

2026:1k

Sida Evaluation

Nordic Consulting Group Denmark

Final Report from the Evaluation of Sida's work with Poverty

Central Evaluation of Sida's work with Poverty



Authors: Carsten Schwensen, Louise Scheibel Smed, John Rand, Anne-Lise Klausen and Ayla-Kristina Olesen Yurtaslan.

The views and interpretations expressed in this report are the authors' and do not necessarily reflect those of the Swedish International Development Cooperation Agency, Sida.

Sida Evaluation 2026:1k

Commissioned by Sida, Evaluation Team.

Published by: Sida

Copyright: Sida and the authors

Date of final report: 2026-06-16

Art.no.: Sida62873en

urn:nbn:se:sida-62873en

This publication can be downloaded/ordered from www.Sida.se/publications

Central Evaluation of Sida's work with Poverty

Final Report

Prepared by Carsten Schwensen, Louise Scheibel Smed,
John Rand, Anne-Lise Klausen and Ayla-Kristina
Olesen Yurtaslan, Nordic Consulting Group Denmark.

June 2026

EXECUTIVE SUMMARY	II
EVALUATION METHOD	II
KEY FINDINGS	II
CONCLUSION	III
RECOMMENDATIONS	IV
ABBREVIATIONS	V
1 INTRODUCTION	1
1.1 EVALUATION FOCUS AND SCOPE	1
1.2 METHODS AND WORK STREAMS	2
1.3 STRUCTURE OF THE REPORT	4
2 SIDA'S CONTRIBUTION TO POVERTY IMPACT	5
2.1 ZAMBIA	5
2.2 SOUTH SUDAN	7
2.3 WIDER SIDA EXPERIENCES	9
3 PREPAREDNESS FOR IMPACT ASSESSMENT	10
4 SHORTCOMINGS IN USE OF EVIDENCE FOR CONTRIBUTION DESIGN	15
5 REASONS FOR UNDERUSE OF EVIDENCE IN SIDA'S CONTRIBUTION MANAGEMENT	20
5.1 DELIVERY PRESSURE AND LIMITED TIME AND SPACE FOR REFLECTION	20
5.2 LIMITED INCENTIVES FOR PURSUING AND USING IMPACT EVIDENCE	20
5.3 STRONG RELIANCE ON PARTNER'S ASSESSMENTS, ANALYSIS AND REPORTING	21
5.4 DEPENDENCY ON STRONG DEVELOPMENT NARRATIVES – LACK OF CLARITY FOR USE OF EVIDENCE	22
5.5 LIMITED OR MODERATE USE OF THEMATIC NETWORKS, HELP DESKS AND SYSTEM TOOLS	23
5.6 LACK OF EVALUABILITY REQUIREMENTS IN CONTRIBUTION DESIGN	24
5.7 EVIDENCE CONSTRAINTS IN FRAGILE AND DATA-POOR CONTEXTS	25
6 INTERNATIONAL EXPERIENCES ON EVALUATION OF DEVELOPMENT IMPACT	26
6.1 A PRACTICAL MENU OF OPTIONS	26
6.2 COST ESTIMATES AND REQUIREMENTS	28
6.3 BINDING CONSTRAINTS AND POSSIBLE WAYS FORWARD	29
6.4 FINAL REMARKS	30
7 CONCLUSIONS	31
7.1 OVERALL CONCLUSIONS	31
7.2 SPECIFIC CONCLUSIONS ON SHORTCOMINGS IN SIDA'S WORK WITH EVIDENCE	33
8 RECOMMENDATIONS	35
8.1 OVERALL RECOMMENDATIONS	35
8.2 SPECIFIC AND ACTIONABLE RECOMMENDATIONS	37

Executive Summary

This report synthesises findings from the “Evaluation of Sida’s Work with Poverty”, conducted between August 2023 and June 2026.¹ The evaluation assesses how evidence is generated and used within Sida’s contribution management, and how this shapes the agency’s ability to achieve and demonstrate development impact. Sida’s work is grounded in a Multidimensional Poverty Approach (MDPA), recognising poverty as encompassing deprivations in resources, power and voice, opportunities and choice, and human security. Within this overall framework the evaluation addresses a central question: how Sida can improve its effectiveness in contributing to sustainable poverty reduction.

The evidence we focus on here relates to effectiveness, cost-effectiveness/efficiency and on “what works”, and less on “what to do” (e.g. relevance, composition of portfolio etc.).

Evaluation method

The evaluation is based on eight impact studies conducted by the evaluation team, four from Zambia (a relatively stable development context) and four from South Sudan (a fragile context),² a synthesis review of 18 additional Sida evaluations conducted between 2018 and 2025, a Sida staff survey (with 221 respondents), interviews with Sida staff at headquarter (HQ) and embassies, and an academic paper (developed by the evaluation team) on wider international experiences and trends related to evaluation of development impact. While, Zambia and South Sudan were focus countries for the impact case studies, the overall geographical scope of the assignment has been global. Although many positive experiences have been reported, focus has been on identifying critical shortcomings, and the learning from these, to inform Sida’s future work.

Key findings

Sida delivers important but uneven results: Across Zambia and South Sudan, Sida-supported programmes generate tangible benefits for poor populations, including improved incomes, food security, and access to essential services. However, these results are often partial, fragile, and insufficient to drive systemic change. In Zambia, interventions perform better as the systems are relatively stable but remain limited in scale and transformative impact. In South Sudan, contributions help sustain basic services under extreme conditions, but outcomes are difficult to verify and sustain and the efforts to build state capacity has fallen short of expectations.

¹ The evaluation team was composed of Carsten Schwensen (Team Leader), Partner Nordic Consulting Group Denmark; Louise Scheibel Smed, Partner Nordic Consulting Group Denmark; Anne-Lise Klausen, Partner Nordic Consulting Group Denmark; Ayla Kristina Olesen Yurtaslan, Partner Nordic Consulting Group Denmark; and John Rand, Professor, University of Copenhagen, Department of Economics.

² All eight impact studies have been published on Sida’s website.

Impact is constrained by design and evidence gaps: A consistent finding is that programme ambitions often exceed what design, context, and evidence systems can support. Interventions frequently combine multiple objectives without clearly specifying causal pathways or conditions for success. This leads to diffuse programme logic, limited focus, and challenges in both delivery and evaluation.

Evidence is available but underused: The evaluation finds that the core challenge is not a lack of evidence, but insufficient integration of evidence into decision-making. Analytical inputs - such as studies, evaluations, and sector analyses - are often available but only partially translated into concrete design choices. Programme design is frequently shaped more by strong development narratives than by context-specific empirical evidence.

Weak evaluability limits learning and accountability: Most programmes are not designed with evaluation in mind. Baseline data (disaggregated), clear theories of change (ToCs), outcome indicators, and comparison frameworks are often missing or misaligned. As a result, impact assessments often rely on reconstruction rather than programme-generated data, limiting the credibility and usefulness of findings, and increasing evaluation costs.

Monitoring systems prioritise reporting over learning: Monitoring systems across Sida's portfolio are generally effective in tracking activities and outputs, but weak in capturing outcomes (such as behavioural change and systemic effects). Heavy reliance on partner-reported data, limited triangulation, and insufficient use of national data systems create a verification gap and constrain deeper analysis.

Organisational factors reinforce underuse of evidence: The underuse of evidence reflects systemic organisational dynamics, including: i) Strong pressure to deliver and disburse funds; ii) Limited incentives for evidence-based adaptation; iii) Strong reliance on partner data and internal experience; iv) Limited tools for translating evidence into decisions; and v) Lack of explicit requirements for evaluability in design. Together, these factors mean that evidence often plays a supporting rather than guiding role in contribution management.

Conclusion

Sida demonstrates a clear ability to deliver development results. However, the strength, credibility, and sustainability of these results vary significantly. Weak integration of evidence into programming contribute to this.

The primary challenge is not generating more evidence, but ensuring that existing evidence systematically informs programme design, implementation, and adaptation. Without stronger integration of evidence, programmes risk remaining over-ambitious, insufficiently targeted, and difficult to evaluate.

Recommendations

To strengthen its contribution to poverty reduction, Sida should prioritise three mutually reinforcing changes:

Embed evaluability and evidence in programme design: i) Treat evaluability as a core requirement from the outset; ii) Develop clear ToCs with testable assumptions; and iii) Align indicators and data systems with intended outcomes.

Strengthen evidence quality and usability: i) Systematically triangulate partner data with independent sources; ii) Improve access to and use of national and administrative data; and iii) Develop practical tools for staff to translate evidence into decision-making.

Align organisational incentives with evidence use: i) Introduce clearer requirements for evidence-based design and adaptation; ii) Create space for reflection and learning within programme cycles; and iii) Strengthen focus on effectiveness, efficiency, and cost-effectiveness.

Overall, enhancing Sida's development impact requires shifting evidence from mainly a reporting tool towards a central driver of decision-making - ensuring that programmes are realistic, focused, and capable of delivering sustained and verifiable change for people living in poverty.

Abbreviations

BGFZ	Beyond the Grid Fund Zambia
CSO	Civil Society Organisation
GBV	Gender-Based Violence
GoZ	Government of Zambia
HMIS	Health Management Information System
HPF	Health Pooled Fund
HQ	Headquarter
HRBA	Human Rights-Based Approach
INCREASE	Increasing Climate Resilience in Energy & Agricultural Systems and Entrepreneurship Project
LCMS	Living Conditions Monitoring Survey
MDPA	Multidimensional Poverty Analysis
MIS	Management Information System
MSD	Market System Development
NGO	Non-Governmental Organisation
NPA	National Programme Assistant
NPO	National Programme Officer
PaCC	Peace and Community Cohesion
RALS	Rural Agricultural Livelihoods Survey
RCT	Randomised controlled trial
SAP	Strengthened Accountability Programme
SEK	Swedish Kroner
SGBV	Sexual & Gender-Based Violence
SRH	Sexual & Reproductive Health
SRHR	Sexual & Reproductive Health Rights
ToC	Theory of Change
ToR	Terms of Reference
UNDP	United Nations Development Fund
USD	United States Dollars
WEE	Women's Economic Empowerment Project

1 Introduction

This report brings together comprehensive experience from the “Evaluation of Sida’s Work with Poverty”, implemented in the period from August 2023 to June 2026.

Sida’s work with poverty takes point of departure in a Multidimensional Poverty Approach (MDPA) recognising that poverty is multi-faceted in various dimensions that go beyond only looking at economic resources.³ Through its MDPA, Sida strives to operationalise poverty as not only being about the lack of *resources* but also include deprivations within three other dimensions: *power and voice*, *opportunities and choice*, and *human security*.

All these poverty dimensions are interlinked. Sida defines a person living in poverty as deprived in the *resource* dimension and one or several of the three other dimensions. Further, Sida emphasises the importance of including *risk* and *vulnerability*, and thus *poverty dynamics*, as an important aspect of its approach to the multidimensional nature of poverty. Consequently, Sida’s target group is broadly defined as *people living in or at risk of falling into extreme poverty*.

1.1 Evaluation focus and scope

The evaluation aims at addressing three key evaluation questions:

- 1) *What are key findings on Sida’s contribution to development impact (including possible shortcomings)?*
- 2) *How well are contributions prepared for generating evidence (for ex-post/ impact assessment)?*
- 3) *Have there been shortcomings/gaps in the assessment/use/quality of evidence (for contribution design)? If yes - why?*

Thus, focus is on **evidence** → **decision-making** → **impact**

While impact is broadly defined by the poverty dimensions referred above, we mainly look at evidence from a “what works” perspective. Hence, we do not focus on “what to do” questions (e.g. relevance, composition of portfolio etc.), including the MDPA.⁴ Furthermore, evidence is here understood as both quantitative and qualitative evidence and mainly with focus on effectiveness and cost-effectiveness/efficiency issues.

While the evaluation team came across many good examples, focus has mainly been on identifying critical shortcomings and the learning from these with a view to inform Sida’s future work.

The evaluation has a global scope but with focus on two country cases that represent two different contexts: *Zambia*, where Sida has a long track record for bilateral development aid

³ This approach was first developed in 2017.

