputs and includes corresponding indicators. Crucially, the process map also identifies potential cost centers—outputs or activities where costs concentrate—which can then become focal points for optimization.

Finally, the monitoring and evaluation and implementation teams reconvene to synthesize insights from the workshop and post-workshop planning. Together, they identify a shortlist of testable implementation variants—models that are either more cost-efficient or higher-performing than the base case. These variants are grounded in real contextual constraints and are designed to be practical to test using short-cycle trials or A/B tests.

By building this step into the scaling process, CARE can ensure that models moving toward scale are not only impactful, but also operationally lean and financially resilient in diverse settings.

## Test lean variants through field trials

With lean implementation variants identified and a contextualized ToC in place, the next step is to set up a structured testing environment. This single-site serves as a laboratory—a controlled, real-world setting where the implementation strategy can be tested, adjusted, and refined, and the optimal fidelity model finalized, before more rigorous testing and scaling.

The first task is to locate a suitable site and population. The selected site and population should meet three criteria:

1. It must be representative of the broader target context, so that findings from the field trials can inform scale-up decisions.
2. It should be small enough to maintain control over delivery and ensure fidelity to the variant designs.
3. It must also be large enough to support adequate sample sizes for short-cycle trials or A/B testing, providing reliable signals on effectiveness and cost-efficiency.

Once the site is selected, the project team proceeds to implement the core solution based on the implementation plan developed in the previous step. This plan outlines the inputs, activities, sequencing, and roles. At the same time, the team must set up a robust monitoring system capable of capturing all relevant data—particularly intermediate outputs and associated costs. This monitoring infrastructure is critical to the evaluation of each variant's performance and cost-efficiency during testing.

With the site functioning as a controlled testbed, the team begins testing the cost-efficient variants identified through the ToC workshop and desk review. The core model, implemented as designed, serves as the baseline.

Each variant modifies the model in deliberate ways—adjusting the quantity, quality, duration, or delivery mode of inputs and activities. These modifications reflect principles of scalable design: simplifying the model, lowering delivery costs, or improving output-to-cost ratios. Each testing round should focus on a manageable number of variants—typically 2 to 4.

The next phase is to run short-cycle field trials, or "microtrials", to test these variants under real conditions. Small clusters (such as villages, staff teams, or service units) are used as the unit of assignment. For fast, directional learning, around 10–15 clusters per variant—or approximately 100–300 individuals—are generally sufficient. If more rigorous testing is required (e.g., for formal A/B testing), larger sample sizes will be necessary.

Trials typically run for 1 to 3 months, during which the monitoring system captures data on reach, outputs, short-term outcomes, and implementation costs. A range of tools may be used: dashboards, mobile reporting systems, and rapid surveys (by phone or in-person). These can be supplemented—or, in some cases, substituted—by qualitative methods such as key informant interviews (KIIs) and focus group discussions (FGDs). Respondents may include frontline staff, participants, and program personnel, depending on the outcome being measured.

After each trial round, the team conducts an analysis to compare the performance of each variant. In well-powered, randomized trials, this may involve formal A/B testing. In smaller-scale tests, the analysis may be more descriptive or indicative—but still valuable for guiding iteration. Variants are evaluated not only for their outputs, but also for their cost-efficiency—i.e., how much output or outcome is achieved per unit of cost.

The results inform a structured iteration and refinement process. Underperforming variants are dropped. High-performing features are retained or adapted. New variants may be defined based on field learning and retested in a second or third cycle. Across 2–3 short testing rounds, the goal is to converge on a refined implementation model—an "optimal fidelity model" that delivers consistent outputs at lower cost and is feasible to deliver at scale.

By treating the field trials as a laboratory and applying a structured, data-driven testing process, CARE can systematically refine its implementation strategy—ensuring that what is scaled is not only impactful, but also efficient, affordable, and tailored to the realities of delivery at scale.

## Final design testing and multi-site roll out

After iterating toward an optimal fidelity model through micro trials and short-cycle testing, the next step is to subject the finalized design to more rigorous validation. This is done in two stages: first, a single-site pilot to confirm performance in a representative context; and second, a multi-site randomized controlled trial (RCT) to establish generalizable evidence for scale.

### Stage 1: Single-site validation of the final design

The single-site pilot serves as a final "proof-of-concept" assessment for the optimized implementation model. While earlier trials prioritized directional learning and iterative refinement, this phase is designed to rigorously test whether the final optimal fidelity model performs consistently in a real-world setting when implemented as intended.

Site selection is critical at this stage. The chosen pilot location should satisfy the following criteria:

- **Operational accessibility:** The site must be easy to reach and oversee, with minimal logistical barriers to implementation, monitoring, or supervision.
- **Presence of core delivery infrastructure:** The site should have the basic resources and enabling conditions required for the intervention—such as local partners, frontline staff, or the places where participants access services.
- **Contextual representativeness:** The site should reflect key features of the broader target population or geography (e.g., socio-economic profile, delivery constraints, institutional landscape), so that findings are relevant to scaling decisions.
- **Generalizability of findings:** The site should be similar enough to other priority contexts that lessons learned are likely to translate, but not so atypical or idealized as to overstate feasibility or performance.
- **Evaluation readiness:** The site must allow for clear identification of a target population, implementation of appropriate data systems, and tracking of key outcomes. Consent procedures and ethical oversight must also be feasible.

Evaluation design depends on available resources and the level of certainty required. Where feasible, the preferred design is a randomized controlled trial (RCT) with sufficient power to detect meaningful effects on key outcomes. However, if resource, ethical, or contextual limitations preclude randomization, quasi-experimental approaches (e.g., matched comparison groups, regression discontinuity) or structured qualitative process evaluations may be used instead.

The minimum requirement is a design capable of credibly assessing whether the final model is achieving its intended outputs and outcomes, and under what conditions.

A "successful" single-site pilot meets three criteria:

1. **Fidelity:** The model is implemented with high adherence to the design, demonstrating that it is operationally feasible and that implementers can consistently deliver it.
2. **Effectiveness:** The intervention achieves intended outputs and short-term outcomes at expected levels, confirming that previous micro trial results hold under more realistic delivery conditions.
3. **Efficiency and acceptability:** The model demonstrates the expected cost-efficiency and is acceptable to both implementers and recipients—validated through cost tracking, implementation logs, and feedback from frontline workers and participants.

Results from this pilot inform the final adjustments before multi-site testing. If critical issues are identified—such as unexpected implementation barriers or context-specific risks—minor redesigns may be made. However, the goal is not further iteration, but confirmation that the model is ready for broader deployment.

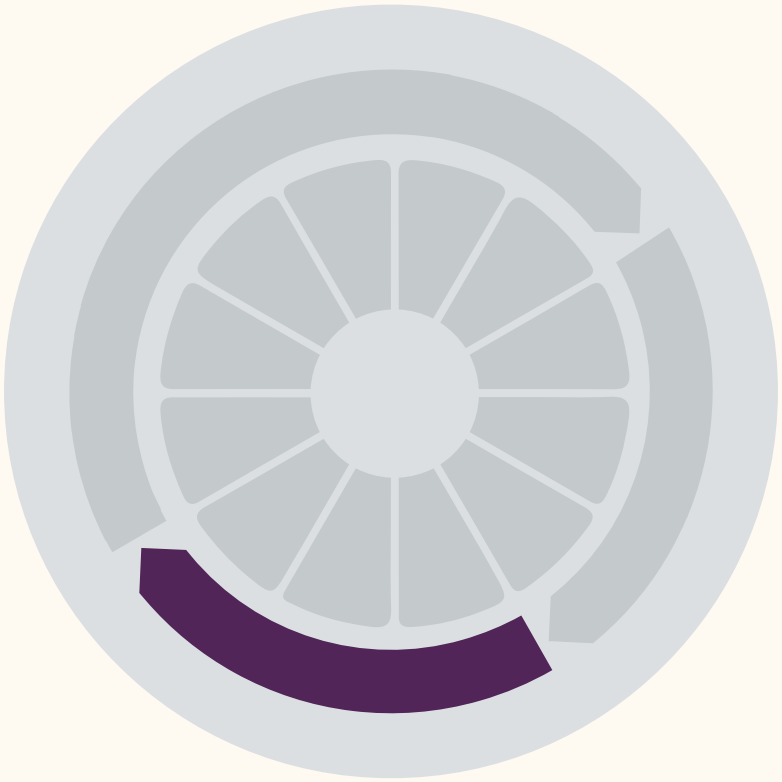
### Stage 2: Multi-site randomized controlled trial

