unicef ✿
for every child

UNICEF EVALUATION OF IMPACT

Strategy and Action Framework 2022–2025

The global map included in this publication is stylized and not to scale. It does not reflect a position by UNICEF on the legal status of any country or area or the delimitation of any frontiers. The dotted line represents approximately the Line of Control agreed upon by India and Pakistan. The final status of Jammu and Kashmir has not yet been agreed upon by the Parties. The final boundary between the Republic of the Sudan and the Republic of South Sudan has not yet been determined. The final status of the Abyei area has not yet been determined.

Table of Contents

LIST OF FIGURES AND TABLE

Abbreviations and Acronyms

3iE	International Initiative for Impact Evaluation
CEP	Costed Evaluation Plan
CO	Country Office
CPD	Country Programme Development
DAC	Development Assistance Committee
DIME	Development Impact Evaluation
EAPR	East Asia and Pacific Region (UNICEF)
ECAR	Europe and Central Asia Region (UNICEF)
ECD	Early childhood development
EISI	Evidence Information Systems Integration (UNICEF internal database of evaluation, research and studies reports)
EO	Evaluation Office
ESAR	Eastern and Southern Africa Region (UNICEF)
FAO	Food and Agriculture Organization of the United Nations
GEROS	Global Evaluation Report Oversight System
HQ	Headquarters
IPA	Innovation for Poverty Action
JPAL	Abdul Latif Jameel Poverty Action Lab
LACR	Latin America and the Caribbean Region (UNICEF)
MENAR	Middle East and North Africa Region (UNICEF)
MSC	most significant change
OECD	Organisation for Economic Co-operation and Development
PME	planning, monitoring and evaluation
PSM	propensity score matching
RBM	results-based management
RCT	Randomised Control Trial
SAR	South Asia Region (UNICEF)
SDG	Sustainable Development Goals
SOP	standard operating procedure
TAG	Technical Advisory Group
ToC	theory of change
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commissioner for Refugees
WASH	Water, Sanitation and Hygiene
WB	World Bank
WCAR	West and Central Africa Region (UNICEF)
WFP	World Food Programme

Strategy at a Glance

- Few completed evaluations focused on impact and outcomes
- Thematic and geographic disparity in coverage
- Long-term programme planning for evaluating impacts

- Data limitations
- High cost and resources
- Limited use of robust methods

- Low awareness, capacity of staff and partners
- Weakness or absence of a theory of change (ToC)
- Misalignment of impact evaluation planning and programme cycle

STRATEGIC PILLAR I

INCREASE INITIATION AND COVERAGE

STRATEGIC PILLAR II

DIVERSIFY METHODS AND INNOVATE

STRATEGIC PILLAR III

IMPROVE LEARNING AND PROGRAMMATIC SYNERGY

ACTIONS

① Launch of the Impact Catalyst Fund
② Technical assistance on the use of OECD/DAC 'impact' criterion in summative evaluations
③ Integrate a requirement for impact evaluation for major donor proposals
④ Promote impact evaluation planning for any new and strategic interventions within CPD process (CEP)

ACTIONS

① The Methods' Innovation Lab tests and promotes new methods, tools and data applications
② The Evaluation Helpdesk provides support in impact evaluation designs, data scoping, application of digital tools among others

ACTIONS

① Develop capacity of UNICEF staff and partners through ongoing learning and programmatic initiatives
② Produce methodological briefs and guidances
③ Facilitate institutional and global learning through dissemination of impact evidence, exchanges on 'best practices' and innovation
④ Promote technical skills upgrading within evaluation function

VISION

BETTER NATIONAL POLICIES

Robust, timely and relevant impact evidence strengthens UNICEF advocacy and supports national partners' decisions to allocate more resources for child-focused policies and programmes

IMPROVED ORGANIZATIONAL EFFECTIVENESS

Impact evidence enables improved programming, demonstrates effectiveness and increases allocative efficiency of its child-focused investments

UNICEF defines impact as the **positive and negative, direct or indirect, primary and secondary, short, medium or long-term change in the lives of children and families produced by an intervention.**

Foreword

At UNICEF, we care about meaningful change in the lives of children. Change that is measurable, sustainable and transformative. Using scientific evidence rather than our ideas, assumptions or aspirations, allows us to make a judgement on the extent to which UNICEF makes a difference.

At UNICEF, we care about meaningful change in the lives of children. Change that is positive, measurable, sustainable and transformative. Using objective scientific evidence rather than our ideas, assumptions or aspirations, allows us to make a judgement on the extent to which UNICEF makes a difference.