⁴ A recent independent evaluation of Sida’s MDPA (see EBA (2025), Evaluation of Sida’s Application of Multidimensional Poverty Analyses) found that the MDPA was well institutionalised and important for decision-making on country portfolios.

and where more long-term planning is possible due to the country context, and *South Sudan* which represents a more volatile and conflict-affected context.

1.2 Methods and work streams

The analytical starting point is that impact, evaluability, and evidence use are fundamentally interdependent. The ability to demonstrate and strengthen impact is not determined solely by what development interventions do, but by how they are designed, how change is conceptualised, and how data systems are structured to capture that change.

Together, these perspectives frame not simply whether Sida contributions generate impact evidence, but whether evidence is effectively integrated into the processes that shape contribution design, implementation, and adaptation. Addressing this aspect has allowed the evaluation to move beyond assessing individual interventions and towards identifying systemic features of Sida's approach to evidence and development impact.

This final report mainly builds on evidence and insight gathered from four complementary work streams in the evaluation:

i) Collection of **comprehensive empirical evidence from Sida contributions in Zambia and South Sudan**. This assessment is based on eight impact case studies - four from a development context in Zambia and four from a humanitarian context in South Sudan (refer Table 1) - as well as a portfolio analysis for each country, conducted by the evaluation team. These assessments have included fieldwork and own data collection in both Zambia and South Sudan, interviews at Sida Headquarter (HQ), and a comprehensive portfolio analysis in each country. Executive summaries from the eight impact case studies are included in Annex 1 and a summary of the impact results are presented in Annex 2. All eight impact case studies have been published and are available in full on Sida's website.

Table 1. Overview of impact case studies in Zambia and South Sudan

Contribution	Sector/subsector
Zambia⁵	
Musika	Agriculture, agribusiness, market system development (MSD)
UN Joint Programme on Social Protection in Zambia II (UN Social Protection Programme)	Social protection, social cash transfer, nutrition
Beyond the Grid Fund Zambia (BGFZ)	Renewable energy, MSD
Diakonia: Strengthened Accountability Programme (SAP)	Governance, gender, conservation farming, child labour, extractive resources
South Sudan⁶	
UNFPA Country Programme	Sexual reproductive health and rights (SRHR), gender-based violence (GBV)
Church of Sweden Nexus Pilot Project	Democratic participation, civil society, livelihood, resilience
UNDP Peace & Community Cohesion Project (PaCC)	Peace building/conflict prevention
Health Pooled Fund (HPF)	Health services and health system strengthening

ii) **A synthesis of evidence generated from Sida’s contributions.** This synthesis drew on the eight impact case studies from Zambia and South Sudan mentioned above, as well as on lessons extracted from 23 additional centralised and decentralised Sida evaluations conducted between 2018 and 2025. Relevant external evaluations, academic and research literature were also included in the analysis in the report. For the synthesis, additional interviews were conducted with Sida HQ staff, particularly to explore reasons why evidence may be underused in Sida’s work. These input from Sida staff provided the basis for development of the online survey questionnaire (see work stream iv below). The synthesis report mainly focused on “what works”, hence it did not discuss “what to do” questions (e.g. relevance, composition of portfolio etc.), including the Multidimensional Poverty Analysis (MDPA). The full synthesis report is available in Annex 3.

⁵ NCG/Sida (2026), Impact study of United Nations Joint Programme on Social Protection in Zambia. A case study as part of the Central Evaluation of Sida’s work with Poverty; Sida/NCG (2025), Impact study of Musika, a Zambian rural project. A case study as part of the Central Evaluation of Sida’s work with Poverty; NCG/Sida (2025), Impact study of Beyond the Grid for Zambia. A case study as part of the Central Evaluation of Sida’s work with Poverty; Sida/NCG (2026), Impact study of the Strengthened Accountability Programme (SAP), a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida’s work with poverty.

⁶ Sida/NCG (2026), Impact study of the Budi Nexus Pilot, South Sudan A case study as part of the Central Evaluation of Sida’s work with Poverty; Sida/NCG (2025), Impact study Health Pooled Fund, South Sudan. A case study as part of the Central Evaluation of Sida’s work with Poverty; Sida/NCG (2026), Impact study of the UNFPA Country Program for South Sudan. A case study as part of the Central Evaluation of Sida’s work with Poverty; Sida/NCG (2026), Impact study of the UNDP South Sudan, Peace and Community Cohesion project. A case study as part of the Central Evaluation of Sida’s work with Poverty.

iii) **An analysis of wider international experiences and trends related to evaluation of development impact**, building on both research and practice from development cooperation (but not from a humanitarian context). The paper was developed as part of this evaluation assignment. See full paper in Annex 4.

iv) **An online survey conducted among Sida staff** (central and decentralised level) with responsibility for contribution management to explore wider Sida staff experiences with evidence use. A total of 220 Sida staff responded to the survey which was conducted during April-May 2026. Annex 5 includes a summary of the survey questions and related responses.

The works streams have been implemented through two phases. The first phase focused on conducting the eight impact case studies as well as the portfolio assessments for Zambia and South Sudan. The second phase focused on verifying and synthesising findings from the Zambia and South Sudan impact case studies within a wider Sida and international context.

Together, these four works streams constitute the backbone of the analysis presented in this report, supplemented with input from other relevant sources.⁷

1.3 Structure of the report

Chapter 1 introduces the report, including a brief presentation of the methods and work streams and the report structure. **Chapter 2** focuses on key findings of Sida's contribution to impact, including possible shortcomings. The assessment is based on evidence from the eight impact case studies conducted by the evaluation team, as well as wider experiences from Sida evaluations. **Chapter 3** includes an assessment of how well the programmes covered in the assessment in Chapter 2 were prepared for impact assessment. **Chapter 4** discusses how well Sida contributions have been prepared for generating evidence (for ex-post/impact assessment), including of possible shortcomings/gaps in the use of evidence for contribution design. The assessment in this chapter is also based on the eight impact case studies, supplemented with wider evidence from Sida evaluations. **Chapter 5** contains a further discussion and assessment of identified main reasons for under-utilisation of evidence in Sida contribution management. This discussion is based on the evaluation team's own work with the impact case studies, as well as on the interviews conducted with Sida staff and the online survey. **Chapter 6** synthesises wider international experiences and trends related to evaluation of development impact and evidence use - based on both research and practice in a development context - with a view to complement the experiences from evidence use in Sida. Finally, in **Chapter 7** the conclusions from the evaluation are presented and in **Chapter 8** key recommendations are outlined.

⁷ E.g. Sida (2026), staff survey: Taking the pulse of Sida's evaluation culture.

2 Sida's contribution to poverty impact

Across the Zambia and South Sudan impact case studies, Sida's contributions have clearly generated important benefits for target groups, yet the depth, consistency, and sustainability of these impacts varied remarkably across programmes and contexts. Annex 1 includes a table overview of key results from the eight impact case studies conducted by the evaluation team. This chapter first discusses empirical evidence from Zambia and South Sudan and then wider experiences from other contexts.

2.1 Zambia

The four impact case studies from Zambia show that the supported programmes have been able to generate tangible improvements in specific areas of wellbeing, yet these gains have generally been partial and insufficient to drive broader structural change in poverty outcomes.

The UN Social Protection Programme⁸ - particularly when combining cash transfers with targeted nutrition support - produced consistent gains, notably in food security, dietary diversity, and child nutrition. These improvements highlight the importance of integrated approaches, as the addition of nutrition-sensitive components significantly enhanced the effectiveness of cash transfers. At the same time, the programme's overall impact was constrained by low transfer values, irregular payments, and behavioural factors such as misconceptions around nutrition, which even led to unintended negative effects. As a result - although the programme has helped cushion vulnerable households - it has not been strong enough to counteract rising poverty trends at the national level.

A similar pattern of concentrated but limited gains is evident in a programme for small-scale miners, the SAP.⁹ Here, increases in household income - driven almost entirely by mining earnings - stand out as the dominant achievement. This income growth was accompanied by improvements in assets, access to water, occupational safety, and reductions in child labour, as well as some progress toward more gender-equitable decision-making within households. However, the programme struggled to diversify livelihoods, as interventions such as conservation farming and goat rearing saw low uptake and minimal effects, largely due to structural constraints. Moreover, while some social outcomes improved, the programme did not manage to reduce the prevalence of gender-based violence (GBV), underscoring the difficulty of translating economic gains into deeper social transformation.

The Musika agricultural market development programme similarly achieved clear positive impacts, particularly for participating smallholder farmers, who benefited from improved market access, better access to finance, and rising incomes, alongside gains in livestock

⁸ NCG/Sida (2026), Impact study of United Nations Joint Programme on Social Protection in Zambia. A case study as part of the Central Evaluation of Sida's work with Poverty.

⁹ NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty.

ownership and food security.¹⁰ These results confirm that facilitating market participation can reduce poverty for engaged producers. However, the programme fell short of its more ambitious goal of driving systemic change in agricultural markets. Likewise, there was no clear evidence of improvements in productivity, technology adoption, or broader asset accumulation, and the observed market effects were limited and fragile, often benefiting actors who were already relatively well positioned. Furthermore, weaknesses in monitoring systems - especially the lack of gender-disaggregated data - meant that important dimensions of impact could not be assessed. As with the other programmes, the scale of positive effects was too limited to offset wider negative trends in national poverty.

Beyond the Grid for Zambia (BGFZ), illustrates both the strengths and limitations of market-based approaches to service delivery.¹¹ The programme achieved impressive scale, enabling more than one million people - many for the first time - to access basic off-grid energy services and experience improvements in daily quality of life. It also demonstrated the feasibility of stimulating private-sector engagement in underserved markets. However, these gains remained largely confined to basic access, with little evidence of knock-on effects in areas such as health, education, or income generation, despite these being central objectives. In addition, the programme struggled to reach the poorest and most remote populations and did not establish a sustainable model for last-mile delivery, limiting its inclusiveness and long-term transformative potential.

In Zambia, the programmes have operated in comparatively stable institutional and market environments, with relatively strong administrative systems, more predictable implementation conditions, and a longer history of sectoral engagement. These factors have created a foundation in which Sida's support could translate more reliably into measurable improvements. Income gains among smallholders in Musika, improvements in food security and child nutrition through the UN Social Protection Programme, and the expansion of basic energy access under BGFZ all illustrate how Sida's contributions produced tangible and verifiable results when embedded in systems with at least some degree of stability and data availability.

At the same time however, the evidence suggests that while the supported programmes were effective at delivering targeted benefits, they were less successful at achieving broader, systemic, and transformative impacts. Common challenges included limited scale or intensity of support, structural barriers to inclusion and uptake, and insufficient integration across sectors. These shortcomings were closely related to programme design. Interventions often combine multiple objectives - such as productivity growth, resilience building, gender empowerment, and climate adaptation - without fully specifying how these objectives interact or the conditions under which they can be achieved. Programmes such as SAP and Increasing

¹⁰ Sida/NCG (2025), Impact study of Musika, a Zambian rural project. A case study as part of the Central Evaluation of Sida's work with Poverty.

¹¹ Impact study of Beyond the Grid for Zambia. A case study as part of the Central Evaluation of Sida's work with Poverty.

Climate Resilience in Energy & Agricultural Systems and Entrepreneurship Project (INCREASE)¹² (another Sida evaluation of a programme designed to strengthen the social, economic and environmental resilience of smallholder farmers in Zambia by integrating climate-smart agriculture, renewable energy solutions, and support to women- and youth-led enterprises) illustrate how complex and multi-layered intervention logic can dilute focus and make it difficult to achieve and measure higher-level outcomes.¹³ In such cases, while specific components may perform well, the overall contribution to systemic change remains unclear.

2.2 South Sudan

In South Sudan, Sida has often made large, strategic funding decisions in environments where rigorous measurement of evidence would come at high costs - if at all possible. Donors, including Sida, when they operate in conflict affected settings calculate with trade-offs between their spending on activities - including saving lives - versus conducting rigorous measurements with advanced and expensive methods, noting that they are well aware of the potential risk of limited impact and sustainability. As a result, even contributions that delivered important service-level gains struggled to demonstrate impact, and several interventions were designed on assumptions that existing research could have challenged.