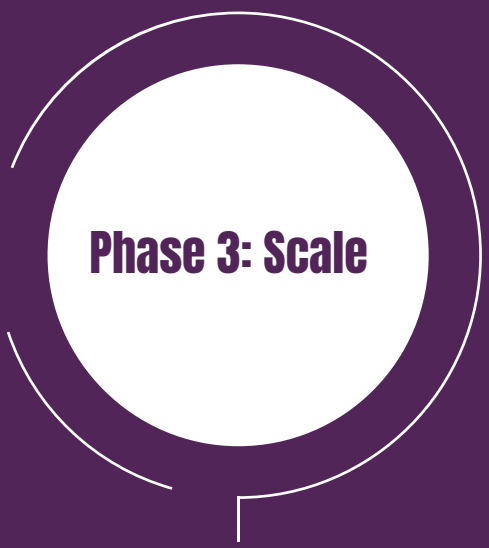
The final stage is a multi-site RCT, designed to establish the generalizability and scalability of the intervention. Sites should be selected to reflect the range of delivery contexts where the program is expected to operate at scale. Stratified randomization can help ensure balance across key variables (e.g., geography, population density, partner capacity).

At this stage, the implementation model is no longer being refined. Instead, the focus is on validating its performance under diverse real-world conditions. The multi-site RCT provides robust evidence on the consistency of results, the adaptability of the model across sites, and the magnitude and cost-effectiveness of its impact at scale.

This final evaluation anchors the case for broader investment and rollout. By concluding Phase 2 with a rigorous, contextually grounded validation process, doers and payers at scale can proceed with confidence—armed with both the evidence and the operational knowledge to deliver durable impact at larger scale. If the multi-context evidence evaluation is completed, and rigorous evidence of the optimal fidelity model's impact in various contexts and geographies generated, it is ready to be pitched to local systems actors.






Phase 3: Scale

# Phase 3: Scale

By Phase 3, your optimal fidelity model, ready for sustainable scale, has been validated. Now, leveraging CARE's Systems Orchestration Playbook, you'll guide host system actors to adopt, implement, and finance the solution at scale. CARE shifts from implementer to orchestrator—activating champions, embedding the model into existing structures, and enabling adaptive management for sustained impact.

## Overview

**01** Preparation: Align stakeholders and define shared ownership. Convene host system actors—across government, private sector, and civil society—to align on pilot results, shared goals, and future roles. Use Systems Orchestration Playbook Worksheets 1.1, 2.1, 3.1, and 4.1–4.2 to prioritize stakeholders, map influence, and clarify responsibilities for leading and supporting scale-up.

**02** Design: Co-create pathways to integrate and deliver. Assess institutional capacity using Systems Orchestration Playbook Worksheets 5.1, 6.1, and 7.1–7.4. Identify system bottlenecks, co-design workplans and training programs, and embed the model into routine delivery systems and budgets to ensure durable ownership.

**03** Planning and Implementation: Secure governance and financing for scale. Use Systems Orchestration Playbook Worksheets 8.1 and 9.1 to establish long-term governance and reflection cycles. Secure financing via public budget lines, private cost-recovery mechanisms, or donor support. Ensure that adaptive management loops are in place to sustain and evolve the model.

# Phase 3 Scale Activity: What to Expect

These activities guide your team through orchestrating host system adoption, integration, and financing of the optimal fidelity model—using CARE's Systems Orchestration Framework to ensure sustainable scale.

**DIFFICULTY**
Medium

**WHAT YOU'LL NEED**
Finalized optimal fidelity model and process map

CARE's Systems Orchestration playbook

Stakeholder ecosystem and RACI templates

Capacity-building curricula and materials

Financing and budget planning templates

Advocacy and communication toolkits

**PARTICIPANTS**
CARE Systems Orchestrator

Government focal points (policy, finance, program leads)

Private sector champions

Civil society and community representatives

Donor liaison

## 1. Preparation: Stakeholder Engagement and Partnership Alignment

**Purpose**

Using Systems Orchestration Playbook Worksheets 1.1 through 4.2, assess CARE's readiness to orchestrate, identify and prioritize key system actors, and clarify their roles in the scale-up process. Convene influential champions from government, private sector, and civil society to align on the model's goals and define their contributions through structured partnership agreements.

## 2. Design: Embed into Systems

**Purpose**

Leverage Systems Orchestration Playbook Worksheets 5.1 to 7.4 to assess system capacity, identify delivery bottlenecks, and co-develop pathways for integration. Build ownership by embedding the model into institutional budgets, training structures, and service workflows through joint planning and coaching-on-the-job.

## 3. Planning and Implementation: Finance and Govern for Scale

**Purpose**

Use Systems Orchestration Playbook Worksheets 8.1 and 9.1 to activate governance structures and feedback loops that will guide long-term implementation. Mobilize reliable, long-term financing by aligning with public budget cycles, establishing cost-recovery strategies, and coordinating with donors. Ensure adaptive management through performance reviews and joint troubleshooting.

***Note*:**
*Carry out these steps in order: first build alignment and ownership, then strengthen system integration and delivery capacity, and finally establish governance and financing to support long-term scale. Use continuous feedback to adapt and improve as you expand.*

## Systems Orchestration Process

CARE's Systems Orchestration process is a nine-step, three-stage journey that equips local systems to fully own, implement, and sustain evidence-backed interventions at scale.

### Stage 1: Preparation – Lay the Foundation

In the Preparation stage, CARE transitions from implementer to orchestrator by first clarifying its new role and identifying the right moment to begin Systems Orchestration—when evidence is strong, partners are aligned, and the policy environment is receptive. Teams then conduct a comprehensive stakeholder mapping exercise, charting government bodies, private sector actors, civil society groups, donors, and community voices across national, regional, and local levels. A rapid Enabling Environment assessment follows, evaluating policies, budgets, and delivery structures, alongside an operational audit of legal, planning, technical, and monitoring and evaluation capacities. Finally, CARE and its stakeholder core team co-define a phased plan—sequencing validation, advocacy, buy-in, institutional capacity-building, and adaptation—with clear timelines and milestones.