Across the four South Sudan cases, Sida-supported programmes did achieve important outcomes: The Health Pooled Fund and the UNFPA Programme contributed to the expansion of life-saving health and SRHR services nationwide; PaCC facilitated local peace dialogues and strengthened community structures. For the Budi Nexus Pilot, there was contradicting evidence on perceived improvements in security and livelihoods.¹⁴ In general, however, the larger programmes fell short of their long-term goal of building a government-led health system, due to weak state capacity, chronic underfunding, and repeated national and sub-national humanitarian crisis and conflict. Moreover, the achievements were difficult to verify, following the same logic across all cases; the tangible, people-oriented outputs and outcomes have in general been more easy to trace (service delivery for example), while the institutional or societal change processes (peacebuilding or capacity development of government institutions) have been more difficult to capture beyond output level.

¹² Sida/Nordic Consulting Group (Carsten Schwensen, Louise Smed) (2023), Midterm Evaluation of Increasing Climate Resilience in Energy & Agricultural Systems and Entrepreneurship (INCREASE).

¹³ NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20. NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty.

¹⁴ Sida/NCG (2026), Impact study of the Budi Nexus Pilot, South Sudan A case study as part of the Central Evaluation of Sida's work with Poverty; Sida/NCG (2025), Impact study Health Pooled Fund, South Sudan. A case study as part of the Central Evaluation of Sida's work with Poverty; Sida/NCG (2026), Impact study of the UNFPA Country Program for South Sudan. A case study as part of the Central Evaluation of Sida's work with Poverty; Sida/NCG (2026), Impact study of the UNDP South Sudan, Peace and Community Cohesion project. A case study as part of the Central Evaluation of Sida's work with Poverty.

The Health Pooled Fund and the UNFPA Programme were the most evaluability-ready contributions among the South Sudan cases, largely because they could draw on the national Health Management Information System (HMIS) and long-standing reporting structures.¹⁵ Even so, both lacked baseline data, relied on outdated national surveys, and operated in one of the weakest data environments globally.

By contrast, the Budi Nexus Pilot was not evaluability-ready: It lacked baseline, outcome indicators, independent data sources, and verification mechanisms.¹⁶ Monitoring was narrative and self-reported, and the evaluation team had no counterfactuals or conflict-trend data to assess change. In the case of PaCC, a baseline study existed (the endline study of an earlier UNDP project) which to some degree covered the PaCC project locations. The systematically collected output data was used by the Phase I Summative Evaluation, both by UNDP and Sida without communicating the limitation of the data and trying to triangulate and verify this with other sources, if and how these would exist. In the case of PaCC this limited the degree to which it could be substantiated whether the PaCC results were attributable to the support.¹⁷

In contrast to the experiences from Zambia, Sida's contribution in South Sudan is primarily visible in the maintenance and expansion of essential services in an environment characterised by fragility, conflict, and weak institutional capacity. In such contexts, the notion of impact needs to be understood differently. Rather than focusing on long-term transformation or systemic change, interventions often aim at stabilising service delivery and prevent deterioration in basic welfare conditions.

Programmes such as the Health Pooled Fund illustrate this dynamic. By financing and coordinating health service delivery across large parts of the country, the programme has contributed to improving access to primary healthcare, including maternal and child health services. Such outcomes are critical in a context where public systems are severely constrained. However, the ability to move beyond service delivery towards sustained system strengthening is limited by structural conditions, including insecurity, reliance on external actors, and weak government capacity. Similar patterns are observed in governance and protection interventions. Programmes supporting access to justice and GBV services improve access and coverage but operate within systems that remain structurally weak. As a result, while immediate outcomes are achieved, longer-term institutional transformation remains uncertain.

Taken together, the four South Sudan impact studies showed that impact, evaluability and evidence use are tightly interconnected, just as in Zambia. However, South Sudan's deep rooted conflict situation (myriads of local and national level conflicts), lack of good data

¹⁵ Sida/NCG (2025), Impact study Health Pooled Fund, South Sudan. A case study as part of the Central Evaluation of Sida's work with Poverty; Sida/NCG (2026), Impact study of the UNFPA Country Program for South Sudan.

¹⁶ Sida/NCG (2026), Impact study of the Budi Nexus Pilot, South Sudan A case study as part of the Central Evaluation of Sida's work with Poverty.

¹⁷ Sida/NCG (2026), Impact study of the UNDP South Sudan, Peace and Community Cohesion project.

availability, and limited access to project locations have constrained designs, even with adaptive project implementation (as was the case of PaCC), and the contributions have struggled both to deliver and to demonstrate impact. Even the strongest interventions produced service-level gains that were difficult to attribute or sustain, while the weakest lacked the foundations needed for learning. The lack of learning points generated in the Budi Nexus Pilot is a real shortcoming, given the project status as a “pilot”.

2.3 Wider Sida experiences

The broader review of 18 additional Sida evaluations reinforces these cross-country patterns. Across regions and sectors, Sida-supported programmes often generated important benefits: improved market access, expanded financial inclusion, strengthened service delivery, enhanced resilience, and increased participation of women and marginalised groups. Yet these gains were frequently fragile, and insufficiently linked to higher-level outcomes such as sustained income growth, systemic transformation, institutional strengthening, and long-term resilience. In many cases, programme ambitions exceeded what the evidence base could support. Designs often assumed that complex behavioural or systemic changes would occur without the enabling conditions required to sustain them. Moreover, the lack of robust monitoring systems limited the ability to track whether early gains were translating into deeper or more durable forms of change. Weaknesses in outcome measurement, limited attention to equity effects, and insufficient tracking of systemic dynamics meant that programmes could demonstrate activity-level progress but struggled to show whether this progress was leading to the intended long-term impacts.

Taken together, these patterns highlight a central insight: Sida’s contributions do generate important results, but the ability to measure and verify these results depends heavily on the quality of the underlying evidence, the realism of the design assumptions, and the robustness of the monitoring systems. Where these elements are strong, results tend to be clearer, more consistent, and more resilient. Where they are weak, results become harder to verify, and more vulnerable to contextual shocks. This underscores the importance of grounding programme ambitions in realistic assessments of context, evidence, and system readiness, and of ensuring that monitoring systems are capable of capturing not only outputs but also the deeper forms of change that Sida seeks to influence.

3 Preparedness for impact assessment

Across the Zambia and South Sudan impact cases, a consistent pattern emerges: even where programmes generated substantial amounts of data, they were rarely designed with ex-post impact assessment in mind (with few exceptions e.g. the cash transfer programme). Evaluability was often an afterthought rather than a foundational design principle. This meant that, when results/impact assessments were eventually commissioned, evaluators frequently had to reconstruct Theories of Change (ToCs) and baselines and rely on secondary data sources rather than programme-generated evidence.

The UN Social Protection Programme in Zambia provides the best example of a case where strong design foundations improve evaluability significantly.¹⁸ The intervention is relatively simple and well-defined, with clearly specified eligibility criteria and a direct transfer mechanism linking inputs to expected outcomes. This clarity is reinforced by the availability of household level baseline data and structured monitoring systems that track changes in consumption, food security, and household resilience. As a result, the programme is able to generate relatively robust evidence on its effects, allowing for assessment of how transfers translate into welfare improvements. Importantly, the simplicity of the intervention logic also facilitates attribution, as the link between intervention and outcome is both observable and measurable.

In contrast, programmes such as SAP and INCREASE highlight the challenges associated with low evaluability.¹⁹ These interventions combine multiple, often interrelated objectives - including productivity increases, market access, climate resilience, and gender empowerment - without clearly specifying how these objectives are causally linked or how progress should be measured at intermediate stages. As a result, the overall programme logic remains diffuse, making it difficult to isolate the contribution of individual components. SAP lacked the basic architecture required for credible impact assessment and an impact target was not explicitly defined. When the evaluation team conducted the impact case study, baseline data were no longer available in raw format, and they were incomplete and misaligned with the outcomes the supported interventions tried to influence - and outcome indicators were either missing or insufficiently defined. As a result, the impact case study depended heavily on retrospective reconstruction and new data collection rather than on a coherent monitoring system established from the outset.

Musika represented a different but equally instructive challenge: Although it accumulated one of the richest datasets in the portfolio - spanning a decade of household surveys - these data

¹⁸ NCG/Sida (2026), Impact study of United Nations Joint Programme on Social Protection in Zambia. A case study as part of the Central Evaluation of Sida's work with Poverty.

¹⁹ NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty; NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20.

were not originally designed for impact evaluation.²⁰ Inconsistencies across survey rounds, missing variables, and methodological shifts limited comparability and required the evaluators to invest significant effort in harmonising the data. The programme's evaluability therefore stemmed more from the sheer volume of data than from deliberate design choices.

BGFZ, meanwhile, invested heavily in real-time operational monitoring, but its systems were oriented toward tracking sales, installations, and service quality rather than poverty and welfare outcomes. This left major impact dimensions unmeasured, despite the programme's stated ambitions around economic and social transformation.²¹

In **South Sudan**, the evaluability challenges were even more acute. Many programmes lacked credible baselines altogether, and monitoring systems were fragmented, irregular, and heavily dependent on partner self-reporting. The ToCs were often implicit, making it difficult to trace contribution pathways and assess whether observed changes could plausibly be attributed to Sida's support. In such contexts, even when programmes delivered outputs or intermediate outcomes, the absence of systematic outcome tracking meant that higher-level effects remained largely speculative. The volatile operating environment further complicated matters: staff turnover, insecurity, and shifting implementation arrangements disrupted data collection and undermined institutional memory. As a result, impact assessments in South Sudan often relied on qualitative inference or triangulation rather than on structured, programme-generated evidence.

The broader set of Sida evaluations reviewed reinforced these findings: Across sectors and countries, the absence of clear ToCs, measurable indicators, and robust monitoring systems consistently limited the ability to assess impact, learn from implementation, and adapt programme design. Many contributions were designed with ambitious objectives but without the evaluability foundations needed to test whether these ambitions were realistic or achieved. In most cases, monitoring systems focused heavily on activities and outputs, with little attention to behavioural change, systemic effects, or sustainability.

Equity considerations, despite being central to Sida's institutional commitments, have often been weakly operationalised. While the contributions in general referenced gender equality, human rights, and inclusion in their design documents, they have often lacked the mechanisms needed to assess who benefited and who did not. Monitoring systems frequently failed to capture differences in outcomes across gender, age, disability, poverty status, and geographic location. As a result, evaluations are often unable to answer basic distributional questions. This limit both accountability and adaptive management, as programmes lacked the evidence needed to adjust strategies in response to emerging inequities. This is particularly visible in

²⁰ Sida/NCG (2025), Impact study of Musika, a Zambian rural project. A case study as part of the Central Evaluation of Sida's work with Poverty.

²¹ Sida/NCG (2025), Impact study of Beyond the Grid for Zambia. A case study as part of the Central Evaluation of Sida's work with Poverty.

programmes with broad or system-wide approaches - such as SAP and INCREASE - where objectives include improving productivity, resilience, and inclusion simultaneously, the absence of clearly defined target groups and disaggregated indicators makes it difficult to track differential impacts.²² Programmes may report improvements in yields, incomes, or participation rates, but without disaggregation by income group, gender, or geographic location, it is not possible to determine whether these gains are reaching the most vulnerable or simply benefiting those already better positioned to take advantage of programme opportunities. By contrast, programmes with more clearly defined targeting mechanisms demonstrate the value of integrating equity into design and monitoring.

In contrast, the UN Social Protection Programme provides an example of how clearly specified eligibility criteria, combined with structured data collection, enable analysis of distributional outcomes. In this case, the alignment between targeting, indicators, and data allows for a more robust assessment of how the programme affects vulnerable households.²³ The difference between such programmes and more diffuse interventions highlights that the issue is not technical feasibility, but the extent to which equity is embedded as a design requirement.

Taken together, these patterns point towards a wider systemic issue where evaluability is rarely embedded as a core design principle within Sida's programming. Instead, it is often treated as a technical add-on, addressed only when an evaluation is commissioned. This limits not only the ability to assess impact ex-post but also the potential for adaptive management during implementation.

Monitoring systems

Monitoring systems across Sida's portfolio are effective in tracking activities and outputs, but they are significantly weaker in capturing outcomes, causal mechanisms, and impact. This imbalance reflects a structural issue where monitoring frameworks are often designed primarily for accountability and reporting rather than for generating evidence that explains how and why change occurs. Large volumes of data are often collected, yet this does not translate into equally strong analytical insight.

A core driver of this problem is a misalignment between monitoring systems and programme logic. When ToCs are broad, implicit, or insufficiently specified, monitoring defaults to what is easiest to measure - activities, participation, and service delivery. This leads to detailed documentation of implementation progress but limited ability to assess whether intended outcomes are being achieved or through which mechanisms. Programmes such as SAP and

²² NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty; NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20.

²³ NCG/Sida (2026), Impact study of United Nations Joint Programme on Social Protection in Zambia. A case study as part of the Central Evaluation of Sida's work with Poverty.

INCREASE illustrate this challenge: although they include numerous indicators, these are not consistently structured around a coherent causal pathway, resulting in fragmented data that lacks analytical value.²⁴ Indicator selection further reinforces this pattern. Output indicators are often prioritised because they are measurable, attributable, and aligned with reporting requirements. Outcome indicators - such as behavioural change, productivity, income, or institutional performance - are more complex to define and measure, especially in multi-component programmes. In the absence of strong design frameworks, they are often excluded or included inconsistently, shaping monitoring practices toward documenting effort rather than assessing results. As a result, even when programmes generate positive changes at the local level, these changes remain difficult to quantify or attribute in a rigorous way.

A defining feature of programme monitoring is partner-reported data. While practical and often necessary in flexible and adaptive programming, this model introduces structural limitations. Data is generated primarily for operational and accountability purposes within partner systems, and indicators are mostly aligned with contractual reporting rather than broader evaluation needs. This influences both what data is collected and how it is interpreted. In this situation, there is an inherent tendency for partner systems to emphasise progress against targets, which may obscure complexities, unintended effects, and areas of weaker performance. At the same time, when verification of partner-reported data is limited, independent sources - national surveys, third-party monitoring, or impact evaluations - may not be consistently integrated. This creates a verification gap in which programmes produce substantial data but lack the means to substantiate claims about outcomes and impact. The survey with Sida staff indicates that assessments of projects often rely on partner reported data and that these are rarely verified.

Contextual differences shape how these challenges manifest. In South Sudan, weak national data systems, restricted access, and limited feasibility of independent verification mean that partner reporting is often the only viable option. One exception is the Health Pooled Fund which supported reporting on service coverage and use in the HMIS, enabling the programme to draw on national data to track service delivery outputs. However, without systematic data collection on quality of care - including clinic/health outcomes of patients - the Health Pooled Fund still struggled to link output tracking to higher-level health outcomes and impact on maternal and infant mortality (with the exception of tracking e.g. child stunting).

In Zambia, by contrast, stronger national data systems - such as the Living Conditions Monitoring Survey (LCMS) and agricultural datasets as the Rural Agricultural Livelihoods Survey (RALS) - offer opportunities for triangulation, but these are not systematically utilised.

²⁴ NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty; NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20.

Monitoring systems often operate in parallel to national systems rather than integrating with them, limiting possibilities for analytical depth even where data availability is not a constraint.

Partners in both the UN Social Protection Programme and Musika did however make use of respectively the LCMS and RALS data in design/monitoring, but these data sets were difficult to get access to by the evaluation team and only because the evaluation team already had access to RALS it was possible to use this data. Attempts to get access to LCMS was unsuccessful. Thus, not only availability but also access can constitute a challenge and requires the right connections.

The absence of well-defined treatment groups further complicates the assessment of impact. In many cases, interventions are implemented through broad-based or market-wide approaches, without clear distinctions between beneficiaries and non-beneficiaries. This limits the ability to construct comparisons, which are essential for establishing causal relationships. At the same time, baseline data is often either unavailable or not aligned with intended outcomes, making it difficult to assess change over time.

Overall, monitoring within Sida's contribution management tends to remain more compliance-oriented than learning-oriented. While effective for accountability and financial oversight, it is less effective as a tool for generating knowledge and improving programme performance. Addressing this imbalance requires shifting the focus from the quantity of data collected to its relevance, coherence, and analytical value. This means aligning indicators with clearly defined ToCs, integrating multiple data sources to support verification, and ensuring that monitoring outputs inform decision-making. Only by strengthening these elements can monitoring systems contribute meaningfully to understanding programme effectiveness and enhancing Sida's impact.

Taken together, the evaluation findings show that Sida's contributions achieve the strongest and most credible impacts where contributions are grounded in robust evidence, realistic intervention designs, and functioning monitoring systems. Where designs are overly ambitious and insufficiently evidence-based, contributions struggle both to deliver and to demonstrate results.

4 Shortcomings in use of evidence for contribution design

The evaluation findings show that evidence is not consistently or systematically used to inform contribution design, even in cases where a substantial body of analytical material is available. Across Sida's portfolio, most interventions are preceded by multiple types of analysis, including sector diagnostics, political economy assessments, market studies, and thematic reviews. However, these inputs are not consistently translated into concrete design decisions regarding programme scope, sequencing, targeting, and prioritisation. A recent survey conducted by Sida itself showed that 40% of staff had not consulted an evaluation as part of preparation of a new contribution/portfolio.²⁵

The Zambia cases illustrate a situation where good evidence has often existed but not been fully utilised. In several sectors, substantial bodies of knowledge were available (e.g. on conservation agriculture, goat rearing, market systems development, and social protection design). Yet, these insights were only partially integrated into contribution choices, resulting in interventions becoming either over-ambitious or misaligned with contextual realities. The consequence of this was not the absence of impact, but rather that impact became narrower, more fragile, and less transformative than intended.

Across both countries, contributions also suffered from evidence gaps that could have been filled with modest effort. Market analyses, and system diagnostics were often missing, even though they would have been feasible to generate early in the programme cycle. The absence of these foundational elements has limited the ability to design interventions grounded in a realistic understanding of context and later made it difficult to assess whether programmes were achieving their intended outcomes. These gaps were not inevitable but reflected a lack of systematic attention to evaluability and evidence needs during the contribution design stage.

A broader pattern emerging from all reviewed evaluations is the dominance of narrative over evidence in contribution choices. Many contributions have been built around compelling development narratives - systemic change, market system development, women's economic empowerment, productive use of energy, financial inclusion, climate-smart agriculture, organic certification, or GBV prevention. These narratives often carried strong intuitive appeal and aligned with Sida's strategic priorities. However, they sometimes overshadowed contradictory evidence or gaps in the underlying logic. Designs frequently began with a conceptual storyline and then selected interventions to fit that storyline, rather than grounding the design in empirical evidence about what was likely to work in the specific context.

One consequence of this stronger reliance on narratives than on evidence has been that interventions often have tended to persist even when early evidence has shown limited traction. Designs have remained overly ambitious, assuming systemic and behavioural change

²⁵ Sida (2026), staff survey: Taking the pulse of Sida's evaluation culture.

without sufficient enabling conditions. In this situation, adaptation has often become reactive rather than proactive, occurring only when problems became too visible to ignore.²⁶

Another important gap concerns a limited integration of cost-effectiveness into programme analysis and decision-making. Across the portfolio, there is relatively little systematic assessment of the relationship between resources invested and results achieved. While data on outputs and, in some cases, outcomes is available, it is rarely combined with cost data in a way that allows for meaningful analysis of efficiency or value for money. This limitation has several implications. At the level of individual programmes, it becomes difficult to assess whether different intervention components contribute proportionately to overall results. In complex programmes - such as the Women's Economic Empowerment (WEE),²⁷ SAP or INCREASE in Zambia where multiple strands of activity are implemented simultaneously - the absence of cost-effectiveness analysis makes it challenging to identify which components are most impactful relative to their cost.²⁸

This, in turn, limits the ability to refine programme design or reallocate resources towards the most effective approaches. At the portfolio level, the lack of cost-effectiveness data constrains Sida's ability to prioritise between interventions. Without comparable information on the costs and benefits of different programmes, decisions about resource allocation are more likely to be based on strategic alignment or historical precedent than on evidence of relative performance. This reduces the potential for evidence to inform not only programme-level decisions but also broader strategic choices.

The implications extend to questions of scaling and sustainability. Interventions that demonstrate positive outcomes cannot be easily assessed in terms of their scalability without understanding their cost structure. For example, even where programmes such as Musika or the UN Social Protection Programme generate measurable benefits, the absence of systematic cost-effectiveness analysis makes it difficult to determine whether similar approaches could be expanded or replicated under different resource constraints. This limits the ability to translate programme-level success into broader systemic impact.

Another critical area concerns adaptation and learning. While Sida operates in diverse and often dynamic environments, adaptive management is not consistently embedded as a

²⁶ NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty; NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20; NCG/Sida (Carsten Schwensen, Louise Smed) (2023), Midterm Evaluation of Women's Economic Empowerment (WEE).

²⁷ Sida/Nordic Consulting Group (Carsten Schwensen, Louise Smed) (2023), Midterm Evaluation of Women's Economic Empowerment (WEE).

²⁸ NCG/Sida (2026), Impact study of the Strengthened Accountability Programme, a programme for small-scale miners in Zambia. A case study as part of the Central Evaluation of Sida's work with poverty; NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20; NCG/Sida (Carsten Schwensen, Louise Smed) (2023), Midterm Evaluation of Women's Economic Empowerment (WEE).

structured component of programme design. Instead, adaptation tends to occur in an informal manner, shaped by context rather than by the systematic use of evidence. In fragile contexts such as South Sudan, programmes often demonstrate a high degree of operational flexibility, adjusting to changes in security, access, and population dynamics.

However, these adjustments are typically reactive, driven by immediate constraints rather than by structured analysis of monitoring data. Monitoring systems are not consistently designed to capture the effects of these adaptations or to generate insights that can inform future decision-making. As a result, learning remains largely implicit, residing in the experience of programme staff rather than being systematically documented or shared. This was mentioned in interviews with Sida embassy and HQ staff and also confirmed by the survey responses: 77% of staff respondents indicated that they relied “a lot” on their own experience while 70% of the respondents indicated relying “a lot” or “somewhat” on advice from national embassy staff.

In more stable contexts such as Zambia, the need for immediate operational adaptation is less pronounced, but this does not necessarily translate into stronger learning processes. Programmes may continue along established design trajectories even when emerging evidence suggests that modifications could improve effectiveness. The absence of formal mechanisms for reviewing and revising programme logic limits the extent to which monitoring data feeds into strategic decisions. This creates a form of path dependency, in which initial design choices continue to shape programme implementation even in the presence of new information. These patterns point to a broader structural issue that has also been highlighted in interviews with Sida staff: monitoring and evidence systems are not consistently designed to support iterative learning. Instead, they are oriented towards documenting pre-defined activities and outputs, with limited feedback loops into decision-making processes. As a result, even where data is available, it does not systematically inform adaptation.

Taken together, the challenges related to equity, cost-effectiveness, and adaptation reveal a common underlying feature. These dimensions are essential for understanding the quality and sustainability of impact, yet they remain peripheral to the core architecture of evidence systems. They are addressed in principle but not consistently embedded in the way programmes are designed, monitored, and evaluated. This has important implications for Sida’s overall effectiveness. Without robust evidence on distributional outcomes, it is difficult to demonstrate a clear contribution to poverty reduction and inclusion. Without systematic analysis of cost-effectiveness, it is not possible to ensure that resources are used in the most efficient way. Without structured approaches to adaptation, programmes are less able to respond to new information or changing conditions.

Addressing these gaps therefore requires more than incremental improvements in data collection. It requires a reorientation of evidence systems towards the key questions that underpin Sida’s mandate: who benefits, at what cost, and under what conditions do interventions lead to sustained change. This, in turn, implies stronger integration of these

dimensions into programme design, clearer alignment between indicators and objectives, and greater emphasis on using data to inform decisions rather than simply to report on activities.