### Stage 2: Design – Co-Create the Strategy

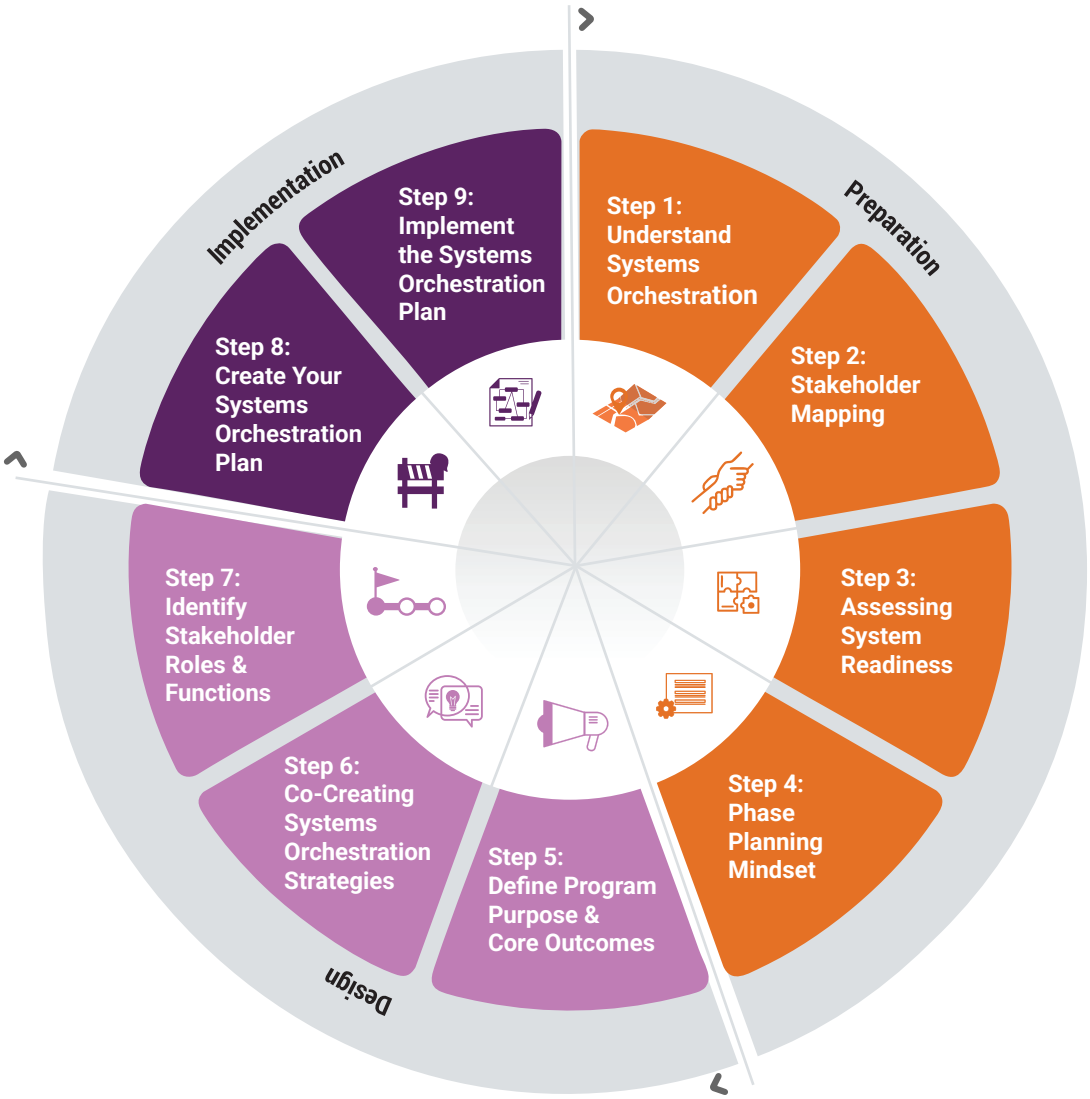
During the Design stage, CARE works hand-in-hand with the broader ecosystem of host-system partners to co-craft the orchestration strategy.

First, the team co-creates a concise program purpose statement and agrees on 3–5 core outcomes—both for participants and system actors—along with measurable indicators. Next, they identify strategic levers such as policy advocacy, capacity strengthening, and financing mechanisms, drafting targeted action plans for each outcome area. Roles and responsibilities are then formalized using RACI matrices and partnership agreements, ensuring every orchestration function—from policy engagement to monitoring and evaluation—is owned by the appropriate institution.

### Stage 3: Implementation – Activate and Embed

In the Implementation stage, the orchestration plan comes to life. CARE and host-system actors launch coordinated advocacy campaigns, joint trainings capacity building efforts, and policy integration efforts while securing reliable, long-term funding through government budget lines, private sector cost-recovery, and donor co-financing. A governance structure—often a steering committee or working group—meets regularly to review progress against indicators, troubleshoot emerging bottlenecks, and adapt the strategy as contexts shift. Through continuous monitoring, iterative learning, and adaptive management, the intervention becomes embedded within existing systems, ensuring that the scaled model is both impactful and enduring.

# Systems Orchestration Process






Best Practices
for Designing
for Scale from
the Start

# Best Practices for Designing for Scale from the Start

The best practices detailed below reflect scaling lessons shared by other leading institutions in the social sector, and they accurately reflect CARE's own experience with scaling. These insights—developed by Save the Children (2023) based on learnings from Brookings Institution, MSI, CASE at Duke and others—provide foundational guidance to teams designing interventions for scale, helping keep scaling pathways, local systems, and doers and payers at scale in mind from the start.

**Simplify the model but maintain effectiveness thresholds**

The art of scaling often revolves around achieving a balance between increasing implementation feasibility by reducing complexity, while still maintaining effectiveness at scale. Simpler models are generally more feasible to scale, particularly in resource constrained settings.

**Scaling is most likely through governments or markets**

Research highlights that commercial markets and institutionalization through governments are the most likely and viable scaling pathways. The latter is particularly relevant where an intervention is targeting more underrepresented  populations.

**Ignore context at our peril**

Scaling teams that design interventions without the context in mind are less likely to succeed. Social, political, cultural, and economic factors all play a role in driving or preventing scale in any given context.

**Understand the incentives at play**

Effective scaling occurs when interventions take into account system incentives. Understanding what motivates and makes it possible for different stakeholders to adopt and continue using an intervention is key to successful scaling.

**Co-develop and bring stakeholders on board early**

Co-development processes, communication plans and reaching out for advice helps bring important stakeholders on board from the outset. It is even better if there is broad stakeholder ownership of the scaling intervention. Remember it takes a system to scale an intervention!

### You will not get it right the first time and will have to adapt frequently as you go

Be prepared to adjust your intervention as new information becomes known. Flexibility and adaptation informed by experience and learning are key to successful scaling.

### Give stakeholders the data they need when and how they need it

Data generated from pilots and other research studies often does meet decision makers' knowledge and learning needs. Additional work is needed to ensure tailored progress and evaluation data is in the right format for important but diverse stakeholders, including policy makers, donors, and participants.

### Document, document, and then keep documenting

Successful scaled interventions often have a history of documenting their experience. This includes capturing implementation process experience, management and team experience and cost, reach and effectiveness data. Creating a culture where data is collected, analyzed, and used to communicate with stakeholders and reach consensus on intervention decision-making pays off.

### Work towards securing the right finance at the right time

Explore and secure a range of different financing at each scaling phase, grants, loans, impact investing and in-kind donations. Start to communicate early with potential donors so that you have greater chance of securing financing for later scaling phases. Donor advocacy may be required to secure funding post pilot, longer funding cycles and additional flexibility to respond to changing challenges and needs as you scale. Scaling is a long-term commitment. Keep funders on board through timely communication, agreeing to scaling milestones and working together to keep stakeholders motivated. This means educating funders on the realities of scaling or identifying and engaging those already familiar with the time and tenacity realities of scaling.

### Team and leadership are critical

The right team makes all the difference. Do not under-estimate the importance of identifying, contracting, and training team members so that you have the right set of skills available to respond to the challenging capacity requirements of scaling. Ensure timely identification of (future) gaps in team skills, capacities and networks, and work towards addressing these with internal and external expertise development or sourcing. Where necessary, bring in the right intermediaries to support your work. Scaling requires a wide range of often highly technical skills. Identify what complementary skills your team will need as you scale and prioritize sourcing and financing these. Dedicated and effective leadership are fundamental to successful scaling. Create the right environment for leaders to act, thrive and grow.

# Conclusion: Charting the Path to Sustainable Scale

To meet its 2030 ambitions and the Sustainable Development Goals, CARE must move beyond grant-funded, direct delivery to become a true Systems Orchestrator, enabling sustainable scale through local actors. This Playbook provides a clear, evidence-driven pathway: begin with a rigorous Evidence Review and Competitive Landscape Assessment; use a comprehensive ToC to isolate core, "optimal fidelity" components; undertake an Enabling Environment Review to select and adapt to the right contexts; and conduct rapid micro-trials to refine a lean, cost-effective model that local systems actors will be able to own, implement, and scale. Finally, by guiding governments, private sector, and civil society to assume the "doer" and "payer" roles, CARE can ensure proven solutions are integrated into local systems—achieving exponential impact without expanding its own footprint.

Realizing this vision demands more than new tools; it requires a mindset shift across CARE and its partners. CARE must be willing to relinquish some control over implementation and outcomes, adapt its operating model (structures, staffing, and revenue), and nurture host-system stakeholders who may expect CARE to lead. Funders will need to embrace process-driven approaches, and CARE teams must build new capacities in evidence generation, advocacy, and systems strengthening. Though challenging—navigating donor expectations, shifting power dynamics, and redefining success—this transition offers immense pride: catalyzing development that truly takes root, flourishes, and endures in the communities we serve.





# Appendix

## Tools and Templates for Action

**(Includes worksheets, sample worksheet templates, workshop guides, consultation prompts, and reflection questions)**

# Phase 1: Review Activity

# Worksheet 1.1: Evidence of Impact Review (Part 1)

Available in IDinsight The Impact Measurement Guide - https://guide.idinsight.org/evidence-reviews/

This worksheet and its linked resources help your team identify, review, and apply relevant research or evaluation insights to strengthen your model. To begin, you'll need a clear understanding of your program's logic—its intended outcomes and how it achieves them—ideally captured in a theory of change (see Worksheet 1.2 if needed). By reviewing findings from past implementations (or similar interventions), teams can base decisions on real-world learning, identify evidence gaps, and ensure the model is realistic, contextually appropriate, and credible to local stakeholders. This template prompts teams to consider what evidence is useful, how strong or relevant it is, and how it aligns with key model assumptions.

Before initiating the evidence of impact review, it is important to decide whether an internal team or an external consultant is best positioned to conduct the review. Here are some things to consider in deciding the best course of action:

Team Capacity: Does someone on your team have time to search, review, and synthesize relevant evidence? AI can help ease time constraints, but a team member is still needed to review, verify, and refine AI output.

Tip: Using AI to Accelerate Your Evidence Review

When using AI for evidence review, opt for "deep research" features (typically subscription-based) and craft specific prompts. Use these criteria to target your search:

- Evidence type: Experimental, quasi-experimental, qualitative, descriptive, or existing reviews/meta-analyses
- Sources: Peer-reviewed articles, working papers, evaluation or project reports
- Geographic focus: Specific countries, regions, or categories like LMICs
- Year: Prioritize recent evidence; studies older than 10 years may be less relevant

Statistical/ research training: Many evaluations simplify findings for accessibility. However, rigorous impact evaluations often include advanced statistical analysis and findings that do require statistical training. A team member with statistical, econometric, or research training will be needed to appropriately assess evidence quality, interpret results, and understand nuances.

Objectivity: How you intend to use the evidence review findings determines whether internal or external review is better. For internal learning, an internal evidence review may suffice. However, for decisions about organizational resource allocation or for funder engagement, an external review can offer greater (real or perceived) objectivity.

# Worksheet 1.1: Evidence of Impact Review (Part 2)

Available in IDinsight The Impact Measurement Guide - https://guide.idinsight.org/evidence-reviews/

The template below – designed to be used separately for each major strategy area or model feature (e.g., delivery mechanism, service design, target group approach) – can guide an internal team conducting the review, or be used to develop a ToR for an external consultant to conduct the review.

1. Identify Potential Evidence Sources: What is the source? What type of evidence does it offer? Example types: peer-reviewed study, government evaluation, NGO report, field diagnostics, internal program data. Include title, author/org, year, and format (e.g., RCT, case study, meta-review, etc.)

2. Assess the Relevance and Context: How similar is the evidence context to your own? Reflect on geography (Was the study conducted in a comparable region e.g Rural vs urban? ), population (Does the population studied align with yours e.g., women entrepreneurs, out-of-school youth?), delivery setting, implementation period (Is the study recent or still relevant?), and intervention similarities. Is the evidence generalizable?

3. Evaluate Findings Relevance: What does the evidence say, and how does it relate to your model? Summarize relevant data points, impact findings, or lessons learned. Note if the source affirms or challenges your model's assumptions, strategy, or delivery mechanism.

4. Strength and Rigor of Evidence: What is the quality of the study? Comment on methodology (e.g., RCT vs. observational), statistical strength, limitations, sample size, or possible bias.

5. Operational Ecosystem Gaps and Implications: What questions are still unanswered? Identify which parts of your model remain untested or unsupported by the literature. What additional data would boost confidence in the scale up?

# Worksheet 1.2: Theory of Change Review (Part 1)

Available in IDinsight The Impact Measurement Guide https://guide.idinsight.org/theory-of-change/

This worksheet helps your team develop or refine a theory of change for your intervention. It should be completed collaboratively, with input from team members who understand the program's goals, delivery mechanisms, and intended outcomes. The theory of change is a visual and narrative model that maps the logical links between activities, outputs, outcomes, and impact. Use this worksheet to define each component, identify gaps or ambiguities, and guide further evidence review, piloting, or refinement. Start with Step 1 (narrative questions), then move to Step 2 (visual mapping). Step 3 is an iterative process to validate logic and improve clarity. Aim to be as detailed and thorough as possible—the goal is to deeply understand the intervention's inputs, outputs, and outcomes.

Step 1: Do a comprehensive review of program documentation and then respond to the following questions

a. What is the goal that the program wants to achieve?

.....

.....

b. What is the long-term objective or "ultimate outcome"? Example: "Improved children's health outcomes"

.....

.....

c. What does the program do?

.....

.....

d. What are the key activities (including prep activities)? Example: "SMS reminders sent to parents before vaccination days"

.....

.....

e. What are the products/services directly produced (outputs)?

.....

.....

f. What are the underlying conditions that must be true for key links in the theory of change to hold (key assumptions)?

.....

.....

# Worksheet 1.2: Theory of Change Review (Part 2)

Available in IDinsight The Impact Measurement Guide https://guide.idinsight.org/theory-of-change/

Step 2: Graphically represent the inputs, outputs, outcomes and goal

Create a flow diagram with labeled boxes and arrows to represent how inputs and activities lead to outputs, outcomes, and the ultimate goal.

- Arrange left to right to show logical progression.
- Draw arrows to represent causal relationships between: Activities -> Outputs -> Outcomes -> Goal. Be sure every arrow tells a logical "because" story.
- List the assumptions beneath the arrows to which they apply. Remember: The assumption describes the underlying conditions that must be true for the causal relationship to hold.
- Use parallel tracks for different delivery streams if needed.
- Indicate timing (e.g., short-term, medium-term, long-term).

Step 3: Collaboratively review and refine the theory of change through team and expert inputs.

Key informant interviews can be used to gather inputs from external experts. For the program team, organizing a theory of change workshop to interrogate and refine the draft theory of change can be a powerful way to gather input and foster alignment. The questions below can serve as a guide for external or internal discussions:

Step 4: Review and refine the theory of change

a. Are all key steps (e.g., inputs, activities, and outputs) included? Any missing links?

.....

.....

b. Are all nodes concise and specific?

.....

.....

c. Are statements written in clear, passive syntax?

.....

.....

d. Are your nodes labeled (e.g., A1.1.1) for future reference?

.....

.....

e. Have you noted feedback loops or two-way relationships?

.....