The gap between evidence availability and evidence use becomes particularly evident in complex, multi-component programmes, where design involves balancing multiple objectives. In Zambia, agricultural initiatives such as the Women Economic Empowerment (WEE) Project (another Sida Evaluation of a project aimed to strengthen women’s agency, decision-making power, access to productive resources, and participation in agricultural value) and INCREASE illustrate how evidence is incorporated at a conceptual level but not operationalised in design choices. Both programmes combine a range of objectives, including improved productivity, climate resilience, market access, and gender empowerment.

However, the design does not clearly specify how these objectives interact within the Zambian context, nor does it prioritise specific pathways for achieving impact. In practice, this results in programme designs that are too broad in scope and diffuse in logic. Rather than selecting a limited number of evidence-supported pathways and investing deeply in their implementation, these programmes incorporated multiple intervention components with multiple partners simultaneously. This has diluted focus and stretches implementation capacity, while also complicating monitoring and evaluation.²⁹

In South Sudan, the use of evidence in design is shaped by a different set of constraints. Programmes operate in environments characterised by instability, weak institutions, and limited data availability. Under these conditions, design processes often rely on established models of service delivery or on international best practice frameworks, particularly in sectors such as health, social protection, and resilience-building. However, while these models provide a necessary operational foundation, they are not always sufficiently adapted to the specific conditions of the local context. For example, programmes focused on resilience or livelihoods often assume some degree of continuity in access to assets, markets, or services. In practice, however, these conditions are frequently disrupted by insecurity, displacement, or environmental shocks. While contextual analyses often identify these risks, the design of programmes does not always adjust objectives or intervention strategies accordingly. Instead, programmes may retain ambitious goals that are difficult to achieve given the volatility of the operating environment. The Budi case provides an example of this.³⁰

Across both Zambia and South Sudan, a broader pattern emerges in which programme design is influenced as much by generalised development narratives as by context-specific evidence. In South Sudan, Sida also often “buy in” to UN or multi-donor led programmes. These narratives - including market systems development, women’s economic empowerment,

²⁹ NCG/Sida (2022), Increasing Resilience in Energy and Agriculture Systems and Entrepreneurship (INCREASE). Sida Decentralised Evaluation 2023:20; NCG/Sida (Carsten Schwensen, Louise Smed) (2023), Midterm Evaluation of Women’s Economic Empowerment (WEE).

³⁰ Sida/NCG (2026), Impact study of the Budi Nexus Pilot, South Sudan A case study as part of the Central Evaluation of Sida’s work with Poverty.

climate-smart agriculture, resilience-building and addressing GBV - provide important strategic direction and coherence across Sida's portfolio. However, when applied without sufficient adaptation, they can lead to a form of "conceptual standardisation," where similar intervention models are replicated across different contexts without fully incorporating local evidence. When Sida takes part in a multi-donor set-up, they often also have limited influence on the design. In this dynamic, evidence tends to play a confirmatory rather than a critical role. Analytical work is often used to support or justify broad strategic directions, rather than to challenge underlying assumptions or refine programme choices. Furthermore, in South Sudan where independent research is more limited, there is limited independent analysis to challenge programme designs. This reduces the extent to which evidence functions as a constraint on design decisions and increases the likelihood that programmes will incorporate elements that are not fully supported by context-specific data.

The consequences of these dynamics are visible across the portfolio. Quite a lot of relevance analysis is being conducted (including MDPA) and programmes are generally well aligned with Sida's strategic priorities and demonstrate coherence at a conceptual level. However, they are not always grounded in a clearly articulated and evidence-tested understanding of "what works" and how change is expected to occur within specific contexts. This contributes to a pattern of over-ambitious designs, diffuse intervention logic, and challenges in both implementation and evaluation. While this does not indicate a failure to value evidence, it reflects a gap between the recognition of evidence as important and the existence of systems and processes that enable its consistent use. Thus, even when evidence is present it is not sufficiently embedded in the mechanisms that translate analysis into decisions.

5 Reasons for underuse of evidence in Sida's contribution management

The staff survey responses - together with interviews conducted with Sida staff - provide a critical lens through which to understand why evidence is not more systematically used in contribution management. It reveals that the underuse of evidence is not primarily a technical issue related to data availability or analytical capacity, but rather the result of a set of interlocking organisational, behavioural, and procedural dynamics. These dynamics shape how evidence is perceived, accessed, and ultimately applied in practice.

5.1 Delivery pressure and limited time and space for reflection

Interviews with Sida staff revealed that the contribution management system place considerable pressure on desk officers to keep contributions moving through the system and meet administrative system requirements. Even when evidence has pointed to weaknesses in a contribution's logic or feasibility, the costs - administrative, political, and temporal - of re-visiting the contribution design has often felt too high, leading to situations where evidence was acknowledged but applied only superficially. This is supported by survey responses, where 44% of the respondents reported feeling at least "somewhat" pressured when assessing partner proposals, while only 26% reported no pressure. Pressure levels were reported highest among HQ and Swedish embassy staff.

This relates to issues with the timing of evidence use within the programme cycle. The survey responses suggest that analytical work is often undertaken early in the design phase, when strategic directions are being defined. However, as programmes progress towards appraisal and approval, the influence of this analysis tends to diminish. Operational considerations - such as budget constraints, implementation arrangements, and alignment with broader strategies - become more prominent, and decisions increasingly shaped by practical requirements rather than by evidence. This creates a temporal disconnect between analysis and design where evidence informs initial concepts but is less influential in the stages where key decisions about programme structure and scope are finalised.

5.2 Limited incentives for pursuing and using impact evidence

Interviews with Sida staff pointed to some deeper constraints in the incentive structures that shape day-to-day contribution management within Sida. Even when staff recognised the value of robust impact evidence, the internal drivers were often found to reward other behaviours. Most contribution managers felt that they were primarily assessed on delivery, disbursement, and compliance with administrative requirements, not on the quality of evidence used to inform design or adaptation. Some also found that the system offered few incentives for taking evaluative risks - such as redesigning a contribution midstream, challenging a long-standing narrative, or commissioning deeper analytical work - because such actions would slow processes, trigger additional approvals, and complicate reporting. In practice, this meant that even when impact evidence was available, staff perceived limited organisational reward for

engaging with it, and in some cases even disincentives for doing so. Time constraints and risk for delaying approvals are mentioned as the most common barriers for using evidence with 39% and 21% respectively indicating “a lot” to constituting a barrier.

This incentive structure has important implications for how programmes evolve over time. It encourages continuity and stability in implementation, which is often necessary for operational reasons, but it can also discourage critical reassessment of programme assumptions. Even where evidence suggests that aspects of a programme could be improved or redesigned, there may be little practical incentive to undertake the additional work required to do so. In effect, the system tends to reward the successful delivery of planned activities more strongly than it rewards learning and adaptation.

5.3 Strong reliance on partner’s assessments, analysis and reporting

The evaluation team finds that Sida’s own assessments of partners in general are quite solidly founded on evidence and while the process has been criticised for being too cumbersome with humanitarian partners and sometimes overlapping with different departments assessing the same organisation,³¹ staff members largely also indicate that selection of partners is founded on solid evidence. This is also confirmed by a Sida assignment on evidence commissioned in July 2025 and again in November 2025.³² At the same time, interviewed Sida staff explained that they tended to rely quite strongly on partner analysis, reporting and internal monitoring because these sources were readily available and aligned with workflows reporting requirements, whereas external evidence requires time, analytical capacity, and institutional continuity that are not always present.

This is strongly confirmed by the staff survey: 63% of survey respondents indicated to rely “a lot” on partner assessments and 81% reported depending “a lot” on progress reporting from partners when they appraise a new project or do a mayor amendment. For comparison only 7% indicated to rely “a lot” on academic literature, 10% on Sida thematic/method briefs and 9% relied “a lot” on Sida Help Desks.

Some staff also expressed that they did not feel confident about how to access and make use of external evidence in practice. Even when evaluations, studies, or external datasets are available, the pathway from “having evidence” to “applying evidence” seems not always clear. The survey responses broadly support this view: while respondents generally perceived Sida as encouraging evidence use in partner selection processes, considerably fewer reported strong encouragement to use evidence for broader strategic reflection, adaptation, and redesign of contributions. This suggests that evidence use tends to remain more embedded in compliance-oriented processes than in iterative learning and adaptive management. Sida staff rarely use external evidence to assess effectiveness and efficiency, and this element is dedicated much

³¹ The national auditor (2025), Saving lives and alleviating suffering, Sida’s work with humanitarian aid, 2025: 33.

³² Sida (2025), internal document: Diskussionspapper: Användning av forskningsdata och oberoende evidens i Sidas interna processer, August 2025.

less attention in Sida's appraisal system.³³ Interviews further revealed that limited analytical capacity in some Sida teams compounded this challenge, making it difficult to identify and interpret large datasets even when these are accessible. This pattern reinforces the reliance on internal and partner-based knowledge. When the same systems that generate programme data also dominate the evidence used for decision-making, there is limited scope for independent validation or critical comparison.

5.4 Dependency on strong development narratives – lack of clarity for use of evidence

Staff explained that they feel Sida's strategic contribution priorities are in some cases framed around strong development narratives - such as systemic change, service delivery, women's economic empowerment and climate resilience - that provide coherence and direction but can overshadow focus on use of contextualised empirical findings. Once a narrative becomes embedded in a contribution, it tends to shape expectations, indicators, and partner relationships in ways that make it difficult for contradictory evidence to shift the trajectory. As a result of this narrative "lock-in", contributions may continue along their original path even when emerging evidence suggests that assumptions were unrealistic or that certain pathways are not delivering the expected results.

Similar findings were found in national audit reports conducted of both Swedish development aid (2024) and humanitarian aid (2025). They point to a lack of clarity and transparency on how evidence is applied in decision making and how previous results inform new decisions. Often decision making seems to be based on political priorities or other considerations than evidence of results.³⁴ This tends to reinforce established approaches and reduce the likelihood that programmes will be fundamentally reconfigured in response to new evidence.

A particular constraint seems to concern the operationalisation of evidence. Staff frequently reported uncertainty about how to translate evidence drawn from different settings into concrete design decisions. For example, findings from academic research or evaluations may offer general insights into what works but require significant judgement to determine how it applies to a particular country or sector. In the absence of clear frameworks or tools to guide this process, evidence is often used in a broad, illustrative manner rather than as a structured input into partner dialogue and decision-making. This reinforces the tendency for programme design to be shaped by general principles or narratives (see above) rather than by detailed empirical analysis.

³³ Sida (2025), internal document: Diskussionspapper: Användning av forskningsdata och oberoende evidens i Sidas interna processer, August 2025.

³⁴ The national auditor (2024), Sweden's International Development Aid – monitoring, evaluation and reporting, 2024:4; The national auditor (2025), Saving lives and alleviating suffering, Sida's work with humanitarian aid, 2025: 33.

5.5 Limited or moderate use of thematic networks, help desks and system tools

Thematic and methods briefs produced by Sida HQ provide examples of good practices, and guidance on what works within a thematic area or a specific method. While these internal guidance documents provide useful evidence on specific topics, many staff found that they offered less concrete support on how to translate evidence into design choices, partner dialogue, and adaptive management. It is however noted that recent briefs have provided clear guidance on partner dialogue which has previously been a challenge.³⁵

Experiences from a recent assessment of evidence showed that few HQ thematic and methods briefs are referenced in appraisals.³⁶ This pattern is also reflected in the staff survey responses, where the use of Sida's thematic and methods briefs seem to vary substantially across staff groups: While 39% of National Programme Officers (NPOs)/ National Programme Assistants (NPAs) respondents reported using the briefs "a lot", this was only the case for 25% of embassy staff and 13% of HQ staff. Nearly half (48%) of HQ respondents reported not using the briefs at all. This suggests that the briefs may not always be sufficiently operational, accessible, and adapted to day-to-day contribution management realities. Interviews with HQ and Swedish embassy staff further revealed that consulting national programme officers with long experience was often considered more relevant than HQ documents.