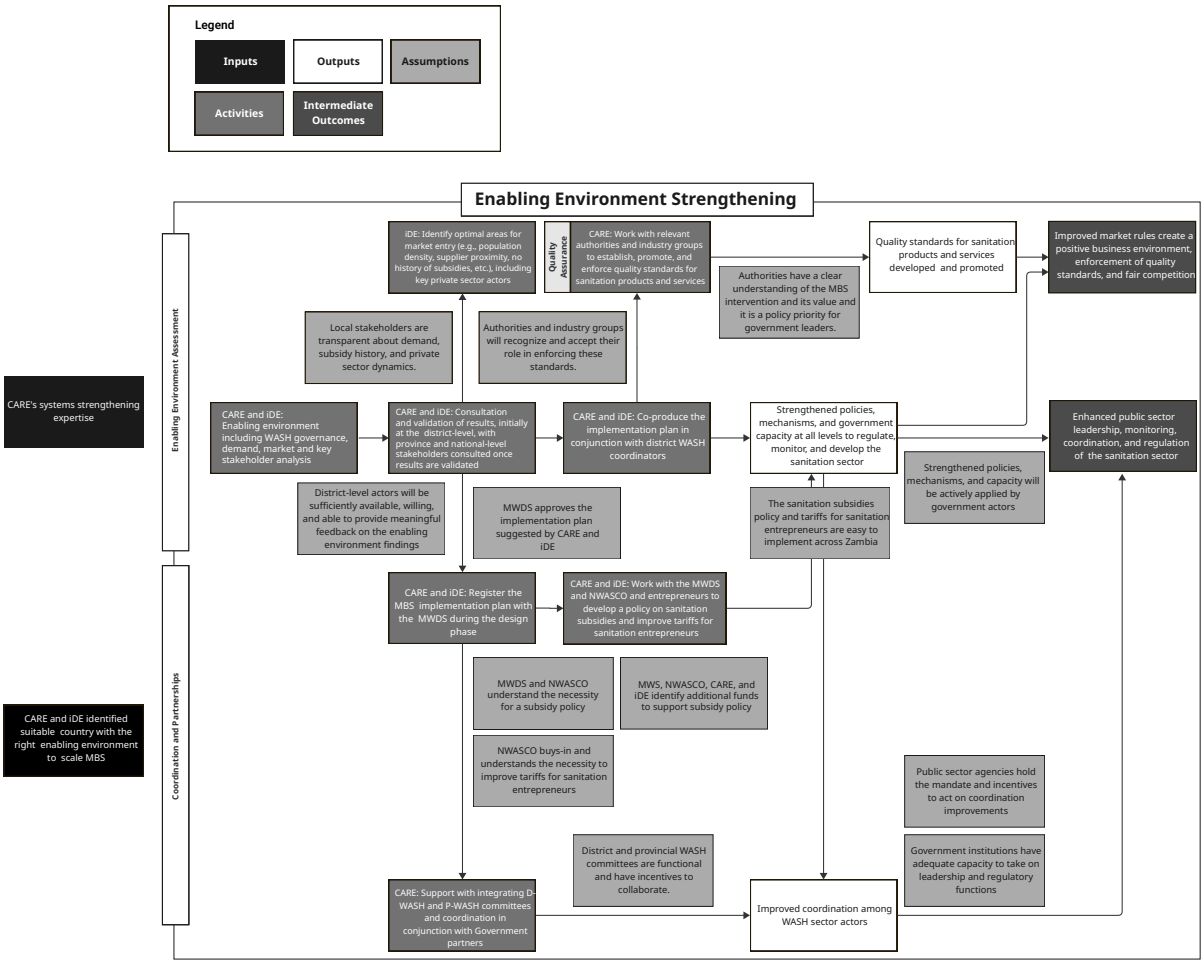
.....

# Worksheet 1.2: Definitions: Theory of Change Review

| Category | Definition and Notes | Examples |
|---|---|---|
| Input | • The financial, human, material and information resources needed to deliver the program successfully and achieve the desired outputs and outcomes | Funding, people, material |
| Activity | • Actions taken or work performed through which inputs are mobilized to produce outputs. | Active verbs such as: procure, hire, build, monitor<br>• Design material that is sensitive to the needs of women and girls |
| Output | • Direct products or services stemming from the activities of a program | • Water and sanitation facilities built/refurbished in rural areas of country X<br>• Community volunteers (f/m) trained to disseminate key messages on essential nutrition and hygiene actions in village Y, X, and Z of country X |
| Immediate Outcome | • A change that is expected to occur once one or more outputs have been delivered by the program.<br>• In terms of time frame and level, these are short-term outcomes, and are usually changes in capacity, such as an increase in knowledge, awareness, skills or abilities, or access to among the target population of the program.<br>• Immediate outcomes represent the first level of change that the target population experiences once the program starts delivering the outputs of a project. | • Improved knowledge of resilient agricultural-production practices among women-smallholder farmers in village X, of country Y<br>• Increased ability of health workers to address the nutrition challenges of women and children, especially girls in county Z |
| Intermediate Outcome | • A change that is expected to logically occur once one or more immediate outcomes have been achieved.<br>• In terms of time frame and level, these are medium-term outcomes that are usually achieved by the end of a program, and are usually changes in behaviour, practice or performance among the program's target population. | • Improved use of essential maternal health services, including those related to sexual and reproductive health, by women in village Y of country X<br>• Enhanced equal access to safe, quality education for girls and boys in crisis-affected province Y of country X<br>• Increased use of business development and financial services by micro enterprises, particularly those led by women, in province Y of country X |
| Ultimate or Final Outcome | • The highest-level change to which a program contributes through the achievement of one or more intermediate outcomes.<br>• The ultimate outcome usually represents the raison d'être of an organization, policy, program, or project, and it takes the form of a durable change among the people the program serves.<br>An ultimate outcome usually occurs after the end of the project, but should, when feasible, still be measured during the life of the project as changes may occur earlier. | • Enhanced economic prosperity for those with low incomes, particularly women and youth, in country X<br>• Increased food security of food insecure populations in region Y of country X<br>• Improved equal health of girls and boys under age five in rural areas of region X<br>• Improved learning outcomes of all girls and boys in crisis-affected province Y of country X<br>• Increased freedom of underrepresented women, men, girls and boys in country X<br>• Enhanced well-being of women in village Y of country Z |

*Note: A well-designed theory of change is a crucial first step in developing a program. Any organization can develop a theory of change, critical thinking, and a deep knowledge of the program and context are all that is required.*

# Worksheet 1.2: Sample: Theory of Change Review (lean MBS)



# Worksheet 1.3: Competitive Landscape Assessment – Developing a Longlist and Shortlist (Part 1)

This worksheet is intended to guide you through the first two steps of the competitive landscape assessment – compiling a longlist of interventions that address the same issue or challenge as yours, and then selecting a shortlist of interventions for deeper review and comparison with your own intervention.

Step 1: Develop the Longlist

In this initial stage, you should include any interventions that are highly similar to CARE's intervention (e.g., same primary objective, same target population, similarities in programming approach, etc.) as well as interventions that are somewhat similar (e.g., same objective, but very different programming approach) or dissimilar (e.g., your primary objective is a secondary outcome of another intervention).

1. Identify competitor interventions from all three categories outlined above by:
    a. Reviewing websites and resources by key issue area networks or organizations. For instance, there are often existing "what works" networks – such as the CALP Network for cash and voucher assistance or the What Works to Prevent Violence Against Women and Girls consortium.

    .....

    .....

    b. Keyword searches on the websites from multilaterals (e.g., UN, World Bank, etc.) and international non-governmental organizations (e.g., CARE, IRC, OXFAM, or WorldVision).

    .....

    .....

    c. Reviewing academic papers and meta-analyses for interventions in relevant issue areas identified through the multilateral/INGO sources or Google Scholar keyword search. J-PAL's library of rigorous evaluations can be found here https://www.povertyactionlab.org/evaluations. Similarly 3ie's https://www.3ieimpact.org/evidence-hub offers existing evidence summaries or evidence gaps maps.

    .....

    .....

    d. Crowd-sourcing interventions through sector experts at CARE and beyond.

    .....

    .....

2. For each of the identified interventions, collect basic information—ideally in an excel sheet—for further selection and comparison purposes, including:
    a. Basic programmatic information: Who is/are the implementing organization(s)? In which countries is the intervention being implemented? What is the reach of current implementations?

    .....

    .....

    b. Evidence of Impact: What is the strongest evidence that this intervention works? What is the evaluation type (RCT, quasi-experimental, internal monitoring and evaluation)? What is the summary of outcomes and key findings?

    .....

    .....

# Worksheet 1.3: Competitive Landscape Assessment – Developing a Longlist and Shortlist (Part 2)

c. Evidence of Scalability: What signals indicate the model can scale beyond its original context? For instance:

.....

.....

i. Simplicity of Model: Is the model easy to deliver and adapt? Is it cost-effective and lean?
ii. Evidence of Cost-Effectiveness: Has the model undergone a cost-effectiveness analysis? What were the results? Is it competitive against other similar interventions?
iii. Adoption by Others: Has the model been taken up by governments, INGOs, or community-based organizations? If so, where and under what conditions?
iv. Flagship Model of a Peer Organization: Is this intervention widely implemented across a major INGO's programs? (e.g., CARE's VSLA model or others recognized in the sector)

Step 2: Identify the Shortlist

Based on the longlist, create a shortlist of 5-6 interventions for deeper review and comparison with CARE's intervention. Select interventions based on the following criteria:

1. For interventions that are very similar to CARE's intervention, shortlisted interventions should have either rigorous evidence of positive impact (e.g., experimental or quasi-experimental) OR evidence of scalability.