Help desks are also established within strategic topics such as gender, environment, human rights, conflict, peace building etc. to support both Sida staff but also partners in considering mainstreaming of strategic perspectives and priority areas. Staff members can request support from help desks to e.g. conduct an analysis of gender or conflict sensitivity for a contribution. According to the Sida staff survey 37% found the Help Desks useful to guide the use of evidence for contribution management and Help Desks are highest rated for systems that support evidence generation.

The contribution management system Tool for Results management and Appraisal of Contributions (TRAC) has long been criticised for not being suitable for project follow-up and for being too cumbersome³⁷ and several attempts to update it have taken place. It is however recognised for requesting an analysis of "Poor People's Perspective", and the manual for Sida's contribution management system (TRAC) specifically instructs programme officers to consider: "the link between the MDPA and the specific contribution and refer to it in the assessment" which has been found to be a clear indication of institutionalisation of MDPA. On the other hand, there are no references in the interface where information is recorded, nor are there any specific questions related to the MDPA model. As the manual is a 300-page document, and help functions are not obligatory to access, instructions regarding the MDPA

³⁵ E.g. Sida (2024), Preventing and responding to gender-based violence, Thematic Support Unit.

³⁶ Sida (2025), internal document: Diskussionspapper: Användning av forskningsdata och oberoende evidens i Sidas interna processer, August 2025.

³⁷ SIPU (2016), Evaluation of the effectiveness of Sida's work with Contribution management.

may be easy to miss.³⁸ In the last quarter of 2025, TRAC was being updated to comply with changes in the Contribution Rules from July 2025.³⁹ According to interviews, a decision has been taken to replace TRAC fully by 2027. Help texts in TRAC was however considered important to guide the use of evidence with 30% of survey respondents indicating “a lot”. 18% and 44% respectively however indicated not at all or somewhat supportive

Thematic networks are also meant to share newest evidence and best practice examples across geographic areas, however according to survey responses 23% use these networks to guide the use of evidence in contribution management. While structures are present in Sida to foster knowledge sharing, their usefulness depends very much on the persons in the networks, and it is less clear to which extent material shared through these networks is being applied in contribution management.⁴⁰

5.6 Lack of evaluability requirements in contribution design

A further explanation provided is that Sida’s contribution management do not require use of evidence nor evaluability at the design stage. There is no mentioning of evidence in the Contribution Rule and while it has been encouraged previously it has not been a requirement, so it has largely rested on the individual programme officer to decide to what extent external evidence has been mentioned.⁴¹ This view is also reflected in survey responses, where only 40% of respondents found that Sida encourages the use of evidence “a lot” when considering future contributions. Contributions can be approved without baseline reports, counterfactual logic, gender-disaggregated indicators, impact targets or system diagnostics and more than 50% of staff respondent indicated that base- and endline reports were not relevant for their contribution. While this could be the case, it is also possible that at least in some cases it could have been relevant but the desk officer may have lacked the skill set to assess this. When these elements are missing at the outset, they are rarely added later, leaving programmes without the data needed to assess impact or adjust course. Even when staff are committed to evidence-driven decision-making, they may simply not have the information required to act on that commitment.

Interviews with Sida staff further indicated that measurement of impact will often be pushed from one phase to another in the recognition that it takes time to achieve. At the same time, survey responses revealed that more than 80% of respondents considered impact targets relevant in contribution design and appraisal. This suggests that the challenge may not be a lack of recognition of the importance, but rather the extent to which impact targets are consistently operationalised in practice, particularly in smaller contributions where impact is perceived to be beyond the direct sphere of influence.

³⁸ EBA (2025), Evaluation of Sida’s Application of Multidimensional Poverty Analyses.

³⁹ Sida (2025), Rule for contribution management, 1st July 2025.

⁴⁰ The national auditor (2024), Sweden’s International Development Aid – monitoring, evaluation and reporting, 2024:4.

⁴¹ Sida (2025), Rule for contribution management, 1st July 2025.

5.7 Evidence constraints in fragile and data-poor contexts

The fact that Sida does not have international staff permanently on the ground - or that staff change frequently - in fragile contexts (such as in South Sudan) was found to create other challenges and less systematic follow-up on contributions.⁴² In these situations, desk officers must rely heavier on remote communication, periodic missions, and partner-generated documentation, which limits their ability to interrogate data quality or contextualise reported results.

Uneven staff on the ground also reduces Sida's ability to build informal knowledge networks, maintain continuity across programme cycles, and detect discrepancies between reported progress and contextual dynamics. In practice, this creates a situation where Sida may have no alternative but to use weak, unverified, or anecdotal evidence simply because it is the only information available within the operating context. With uneven staff on the ground to validate findings and sense-check assumptions, Sida's ability to critically assess evidence is also further constrained. Still, in very fragile or conflict-affected contexts, even though Sida has an in-country presence, access to project areas may still be limited.

⁴² Also a finding in the national auditor (2025), *Saving lives and alleviating suffering, Sida's work with humanitarian aid*, 2025: 33.

6 International experiences on evaluation of development impact⁴³

Over the past several decades, impact evaluation in development has undergone a profound transformation. What began in the 1970s as a relatively narrow, academic exercise gradually evolved into a central component of evidence-based policymaking. Governments have moved from focusing primarily on how much money was spent to asking what results were actually achieved. This shift, accelerated in the 1990s “results agenda” and the subsequent “evidence revolution,” led to the widespread adoption of rigorous causal methods - most notably randomised controlled trials (RCTs). Over time, evaluation became institutionalised within governments, supported by a growing ecosystem of funders, intermediaries, and knowledge brokers, and complemented by new tools such as systematic reviews and evidence gap maps.

While the technical quality of evaluation has improved, its practical usefulness has not increased at the same pace, largely because evaluations are often disconnected from the realities of programme design and implementation. A major reason for this gap is that development programmes themselves have changed. Many are no longer discrete, well-bounded interventions with clearly defined beneficiaries and timelines. Instead, they are complex, adaptive systems involving multiple actors, overlapping components, and evolving objectives. In such contexts, the idea that one can apply a single, clean causal design *ex post* becomes unrealistic. Instead, a more flexible understanding of evaluation as a “portfolio of approaches” may be considered, each suited to different types of interventions and constraints.

This leads to one of the central insights: the main bottleneck in impact evaluation is not the lack of sophisticated methods, but the lack of “design-readiness.” Credible evaluations depend on three foundational elements established early on: a clear and explicit ToC, baseline or pre-intervention data, clear identification of who receives what intervention and when, and data systems that allow outcomes to be tracked and linked. When these elements are missing, evaluators must reconstruct them retrospectively - often relying on incomplete records, recall data, or indirect proxies. The result is not simply weaker evidence, but evidence that is structurally limited in what it can claim, regardless of the sophistication of the methods applied.

6.1 A practical menu of options

Repeated quasi-experimental designs (difference-in-differences (DiD) type designs) are among the most useful “middle” options when the programme has a clear implementation date and repeated outcome data exist before and after implementation, ideally for both treated and untreated or later-treated groups. Their marginal cost can be relatively low when they rely on existing monitoring or administrative data, and they are often much simpler operationally than fielding a new survey-based RCT. Their main bottlenecks are weak or noisy routine data,

⁴³ Extracted from Rand J. (2026), How Can Evaluation Commissioners Think About Impact Assessments Going Forward? This paper was developed as a specific input to this evaluation, see full version in Annex 4.

unclear intervention timing, insufficient pre-intervention observations, and difficulty finding a comparison group unaffected by spillovers.

Threshold, rollout, and allocation-rule-based designs (regression discontinuity design (RDD), phased rollout comparisons, synthetic control methods) are often the strongest non-randomized options when programme rules or implementation realities already create a comparison logic. RDD is especially credible when eligibility is determined by a threshold and there are enough observations close to the cutoff. Synthetic control is particularly useful for area-level interventions when rich pre-intervention secondary data exist. Matching approaches can also be valuable, but only when one has rich covariate data on participants and nonparticipants; otherwise, unobserved selection remains a serious threat. These designs are simpler than an RCT in operational terms, but they are not “cheap tricks”: they demand clean assignment rules, good documentation, a lot of data and competent analytical skills.

*Repeated baseline-endline assessments combined with process evaluation*⁴⁴ are a useful alternative when a programme already has or can still collect pre-intervention and post-intervention data but cannot construct a convincing comparison group. The literature is explicit that such designs should not be oversold as definitive causal attribution. Their value comes from improving on raw stakeholder perception by documenting change over time, implementation reliability, reach, exclusion, and identification of plausible mechanisms. They become much more credible when integrated with ToC testing, triangulation, and monitoring-system data. In some portfolios, these designs are the most realistic option short of doing no impact-focused work at all.

Theory-based and contribution-oriented approaches become relevant when interventions are complex, context-dependent, politically embedded, or difficult to evaluate using experimental or quasi-experimental designs. Rather than estimating a precise net effect, these approaches seek to develop causal claims about whether, how, and under what conditions an intervention contributed to observed change. Their strength comes from explicit ToCs, systematic testing of causal mechanisms, triangulation across data sources, and active consideration of alternative explanations. This family of approaches includes contribution analysis, process tracing, comparative case studies, and participatory methods. Although these differ in emphasis and analytic strategy, they are often used in combination and are particularly useful when programmes are heterogeneous, sample sizes are small, outcomes emerge through multiple institutions or pathways, or context is central to explanation. While these approaches typically cost less and do not require a conventional control-group architecture, they are not “light-touch” alternatives. Credible application requires careful case selection, transparent analytic procedures, systematic verification of claims, and teams capable of combining qualitative and quantitative reasoning.

⁴⁴ Process evaluation refers to assessing how and why a programme was implemented and functioned, alongside descriptive baseline-endline outcome assessment in the absence of a comparison group.

Finally, *evidence synthesis products* (systematic reviews and evidence gap maps) are not substitutes for a programme-level impact evaluation, but they are a critical part of the options set because they determine whether a new evaluation is worth doing, where it will add the most value, and which design should be favoured. Systematic reviews tell commissioners what is already known and evidence gap maps show where strong evidence clusters already exist and where strategically important gaps remain. These products are increasingly central to research prioritisation, funding choices, and evidence-informed decision-making.

6.2 Cost estimates and requirements

The most important point on costs is that a method's price is driven less by the label attached to it than by whether the design can piggyback on data and programme features that already exist. The cheapest credible options are usually those that exploit existing administrative and monitoring data, clear treatment flags, and observable rollouts. Costs rise sharply when evaluators need to create those foundations from scratch through surveys, retrospective reconstruction, or prolonged access negotiations.

Administrative data are especially attractive because they can be less expensive and logistically easier than collecting new primary data, impose less burden on respondents, and often support long-run follow-up. But the hidden costs can be significant. Data-use agreements for identifiable data typically involve substantial review; gaining access should begin in the design phase; providers may take months to extract and transfer files; and iteration over matching rules and file definitions is common. Even in relatively well-institutionalised settings, administrative files can carry nontrivial monetary fees.

Theory- and case-based approaches tend to shift cost from surveys and sample logistics toward senior analytical time, theory development, fieldwork quality, and synthesis discipline. The literature repeatedly describes these methods as specialist, time-consuming, and, if done well, resource intensive. That does not make them unattractive. It means they must be commissioned with the right skill mix and realistic timelines rather than treated as a cheap qualitative add-on. Their comparative advantage is not low price; it is that they remain feasible and policy-relevant where true counterfactual designs are blocked by ethics, politics, small samples, or complex causal chains.

Process evaluation is often the highest-return complement in this middle space. Process evaluation strengthens impact evaluation by revealing whether weak results reflect design failure, implementation failure, institutional constraints, or external shocks. It can also help explain effects, identify excluded groups, and feed back into programme improvement. In practice, the bottleneck is usually not conceptual acceptance but under-budgeting: commissioners want causal answers but do not allocate enough time or resources to examine mechanisms.

Preparatory products are often underused because they are perceived as non-essential overhead, yet the literature suggests the opposite. Evaluability assessments usually represent a small share of total evaluation costs and can pay for themselves by preventing premature,

underspecified, or data-impossible studies. Systematic reviews take time and evidence gap maps are also resource-intensive and vulnerable to coding subjectivity, but they are increasingly supported by automation and better software. In other words, these should not be optional in future impact assessments as they are the cost-effective architecture around later impact work.