.....

.....

2. For interventions that are somewhat similar or dissimilar to CARE's interventions, shortlisted interventions should have both rigorous evidence of positive impact AND evidence of scalability.

.....

.....

This shortlisting may still result in a list longer than 5-6 interventions. If that is the case, further narrow down by:

- Eliminating 'duplicate' interventions – that is, multiple competitor interventions that are all very similar in their design and programming approach. This ensures a diverse representation of intervention types that are relevant and comparable to your intervention.
- Consulting CARE experts with a good understanding of the sector and issue area, who can advise on which interventions to focus on as the most plausible competitors to your own intervention.

# Worksheet 1.3: Worksheet 1.3: Competitive Landscape Assessment – Shortlist Deep Dive (Part 3)

The guidance in this worksheet should be applied to each of the shortlisted interventions. Fill in the sections below using available documentation, evaluation reports, and interviews with program leads or partners. Consider using AI to help with this work, making sure to quality check AI outputs.

1. Summarize the program context and program logic. What is the theory of change of the model? How is the intervention achieving its results/ impact? What are the main components or activities of the intervention? What is the implementation context and are there any unique and context-specific aspects? [Refer to the theory of change Review (Part 1) worksheets for tips and questions on how to unpack the program logic].

   .....

   .....

2. Summarize the evidence of impact and cost-effectiveness. Use the following prompts (adapted from the Evidence of Impact Review worksheet) to guide your review of each shortlisted intervention: What sources and types of evidence are available (e.g., RCT, quasi-experimental, qualitative, meta-review)? What sources and types of evidence exist (e.g., RCT, quasi-experimental, qualitative, meta-review)? Is the evidence contextually similar (geography, population, delivery setting, timing)? What outcomes or lessons were reported? Did they affirm or challenge assumptions? How strong is the methodology and analysis? Any limitations or biases? What cost data exist (per output or participant)? Is it financially viable for scale?

   .....

   .....

3. Conduct a SWOT analysis. Identify the strengths, weaknesses, opportunities, and threats of the competitor intervention relative to the CARE intervention. For strengths and weaknesses, directly examine how internal program attributes – those within the control of the implementing organization – of the competitor intervention compare to those of the CARE intervention. For opportunities and threats, analyze how external factors – those beyond the control of the implementing organization – may influence the competitor intervention.

   .....

   .....

The information gathered through this deep dive can be used in an unbiased assessment of whether it is best to partner with another organization to scale their intervention, pivot by incorporating elements of competitor interventions that make them more scalable, or proceed with scaling the CARE intervention.

# Worksheet 1.4: Cost Review (Part 1)

Use this worksheet to assess the financial feasibility of scaling your intervention. Complete one worksheet per country, delivery model, or institutional context where scale is being considered.

To conduct a thorough cost review, you need granular program cost data. Ideally, this includes phase-specific costs (e.g., inception, pilot, delivery) and clear unit costs per input, activity, and output (e.g., cost per VSLA member or farmer trained).
Ensure you're using data about actual costs incurred—what it actually cost to deliver a given activity—not just original budget estimates.

Of the different valuable cost review activities, there are some that program teams can do on their own, and others that require expert economic evaluation support.

The following can be done by program teams without expert support:

1. Cost Data Compilation:

   Compile all available cost data, separating one-off inception costs from delivery costs.

   - One-off costs are those that do not scale as the intervention scales and that are not recurring. These costs, which you will not continue to incur once scaling begins within a given country or context, include things like:
     - » The initial cost of designing a new intervention, or of adapting an existing intervention to a new country context
     - » Costs of securing initial government buy-in or support for implementation of the intervention
   - Delivery costs can be both fixed and variable, and do continue to be incurred as the intervention scales. These costs include things like staff salaries, training venue cost, participant per diems, fuel expenses, etc.

Organizing cost data in this way allows you to later accurately model the costs of scaling the intervention. One-off costs will not be included in scaling costs, fixed delivery costs are part of scaling costs but do not increase as scale increases, and variable delivery costs do continue to increase as scale or reach increase.

2. Qualitative Affordability Signals:

   - For government scaling, a quick way to assess its viability as a "payer at scale" is to compare your program's cost with the relevant ministry's annual budget—either in total or per participant. For example, if your program costs $1M annually to serve two districts, and the ministry's national budget is $10M, your program already consumes 10% of that budget despite limited coverage. Similarly, if your cost per participant exceeds the government's typical per-participant spending, it's unlikely the government can fund the intervention at scale.
   - For private sector scaling through market mechanisms, you can compare the cost of the good or service being offered with participant willingness to pay. For example, if your offer costs $100 per participant, but research show that the average participant is only willing to pay $50, your model may not be viable without subsidies or cost reductions. This gap offers an indication of whether the market can sustainably support your intervention at scale.
   - For scaling through local NGOs or civil society organizations, comparing program cost with their typical operating budgets offers a quick way to assess viability. For instance, if your program costs $500,000 annually, but local NGOs in your target region typically operate on budgets of $100,000–$200,000 per year, it's unlikely they could adopt and sustain your intervention without significant external funding or cost reductions. This comparison helps gauge whether these organizations can realistically serve as implementers or funders at scale.

# Worksheet 1.4: Cost Review (Part 2)

3. Cost-Efficiency Analysis

   Where granular cost data is available, analyze cost per output (e.g., cost per latrine sold, per savings group established). This is not a one-off exercise, but something that should be part of routine monitoring as you experiment with strategies to improve efficiency.

   - Output:

   .....

   - Cost per unit:

   .....

4. Benchmarking

   If data regarding the per-unit costs of similar interventions (cost per teacher trained, latrine sold, child vaccinated etc.) is available, determine the maximum cost per participant or output that would allow your intervention to remain competitive. This can help determine how cost-competitive your intervention is relative to alternatives and can inform decisions about program design and iteration.

   - Comparator intervention and cost per unit of output or per participant:

   .....

   - Competitive cost per unit/participant threshold:

   .....

The following analyses will likely require expert economic evaluation support:

1. Cost-Effectiveness Analysis: Calculate cost per unit of outcome achieved (e.g., cost per percentage-point reduction in malnutrition).
2. Scale-Scenario Projections: Model how economies of scale, integration with government systems, or technological innovations could alter costs and obtain projections of the total funding required to implement at scale.

*Note: Use the insights gathered here to inform your Phase 3 Systems Orchestration strategy—particularly around funding pathways, stakeholder engagement, and implementation phasing.*

# Worksheet 1.5: Scaled Theory of Change Design (Part 1)

Use this worksheet to identify opportunities to simplify the intervention and develop a lean "optimal fidelity model." Work with your team and stakeholders to reflect on how the current theory of change might be simplified for scale without compromising core impact. This process helps identify which components are essential ("non-negotiables"), and which are context-dependent or optional ("nice-to-haves").

1. Prioritize Core Components of Your Theory of Change: Review your existing or "status quo" theory of change. Identify which activities and outputs are essential for achieving impact—and which may be simplified, adjusted, or removed for scale.

   .....

   .....

   Workshop Activity 1: Each participant reviews the theory of change independently and votes on each component:
   - Essential to achieving impact
   - Contextual and may require adaptation
   - Optional or non-core (e.g., "bells and whistles")

   Workshop Activity 2: Facilitate a group discussion to align on which elements are core and which can be adjusted or removed. Document the consensus.

   *Note: Use the group's findings to develop a streamlined version of the theory of change—focusing on simplicity, affordability, and what's truly required to drive outcomes at scale.*

2. Consider Opportunities in the Enabling Environment

   Identify external conditions that could support simplification. For example:
   - Can existing government systems or personnel deliver parts of the program (e.g., health workers, ag extension officers)?
   - Are there delivery partners or networks that could reduce implementation costs? Be reflected in your theory of change.

   .....

   .....

3. Revisit Key Assumptions for Scale

   Which assumptions from your current model might not hold at scale? (Even if they work in your current pilot or implementation phase.)
   - Are there assumptions about staffing, timing, infrastructure, engagement, or costs that need re-testing?

   .....

   .....

# Worksheet 1.5: Scaled Theory of Change Design: Assumptions Review (Part 2)

Use this worksheet to document and assess key assumptions between steps in your theory of change. This will help surface potential risks and prioritize areas for testing or further validation. For each assumption linking activities or steps in your theory of change, answer the questions below. Focus on assumptions that could critically affect implementation or impact.

1. Which theory of change steps does this assumption connect? (List the "From" and "To" steps the assumption bridges)

2. What is the assumption? (State the assumption clearly and specifically)

3. Is this assumption testable? Yes or No? (If no, explain why)

4. What evidence currently supports this assumption? (List any data, documents, or experience that suggest the assumption holds true)

5. What evidence contradicts this assumption (if any)? (List any data, documents, or experience that suggest the assumption holds true)

6. How confident are you that this assumption holds true? High or Medium or Low (Briefly explain your rating)

7. If the assumption fails, how disruptive would it be? High or Medium or Low (E.g., Would it delay implementation? Undermine outcomes?)

8. Should this be a priority for testing or mitigation? Yes or No? (Why or why not?)

# Sample Worksheet 1.5: Scaled Theory of Change Design: Assumptions Review (Part 2)

| # ToC node | Starting node | Node type | # ToC node | Ending node | Node type | # ToC link | Assumptions<br>Critical factors or conditions for the link between two nodes to hold | Priority<br>Prioritized if likelihood of not holding is medium or high, and potential for disruption is high | Likelihood assumption does not hold | Level of disruption if assumption doesn't hold |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Identifying a viable implementation area | Activity | 2 | Seeking approval from district and sector-level gatekeepers | Activity | i | There is information on the prevalence of IPV at the required level of target areas | NO | Medium | Medium |
| 2 | Seeking approval from district and sector-level gatekeepers | Activity | 3 | Mapping active VSLAs | Activity | ii | Authorities have a clear understanding of the program and its value | YES | Medium | High |
| 8 | Pre-testing materials | Activity | 9 | Identifying and selecting eligible participants | Activity | viii | There are clear eligibility criteria for participant selection | NO | Low | Low |
| 11 | Post-curricula ongoing support | Activity | 13 | - Couples become aware of negative effects of IPV | Output | xii | Couples regularly participate and actively engage in all training exercises/ activities | YES | Medium | High |
| | | | 14 | - Couples aware of non-violent conflict resolution techniques | | | | | | |
| | | | 15 | - Increased knowledge on ways to respond to IPV | | | | | | |
| 13 | - Couples become aware of negative effects of IPV | Output | 34 | - Better communication | Outcome | xiii | Couples willing to apply acquired knowledge in their relationships | YES | Medium | High |
| 14 | - Couples aware of non-violent conflict resolution techniques | | 35 | - Increased trust among couples | | | | | | |
| 15 | - Increased knowledge on ways to respond to IPV | | 36 | - Share household decision making among couples | | | | | | |
| 16 | - Couples aware of IPV triggers and ways of managing them | | | | | | | | | |
| 44 | Increased IPV reporting and service uptake among IPV survivors | Outcome | 56 | Improved physical and mental health of women and children | Final Outcome | xxxv | Victims are supported to recover | NO | Medium | Low |

# Worksheet 1.6: Enabling Environment Review

Use this worksheet to assess the broader ecosystem factors that influence whether your intervention can scale sustainably. It is particularly useful in the lead-up to Phase 3 (Systems Orchestration), where policy integration, financing, and delivery readiness become critical. Complete one worksheet per country, system, or institutional context where scale is being considered.

1. Policy Environment: Are there national or sub-national policies, plans, and strategies that support the intervention? What are these?

2. Doers at Scale: Which institutions (e.g., government, civil society, private sector) have the mandate and capacity to deliver the intervention at scale? Are they engaged? If not, how could they be engaged?

3. Payers at Scale: Who are the potential funders of the intervention at scale (e.g., government budgets, donor mechanisms, corporate social responsibility (CSR) initiatives, user fees)? How reliable are these funding sources?

4. Systemic Incentives and Risks: What political, economic, or sector-specific incentives might support—or obstruct—scale-up and sustained implementation? Include risks such as leadership turnover, elections, or cross-sector misalignment.

5. Operational Ecosystem
   What existing systems (e.g., HR, procurement, digital platforms) need to be leveraged or strengthened for successful delivery at scale?

*Note: Use the insights gathered here to inform your Phase 3 Systems Orchestration strategy—especially around stakeholder engagement, design plans, and implementation phasing.*

## Phase 2: Test

# Workshop Session 2.1: Identify Testable Variants of the Scaled Theory of Change

This worksheet supports teams in identifying specific, testable variants of a scaled theory of change that can be refined and tested in lean implementation trials. The goal is to create an optimized model that balances cost efficiency, feasibility, and impact. This session can be embedded in a scaled theory of change design workshop or run as a follow-on session.

Use the tools and facilitation steps below:

1. Dot Voting for SWOT Assessment: Provide participants with color-coded dots and ask them to tag theory of change nodes or links as a Strength, Weakness, Opportunity, or Threat. Remind the group that the purpose of dot voting is not to reach a final decision, but to surface priorities and shared perceptions. Afterward, facilitate a discussion on the "whys" behind commonly tagged assumptions and nodes.
2. Crowd sourced Efficiency Hypotheses: Prompt participants to suggest which inputs/activities might be done at lower cost without reducing effectiveness or, alternatively, to identify inputs/activities where small increases lead to relatively large outputs or intermediate outcomes.
3. Commenting Exercise: Participants place sticky notes or comments on specific theory of change components where cost-efficiency gains are likely possible.
4. Synthesis: Gather insights and observations from the workshop and pass to monitoring and evaluation and implementation leads for further analysis and strategic planning.

Post-Workshop Analysis and Planning

Following the workshop, the monitoring and evaluation and implementation teams should carry out:

1. Enabling Environment Recheck: Match the proposed inputs and activities in the scaled theory of change with the real constraints and enablers in the implementation context.
2. Monitoring Framework Setup: Identify intermediate outputs and define corresponding indicators and metrics. Identify specific activities that lead to the intermediate outputs, and group them into well-defined cost centers. Create protocols and templates for collecting cost data.
    - Implementation Plan Development:
    - Finalize the inputs, activities, time-frame, and delivery structure.
    - Incorporate a RACI framework for roles and accountability.
3. Process Mapping / Log-frame Development: Translate the implementation plan into a formal log-frame or workflow diagram that defines output indicators and tracks associated cost centers.
4. Synthesis Meeting: Bring together the monitoring and evaluation and implementation leads to shortlist testable variants that are feasible, lower- cost or high-yield, and adaptable to local constraints.

# Worksheet 2.2: Test Lean Variants through Field Trials (Part 1)

This worksheet guides your team in designing and executing small-scale trials of cost-efficient intervention variants. These lean variants should build on your scaled theory of change and may be tested through a series of microtrials to validate feasibility, affordability, and effectiveness.

For each step, document key decisions, learning questions, and indicators to monitor. This process will help you iterate toward an optimal fidelity model—a version of the intervention that balances effectiveness, cost-efficiency, and scalability in context.

1. Define the Core Intervention and Variants
    - What is the base core intervention?
      *(Briefly describe the standard intervention as currently designed. E.g., "Providing farmers with soil health cards after soil testing" or "Delivering 4 weekly financial literacy sessions.")*

      .....

      .....
    - What 2–4 variants will you test?
      *(Specify differences in quantity (dose/frequency), quality (level of intensity or expertise), delivery (channel, provider, or setting), or duration (short vs. long exposure). Avoid making too many changes at once.)*

      [Variant 1:] .....
      [Variant 2:].....
      [Variant 3:] .....
      [Variant 4:] .....

      Example:
        - Variant 1: SMS reminders only
        - Variant 2: SMS + one community meeting
        - Variant 3: Community meeting only
        - Variant 4: SMS + community meeting + demonstration plot

    - How do these variants adhere to design principles (simplicity, affordability, scalability)?
      *(Explain how each variant maintains feasibility for large-scale roll-out, avoids unnecessary complexity, and remains low-cost.)*

      .....