6.3 Binding constraints and possible ways forward

The *first* binding constraint is late commissioning of evaluations. When the evaluation is designed only after programme implementation is underway, the programme usually lacks baseline measures, pre-randomisation identifiers, treatment flags that can be linked to external data, and sometimes even a stable statement of who exactly the intervention was meant to reach. The consequence is weaker retrospective designs, greater ambiguity about whether groups were comparable, and limited ability to explain results. Solutions are prospective design at programme inception, baseline or intake data collection, early treatment/comparison logic, and collection of identifiers before exposure so later linkage does not itself introduce bias.

The *second* binding constraint is weak ToC and indicator architecture. Many programmes have output indicators but lack intermediate, process, and mechanism indicators, or they use log frames that list activities and outputs without explaining how change is expected to happen. The result is an evaluation that can count activity but cannot diagnose where the causal chain broke. The literature is explicit that ToC should guide data collection, analysis, and reporting, and that each element of intervention logic should be associated with indicators and data collection or verification methods. The practical fix is to map each causal step in advance and assign a preferred data source (M&E, administrative records, surveys, qualitative inquiry, and/or digital trace) to each step.

The *third* binding constraint is fragmented or inaccessible data. Many donor-funded programmes could support stronger evaluation if their M&E records were linked to national administrative databases, but in practice those datasets often sit in silos, use different identifiers, operate under unclear legal permissions, or are held by agencies that lack the staff time to support extraction and matching. The route around this bottleneck would include data audits before commissioning, memoranda of understanding or data-use agreements, clear data-flow diagrams, agreed extraction schedules, realistic timelines, and willingness to support national data owners with staff time, in-kind analytical help, or other cooperation incentives. In some cases, on-site linkage at the data owner's premises is the most feasible compromise between access and confidentiality.

The *fourth* binding constraint is lack of a credible comparator. Programmes are often national, politically sensitive, and already rolled out everywhere. In those cases, the best response is not to default to generic qualitative opinion gathering. It is to look systematically for alternative comparison logic; rollout phasing, delays due to administrative factors, eligibility scores, geographic thresholds, synthetic controls, or natural experiments. If none exist, the evaluation should pivot honestly toward theory-based contribution analysis, process tracing, and comparative case work rather than pretending that net impact has been estimated. The

constraint is not simply “no control group”; it is failure to adapt the research question to the level of causal leverage that is actually feasible.

The *fifth* binding constraint is capability. Administrative-data evaluation requires people who understand identifiers, linkage, data quality, privacy, extraction logic, and statistical assumptions. Theory-based work requires people who can specify a credible ToC, formulate rival explanations, and synthesise mixed evidence. New digital methods require computational literacy and ethical judgment. The literature on government analytics and AI is clear that both data infrastructure and human capital matter. Without high-quality foundational data and the capacity to process, analyse, visualise, and interpret them, “analytics” becomes unreliable or performative.

These constraints reinforce each other, creating a cycle where evaluations are expected to deliver strong causal claims without the conditions required to support them.

6.4 Final remarks

Individual evaluations are only one part of a broader evidence system. Systematic reviews, meta-analyses, and evidence gap maps play an increasingly central role in shaping policy because they aggregate findings across contexts and identify where evidence is strong or weak. This marks a shift from a model where each evaluation is expected to provide definitive answers, to one where knowledge accumulates across studies. As a result, commissioning a new evaluation without first understanding the existing evidence base is both inefficient and potentially redundant.

The fundamental problem in development impact evaluation is not choosing the wrong methods but treating evaluation as an end-stage activity rather than an integral part of programme design. The most effective reform is therefore to move evaluation upstream. Before commissioning new impact assessments, agencies should review existing evidence, assess evaluability, map available data, secure data access, and align indicators with the programme’s ToC.

Ultimately, the goal is not methodological perfection, but proportionality - matching evaluation ambition to programme scale, complexity, and decision-making needs. By embedding evaluation early, strengthening data systems, and prioritizing synthesis and practical design choices, impact assessment can better fulfil its purpose: generating insights that inform policy and improve development outcomes.

7 Conclusions

Below are first presented some overall conclusions based on the evaluation findings presented in the report (section 7.1). This is followed by some more specific conclusions related to identified shortcomings in Sida's work with evidence in contribution management (section 7.2).

7.1 Overall conclusions

Sida's portfolio demonstrates a clear capacity to deliver development results across a wide range of sectors and contexts. From measurable improvements in income, food security, and service access in Zambia to the delivery of essential health and protection services in South Sudan, Sida-supported interventions contribute in important and tangible ways to improving living conditions for poor people. These achievements reflect both the relevance of Sida's strategic priorities and the effectiveness of its implementing partnerships. Yet, the evaluation also shows that the depth, credibility, and interpretability of these results vary significantly across the portfolio. Understanding why this variation occurs is central to strengthening Sida's overall contribution to impact.

The evaluation finds that the main challenge in Sida's contribution management is not a lack of information but rather a limited extent to which existing evidence is integrated into decision-making. While evidence from evaluations, research, partner data, and national systems is often available, this information does not consistently shape programme design, indicator selection, and adaptive management. This disconnect is not due to one single factor but arises from the interaction of design practices, monitoring systems, organisational incentives, and contextual constraints.

The evaluation also shows that timing is a critical factor. Many limitations in evaluability and impact assessment originate in the design phase, where key assumptions, causal pathways, and indicators are established. Contribution designs often do not include explicit causal logic or testable assumptions. As a result, evaluability is often weak from the outset. When ToCs are broad, implicit, or overly ambitious, this limits the ability to assess effectiveness and understand how change occurs. The evaluation shows that the quality of evidence generated during implementation is largely determined at inception and that design weaknesses tend to cascade through the entire programme cycle.

Monitoring systems reinforce this pattern. These systems are generally robust in documenting outputs but far less effective in capturing outcomes, behavioural change, or system-level effects. Heavy reliance on partner-reported data, limited triangulation with independent sources, and weak alignment between indicators and programme logic reduce the analytical value of monitoring information. Lack of consistent and disaggregated baseline data further challenge this. In fragile contexts (like South Sudan) the challenges are amplified by restricted access and weak national data systems; in data-rich contexts (like Zambia), opportunities for triangulation are not consistently utilised and national data is not always easy or cheap to

access. The result is a verification gap that constrains the ability to substantiate claims about outcomes and impact.

Organisational incentives further shape how evidence is used. Sida's contribution management processes tend to prioritise delivery, compliance, and disbursement over analytical reflection or evidence-based adaptation. Staff often face compressed timelines, competing demands, and limited space for structured review. In this environment, evidence tends to play a supporting rather than a guiding role. While the evaluation findings show that adaptation does occur in contribution management, it is mostly driven by operational experience rather than in combination with systematic analysis of data. An insufficient attention to cost-effectiveness issues often constrains the ability to prioritise among competing approaches, limiting Sida's ability to ensure that resources are both efficiently allocated and grounded in context-specific evidence.

Equally important is the usability of evidence. Evidence is often produced in formats - such as lengthy reports, technical studies, or global research syntheses - which do not directly correspond to the kinds of decisions programme managers need to make. Translating such material into actionable insights requires both time and specialised interpretation capacity. In the absence of practical tools, guidance, or summarised formats, staff are more likely to rely on immediately accessible sources such as partner documentation, internal experience and operational considerations. As a result, even strong evidence may remain underutilised, while more accessible but less robust information often tends to play a larger role.

Taken together, these findings show that evidence is not yet functioning as a central foundation to shape intervention logic within Sida's contribution management. Strengthening Sida's contribution to development impact therefore requires a shift in how evidence is being conceptualised and used. Evidence must actively shape what programmes aim to achieve, how they are structured, how progress is measured, and how decisions are made throughout the programme cycle.

Ultimately, strengthening Sida's contribution to poverty reduction depends not only on what programmes do, but also on how they learn from what they do. Embedding evidence more systematically within contribution management is therefore essential for improving both effectiveness and credibility. From an evidence perspective, the key issue is not the absence of relevant knowledge, but the absence of structured processes to translate that knowledge into strategic choices about what to do, where to do it (country/location), and how to sequence interventions over time. Such shift will require changes in design practices, monitoring systems, organisational incentives, and the practical tools available to staff.

Thus, the findings across the evaluation point to a clear and consistent conclusion: **the primary challenge facing Sida is not the production of additional evidence, but a more effective integration of existing evidence into contribution management processes.**

7.2 Specific conclusions on shortcomings in Sida's work with evidence

Below, more specific conclusions are presented with regards to identified shortcomings in Sida's work with evidence in contribution management. The conclusions are based on the findings from the eight impact case studies, the synthesis of wider evidence from Sida evaluations and the Sida staff survey.

- **Strong reliance on partner-generated evidence with limited verification.** Sida's contributions rely heavily on partner assessments, data, and reporting, with limited systematic cross-checking. This creates risks related to bias and weakens confidence in reported results. In South Sudan, reliance on partner reporting was particularly strong. This was, for instance, the case in the Budi Nexus Pilot, where monitoring depended almost entirely on partner reporting. The evaluation team's assessment of the Budi Nexus Pilot also illustrated what triangulation of partner reporting could do, as potential discrepancies were found when checked with external conflict data.
- **Baseline data (particularly micro-level data) is often inadequate, misaligned, or inaccessible.** Baseline data frequently suffers from conceptual and practical limitations. In many cases, available data are aggregate indicators, which are insufficient for assessing programme-level change. Relevant micro-level, disaggregated data aligned with target populations and outcomes is often inaccessible or unavailable (in raw format), incomplete, and/or misaligned with outcomes the interventions supported tried to influence. This significantly constrains the ability to measure change over time.
- **Access barriers limit the use of existing national administrative/statistic data.** Even when relevant datasets exist, access constraints reduce their utility. National datasets (such as LCMS and RALS in Zambia) were difficult to obtain access to for the evaluation team, and their use often depended on pre-existing connections. Evidence use is thus constrained not only by availability but also by institutional access conditions which may in some cases be outside Sida's control.
- **Monitoring systems are predominantly compliance-driven and insufficiently designed to support learning and equity.** While monitoring systems are generally effective for accountability and financial oversight, they tend to prioritise reporting requirements over generating actionable insights for decision-making and adaptive management. Often indicators are not clearly aligned with ToCs, and do not integrate multiple data sources to support verification. At the same time, limited integration of distributional perspectives means they often fail to capture differences in outcomes across key population groups (e.g. with regards to gender, age, disability, poverty status, location etc.). Together, these weaknesses constrain the ability of monitoring systems to inform programme improvement and ensure inclusive, evidence-based decision-making.

- **Programme design is frequently driven by compelling narratives rather than robust, context-specific evidence, resulting in overly ambitious interventions with weakly substantiated logic.** A broader pattern emerging from all Sida reviewed evaluations is the dominance of narrative over evidence in contribution choices. Many contributions have been built around compelling development narratives - systemic change, market system development, women’s economic empowerment, productive use of energy, financial inclusion, climate-smart agriculture, organic certification, or GBV prevention. While these narratives often carried strong intuitive appeal and are aligned with Sida’s strategic priorities, they sometimes overshadowed contradictory evidence or gaps in the underlying intervention logic. For instance, the Zambia cases illustrated examples where good evidence often existed but had not been clearly reflected in contribution choices.
- **Limited integration of cost-effectiveness into programme analysis limits Sida’s ability to assess scalability and sustainability.** Without systematically linking costs to results, it is difficult to identify which components deliver the greatest impact, particularly in complex programmes. This weakens both programme refinement and portfolio prioritisation, ultimately constraining the translation of programme success into wider systemic impact.

The consequences of these shortcomings are visible across the Sida portfolio: Programmes generally do quite a lot of relevance analysis (linked to the MDPA), are well aligned with Sida’s strategic priorities and demonstrate coherence at a conceptual level. However, they are not always grounded in a clearly articulated and evidence-tested understanding of “what works” and how change is expected to occur within specific contexts. This contributes to a pattern of over-ambitious designs, unclear intervention logics, and challenges in both implementation and evaluation. While this does not indicate a failure to value evidence, it reflects a gap between the recognition of evidence as important and the existence of systems and processes that enable its consistent use. Thus, even when evidence is present, often it is not sufficiently embedded in the mechanisms that translate analysis into decisions.