      .....
2. Design Microtrials
    - Describe the target population and sample size for each variant:
      *(Be explicit: "30 farmers per variant" or "4 schools, each with 50 students." The sample should be 'just big enough' to generate directional insights—statistical power is less critical than speed of learning at this stage.)*

      .....

      .....

# Worksheet 2.2: Test Lean Variants through Field Trials (Part 2)

- What will serve as the comparison/control unit?

  *(Specify whether you will use:*

  *– Status quo program*

  *– No intervention group*

  *– Another variant as implicit comparison)*

  .....

  .....

3. Implement Trials and Monitor
   - Duration of each microtrial (e.g., 1–3 months):

     *(Choose a period short enough for rapid feedback, but long enough to observe early signals of feasibility, adoption, or satisfaction.)*

     .....

     .....

     *(Examples: digital dashboards, mobile surveys, direct observation checklists, FGDs/IDIs, low-cost outcome measurement.)*

     .....

     .....

4. Analyze and Compare
   - Analysis plan (A/B test, descriptive comparison, etc.):

     *(Keep methods lean. Even simple cross-tabulations or difference-in-means can generate actionable insights. Reserve rigorous causal analysis for later stages.)*

     .....

     .....

   - What indicators will you track (outputs, outcomes, cost)?
     - Outputs: e.g., # of participants attending, # of inputs delivered
     - Short-term outcomes: e.g., knowledge gain, initial behavior change, satisfaction
     - Costs: delivery cost per participant, cost per unit of change

     .....

     .....

5. Iterate and Refine
   - Which variants will be dropped or improved upon?

     *(Decisions should be based on early signals of feasibility, uptake, and cost, even if long-term outcomes aren't yet visible.)*

     .....

     .....

   - What changes will be made to the core model based on insights?

     *(E.g., dropping high-cost components, simplifying delivery, adjusting frequency.)*

     .....

     .....

# Worksheet 2.2: Test Lean Variants through Field Trials (Part 3)

6. Repeat for 2–3 Cycles
   - What adaptations will you test in the next cycle?
     *(Build incrementally. Cycle 2 variants should reflect refinements based on Cycle 1 evidence.)*

     .....

     .....
   - What evidence do you hope to validate?
     *(Examples: "Demonstration plots increase adoption compared to SMS only" or "Low-cost peer-led delivery performs as well as expert delivery.")*

     .....

     .....
7. Finalize the Optimal Fidelity Model
   - Which variant(s) proved to be most cost-effective and scalable?

     .....

     .....
   - What evidence supports your final selection?
     *(Summarize learning across cycles, showing how the evidence converged toward a model that balances effectiveness with affordability and scalability.)*

     .....

     .....

# Worksheet 2.3: Final Design Testing and Multi-Site Roll Out (Part 1)

Use this worksheet guides your team through validating the final implementation model before scaling. Validation ensures that your chosen intervention design is feasible, effective, and cost-efficient in practice, and that it can be generalized across contexts.

The process has two stages:

1. Single-Site Validation — a "proof of concept" trial in one representative setting
2. Multi-Site RCT — testing generalizability across diverse contexts

Each section prompts you to record reasoning, decisions, and evidence. Insights from this process should directly inform your Phase 3 scaling strategy and investment decisions.

Stage 1: Single-Site Validation

1. Site Selection Criteria
   - Why was this site selected for validation?
     *(Explain the rationale: Does it represent a "typical" site? Was it chosen for ease of access, strong partner buy-in, or existing data infrastructure?)*

     .....

     .....
   - Which features make it suitable?
     *(Examples: availability of trained staff, good baseline data systems, accessibility for monitoring, demographic representativeness.)*

     .....

     .....

2. Evaluation Design Options
   - Select the most appropriate design (circle/highlight one):
     - Randomized Controlled Trial (preferred) – e.g., randomizing households, schools, or communities
     - Quasi-experimental – e.g., matched comparison groups, regression discontinuity, stepped-wedge rollout
     - Structured qualitative process evaluation – e.g., in-depth interviews, focus groups, implementation fidelity reviews

   - Reasoning:
     *(Why is this design appropriate given the resources, timeline, and decision-making needs? E.g., "A short RCT is feasible because we can randomize across farmer groups," or "Qualitative validation is needed first to refine outcome measures before randomization.")*

     .....

     .....

# Worksheet 2.3: Final Design Testing and Multi-Site Roll Out (Part 2)

3. Validation Criteria
    - How will you measure:

    a. Fidelity of implementation?
    *(Examples: proportion of sessions delivered as planned, adherence to protocols, delivery consistency across staff.)*

    b. Effectiveness of outcomes?
    *(Examples: knowledge gain scores, adoption of practices, uptake of services, short-term yield or income changes. Use measurable, early outcomes aligned with your theory of change.)*

    c. Efficiency and acceptability?
    *(Examples: cost per participant, staff time requirements, participant satisfaction surveys, dropout/non-compliance rates. Ensure data on both implementer burden and participant experience is captured.)*

Stage 2: Multi-Site Randomized Controlled Trial (RCT)

This stage tests whether the validated intervention model from Stage 1 works consistently across diverse contexts. A multi-site RCT helps establish generalizability, ensures results are not driven by unique site conditions, and generates robust evidence for scaling and investment decisions.

1. Site Selection
    - What criteria will you use to select sites for multi-site testing?
    *(Examples: diversity in geography, socio-economic conditions, partner capacity, or delivery infrastructure. Aim for variation that reflects the contexts where scaling is likely to occur.)*
    - How will you ensure diversity and representativeness across sites?
    *(Consider balancing rural/urban, high/low resource settings, or regions with different cultural norms. Avoid over-representing "easy" sites only.)*

# Worksheet 2.3: Final Design Testing and Multi-Site Roll Out (Part 3)

2. Implementation Planning
    - What steps are required to ensure implementation readiness at each site?
      *(Examples: staff training, material adaptation to local language, alignment with local authorities, ensuring data systems are functional before rollout.)*

    - What ethical approvals or protocols are needed?
      *(Examples: IRB/ethical board approval, informed consent processes, child protection protocols, data privacy compliance. Anticipate differences in national regulations across sites.)*

3. Indicators and Analysis
    - What impact indicators will you use?
      *(Specify primary and secondary outcomes, e.g., adoption of practices, learning gains, health outcomes. Ensure indicators are harmonized across sites for comparability.)*

    - How will cost-effectiveness and consistency be evaluated?
      *(Examples: cost per participant, cost per unit of outcome, incremental cost-effectiveness ratio (ICER). Consider subgroup analyses to test heterogeneity of effects across sites.)*

4. Synthesis and Use
    - How will you use these findings to guide investment and broader rollout?
      *(Examples: identify contexts where the model performs best, adapt for lower-performing contexts, provide funders with robust business cases for scaling. Findings may also inform Phase 3 investment strategy, partner selection, or refinement of theory of change assumptions.)*

# ACKNOWLEDGEMENTS

This Scaling Playbook was developed to help teams across CARE design and implement scalable solutions that deliver catalytic impact. Its creation was made possible through a collaborative process that drew on prior learning from ongoing workstreams supported by the learning partnership between CARE's Impact at Scale team and IDinsight.

The Playbook builds on practical experience and experimentation from teams across CARE and partner organizations. We would like to thank the CARE Program Strategy and Innovation colleagues who allowed us to dive deeply into their interventions, providing testing grounds, feedback loops, and honest reflections that shaped the evolution of the work and methods reflected here. We're also grateful to the many contributors who shared their insights, reviewed drafts, and helped bring this resource to life.

**Authors and Contributors**
Anita Sundari Akella, CARE Impact at Scale
Zahra Khan, CARE Impact at Scale
Rico Bergemann, IDinsight
Dr. Will Thompson, IDinsight
Julie Ramambason, IDinsight
Brittany Dernberger, CARE Impact, Learning, Knowledge and Accountability
Meg Watkins, CARE Transformation

**UX and Graphic Design**
Su Kutalek - IAS Design LLC
Patrick Mungai - Graphic Designer


care®