8 Recommendations

Below are first presented (section 8.1) some overall recommendations based on the conclusions presented in Chapter 7. This is followed by some more specific and actionable recommendations to Sida management (section 8.2).

8.1 Overall recommendations

The findings and conclusions from this evaluation point to the need for a systematic strengthening of how evidence is embedded within Sida's contribution management. Addressing the challenges identified across design, monitoring, and organisational practice requires a set of mutually reinforcing changes. These changes do not primarily involve generating more data but rather ensuring that existing evidence is effectively translated into decisions that shape contribution design, implementation, and adaptation.

The overall recommendations that follow below outline how Sida can move toward a more integrated, evidence-driven approach that enhances its ability not only to deliver results, but to understand, verify, and strengthen them over time, in a cost-effective manner.

First, contribution design practices should become more explicitly evidence-driven with enhanced focus on evaluability. The evaluation findings show that the quality of evidence generated is largely determined at inception. This implies that evaluability should not be treated as a secondary or technical consideration, but rather as a core design requirement. Approval processes should systematically require programmes to articulate causal pathways, define key assumptions, and specify how these assumptions will be tested through data. While not all contributions have to generate new data or conduct rigorous evaluations, the need for data/information/evaluation should always be assessed early.

In practical terms, this also implies strengthening the link between ToC and data systems already at design stage. For each step in the causal chain, programmes should identify what indicators will be tracked, from which data sources, at what frequency, and whether these data can be linked to existing administrative or external datasets. At the same time, this should be approached pragmatically, particularly for programmes that are already underway or where evaluation needs emerge retrospectively, as well as in fragile and conflict affected contexts where adaptation and flexibility should be possible, recognising that programme logics can be affected by contextual changes. In such cases, the focus should be on improving evaluability ex-post through clearer reconstruction of programme logic, targeted data mapping, and prioritisation of feasible causal questions.

Closely linked to this is the need to align monitoring systems with programme logic and intended outcomes. This will require a shift towards better capturing of outcomes, behavioural changes, and system-level effects. This involves selecting indicators that directly reflect the causal pathways outlined in programme design and ensuring that these indicators are measurable within the constraints of the operating context. In addition, monitoring systems

should be explicitly designed to support both evaluation and adaptive management, enabling programmes to adjust in response to emerging evidence.

Second, there is need for strengthening of evidence through triangulation. The current heavy reliance on partner-reported data limits both verification and analytical depth. While partner systems will remain central to data collection, they should be complemented more systematically by independent and external sources of evidence, including national statistics, third-party monitoring, administrative data, and existing research. Integrating multiple sources of data - although not always easy in fragile and conflict affected countries like South Sudan - would not only improve credibility but also provide a more comprehensive understanding of programme effects.

As part of this, improving the usability of evidence for staff is also critical (staff indicate that one of the main barriers to evidence use is uncertainty about how to translate available knowledge/evidence into practical decisions and make it “actionable”). This will require development of practical tools, guidance, and processes that support the application of evidence in design and management. This could include structured templates for linking evidence to programme logic, summaries of relevant research tailored to specific sectors or contexts, and examples of how evidence has been applied in comparable programmes. It also suggests a stronger role for synthesis products - such as evidence reviews or gap maps - to help staff understand what is already known and where additional evidence is most needed.

Third - at the organisational level - there is need for better aligning incentives with evidence-informed decision-making. This should include introduction of clearer expectations around the use of evidence in contribution design and adaptation. This does not imply reducing the importance of delivery but rather complementing it with a stronger emphasis on effectiveness, efficiency and learning. This also points to the importance of creating more space for reflection and adaptation within programme cycles. Current processes seem often to compress design and implementation timelines in ways that limit opportunities for analysis and revision. Introducing structured points for review - where monitoring data and emerging evidence are explicitly considered - would support more systematic learning.

Strengthening Sida’s effectiveness and efficiency in delivering results would also require a more explicit integration of cost-effectiveness considerations alongside a more critical use of narratives in programme design. This would imply a more systematic assessment of the relationship between costs and results across intervention components and using this analysis to inform prioritisation, scaling, and adaptation decisions. At the same time, narratives should be reframed as hypotheses to be tested rather than assumptions to be confirmed, ensuring that programme choices are continuously challenged and refined through evidence. Embedding these practices would support more realistic, focused, and efficient interventions, and enhance Sida’s ability to translate resources into sustained and verifiable development impact.

Strengthening evidence use requires a shift in how evidence is conceptualised within Sida's overall approach. Rather than being treated as a supporting input or a compliance requirement, evidence needs to function as a central organising principle within contribution management. This means that evidence should actively shape what programmes do, how they do it, and how success is assessed. It should define priorities and provide the basis for adaptation over time.

8.2 Specific and actionable recommendations

Building on the overall recommendations presented above in section 8.1, below is presented more specific and actionable recommendations for Sida management to consider for enhancement of evidence use and evaluability in contribution management, with a view to cost-effectiveness and efficiency:

Specific Recommendation 1: Contributions should always be designed as evaluation-ready as possible from inception, with clear causal logic and aligned data systems (requirements to be differentiated in view of partner capacity and budget volume).

Proposed actions for contribution design:

- **A ToC with clearly defined causal links** and explicit, testable assumptions.
- **Map each step of the ToC** to concrete indicators and data source (a ToC-to-data mapping table specifying indicators, sources, frequency, and verification approach).
- **Baseline data disaggregated** – and micro-level data if feasible - in accordance with specific target groups/locations and, if possible, including a comparison group (when relevant, since some contributions are not directed towards a specific target group, but rather whole populations, e.g. support to election processes, strengthening of institutions, research, support to a national development plans, etc.). Continued access to raw data and key identifiers is important in this regard.
- **Link to national administrative data/statistics or external data sources (if feasible).** Explore possibilities for **piggybacking** on existing data systems and make sure that baseline surveys will ask **same questions** as in national surveys (to allow for comparison across waves). Small upstream investments in comparable measures and linkable data architectures often determine whether later evaluation can rely on existing systems or must resort to more expensive retrospective reconstruction.
- Definition and inclusion of outcome- and mechanism-focused monitoring indicators (e.g. related to behavioural changes and system-level effects) ensuring that **monitoring plans/systems** will be set up to capture defined outcomes (and not only activities and outputs).
- Identification of any possible **data gaps/challenges**.

Proposed actions for choice of evaluation design/approach:

- Evaluation planning should be informed by existing evidence through **systematic reviews and gap maps**. This should be followed by a recommendation on what information would be needed to determine the difference a given contribution has made. This will ensure cost-effectiveness and value of new studies.

- Where good-quality time series data and a credible comparison group exist, **quasi-experimental designs** such as difference-in-differences (D-i-D) should be the default choice due to their relatively low cost and strong causal potential. This requires early investment in robust routine data systems and clear documentation of intervention timing. When comparison groups are not feasible, **baseline-endline designs** can still add value if combined with process evaluation, ToC analysis, and triangulation - though they must be presented transparently as non-causal.
- For complex or system-level interventions, **theory-based approaches** should be pursued, focusing on understanding contribution and causal pathways rather than estimating precise effects.
- Across all approaches, **triangulation of multiple data sources and methods** should always be ensured to strengthen credibility.
- It should be considered to commission (selected) **ex-post (impact) evaluations**.⁴⁵ Usually, there is no funding for ex-post impact assessment within the programme budget, as this will generally not be done during the programme period. Impact assessments should usually not be made at the time of programme closure but preferably some time after.
- Overall, a **fit-for-purpose, realistic, and mixed-methods approach** should always be preferred to over-ambitious designs that cannot be implemented rigorously.

What will this change?:

- It will shift evaluability from a technical afterthought to a binding design condition.
- It will ensure that monitoring systems become better aligned with causal logic and in support of adaptive management rather than being based mainly on reporting needs.
- It will enable more credible and cost-effective evaluations by addressing data constraints early rather than retrospectively.
- It will help enhance efficiency, effectiveness and learning from implementation.

Specific Recommendation 2: Strengthen credibility and usability of evidence through triangulation and implementation of practical tools to translate evidence into decisions.

Proposed actions:

- Develop and roll out practical “**evidence-to-action**” toolkits. Create standardised, user-friendly tools - such as templates linking evidence to programme logic, decision-making guides, and step-by-step workflows - that help staff translate evidence into design and management choices. Pilot these toolkits in selected countries/teams and refine based on feedback before scaling up.

⁴⁵ The Swiss Development Cooperation (SDC) is currently piloting an ex-post evaluation framework, where some projects are randomly selected for ex-post (impact) assessment 4-6 years after their completion.

- Develop “**rapid evidence scan guides**” (short, step-by-step guidance) on how to quickly identify and synthesise relevant evidence to inform emerging decisions without requiring full-scale research (e.g. focus on applied evidence use from comparable programmes/interventions).
- Develop “**evidence quality checklists**” (simple tools) to help staff quickly assess the reliability and relevance of different data sources (e.g. partner reports vs external studies). This will support more critical, consistent use of evidence in decision-making.
- Establish a **centralised evidence synthesis function** to support producing synthesis products such as evidence reviews and gap maps. Ensure these outputs are easily accessible and actively disseminated and link them to programme planning processes so staff can systematically use them when identifying priorities and designing interventions.

What will this change?:

- It will facilitate a move from reliance on single-source, partner-driven evidence to verified, multi-source evidence.
- It will improve both credibility (through triangulation) and usability (through practical tools).
- It will enable staff to more consistently translate evidence into concrete design and management decisions.
- It will facilitate that contribution designs to a larger extent become grounded in existing evidence than in compelling narratives.

Specific Recommendation 3: Enhance organisational incentives and processes to better align with evidence-based and cost-effective decision-making.

Proposed actions:

- Review - and possibly adjust - the current **balance between system requirements** (and staff time allocation) for, respectively, i) relevance and coherence analysis; and ii) “what works” (effectiveness, efficiency, learning/adaptation), with a view to enable larger focus and staff dedication to evidence use for “what works”. A possible closer linking between the two parts (building further on already well-established processes related to the MDPA) should also be explored.
- **Set mandatory evidence-use standards in contribution design.** Require all new contributions and major revisions to explicitly document the evidence informing design choices and how evidence will be used for adaptation. Introduce a short compliance template and make this a formal approval criterion. Important, however, that “evidence” will not just become another “compliance check box” but that focus will be on quality aspects and action-orientation.
- **Align incentives and performance management** with learning and adaptation. Adjust staff and management reviews to include measurable expectations on evidence use (e.g., quality of evidence integration, adaptation decisions taken, lessons documented and applied).

What will this change?:

- It will move incentives from delivery and compliance towards effectiveness, efficiency, and learning.
- It will ensure that evidence becomes more actively used in decision-making (and not primarily in reporting).
- It will strengthen Sida's ability to prioritise resources and adapt programming based on results and wider experiences.
- It will provide a more equal balance in contribution management between, respectively, relevance/coherence and effectiveness/efficiency/learning.

Final Report from the Evaluation of Sida's work with Poverty

Main evaluation method: mixed-methods, synthesis analysis, literature review.

Positives: Across Zambia and South Sudan, Sida-supported interventions generated tangible benefits for poor populations, including improved incomes, food security, and access to essential services. In Zambia, interventions performed better as the systems are relatively stable while in South Sudan contributions have helped sustain basic services under extreme and fragile conditions.

Shortcomings: The impact results from Sida's work were uneven and remained limited in scale and transformative impact. Impact measures were often constrained by underuse of existing evidence in contribution designs and weak evaluability. Heavy reliance on partner-reported data, limited triangulation, and insufficient use of national data systems further created a verification gap and constrained deeper analysis.

SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY

Visiting address: Rissneleden 110, 174 57 Sundbyberg
Postal address: Box 2025, SE-174 02 Sundbyberg, Sweden
Telephone: +46 (0)8-698 50 00. Telefax: +46 (0)8-20 88 64
E-mail: sida@sida.se Web: sida.se/en