As we enter the last decade of Agenda 2030 for the Sustainable Development Goals (SDGs), it becomes even more critical to evaluate our organizational footprint in achieving concrete results through scale-up of the most transformative interventions. It also becomes essential to make our programmatic efforts more efficient and strategically focused as UNICEF faces a shift in donor support towards more earmarked, thematic funding. This is why the new UNICEF Strategic Plan 2022–2025 focuses on outcomes and impact-level change to accelerate progress towards the SDGs and on greater accountability for investment of limited public resources. This reinforces the need for a coherent, comprehensive, and forward-looking strategy to evaluate outcomes and impacts of UNICEF interventions.

The Strategy reflects the long-term vision for a more harmonized approach to evaluating impact at UNICEF. It outlines the rationale, concepts, priorities and actions to be taken by the evaluation function at all levels and by the organization as a whole.

It sets methodological principles that support purpose-driven and a user-focused approach with a combination of methods available to evaluation professionals today. Rigorous and ethical impact evaluation is at the core of this Strategy. Yet it is not the only modality to evaluate long-term change, particularly in areas of upstream work that are important to 'move the needle for children'. The Strategy provides ideas and directions for the evaluation function to stay abreast of innovation, including technological advances. It also underscores the importance of forging partnerships with many prominent actors in the field to achieve more sustainable results and exchange best practices.

The enthusiasm, leadership and commitment of the UNICEF evaluation function at all levels will ensure that the Strategy is adopted and integrated across UNICEF. Equally, we count on the support of UNICEF programmes, as well as good coordination and cross-fertilization within UNICEF's evidence functions, and active engagement of our external partners. Together we have the power to make this journey purposeful and to assess and achieve impact for the organization and for children.

Robert McCouch
Director
Evaluation Office
UNICEF

1

Rationale and Vision

Meaningful change in the lives of children does not happen overnight. With less than a decade left to report on progress on Sustainable Development Goals (SDGs) as set by Agenda 2030, UNICEF allocates substantial resources and efforts towards achieving child well-being outcomes as set by Agenda 2030. Knowing if these efforts make a difference is an urgent task and our shared responsibility.

2017–2021

$ 23 billion

spent on programmes

36

impact evaluations produced

With increased global uncertainty, and a tightened and more earmarked financial envelope, it is time to leverage rigorous evidence that helps to advocate for, and target limited resources to, the greatest needs. With this Strategy and Action Framework the UNICEF evaluation function responds to the persistent demand for more strategic and consistent institutional efforts to deliver feedback on results of UNICEF and government-led programmes.

Towards greater accountability and transparency

Every year, UNICEF invests millions of dollars in improving delivery of basic services for millions of children and families, protecting the most vulnerable and driving system-wide policy change globally. It does so by scaling up well-tested and new solutions to social problems. As a mission-driven organization, UNICEF has responsibility to maximize social welfare of children with the public funds it receives.

Between 2018 and 2021, UNICEF invested just over 23 billion dollars on programmes and interventions in five thematic areas. With only 36 rigorous impact evaluations conducted over about the same period, the scale of rigorous evidence to date falls short of the scale of UNICEF development efforts and does not allow a sound public judgement on the relative effectiveness of chosen models and approaches to deliver benefits to millions of children.

Improved effectiveness, decision-making and organizational learning at all levels

Increasingly, UNICEF programmes support governments to address multi-faceted problems and work in complex environments including fragile and humanitarian contexts. As a result, interventions become more integrated and more innovative in their approaches. For example, in 2020 UNICEF initiated Giga, a global school connectivity programme in 13 countries, and expanded drone-focused projects from vaccine delivery to better emergency preparedness. No social solution is perfect from the outset. By testing and identifying early what works and what does not we can avoid wasting years of human effort and inefficient spending on interventions that can only be carried out for thousands rather than delivering benefits to millions. We should not lose sight of what matters.[1]

Rigorous and cost-effective impact evaluations and outcome analysis should become a part of the programme journey to help us sustain a focus on results.

Ambitious, results-oriented UNICEF 2022–2025 agenda

Over the last decade, UNICEF has utilized evidence on the positive impacts of important interventions such as unconditional cash transfers in Africa to catalyze policy change.[2] New programmatic priorities (e.g. building resilience, supporting children's and adolescents' mental health, equal access to digital services) as well as old challenges placed in a new context (e.g. intensified climate shocks, post-COVID-19 economic landscape, conflicts) exposed evidence gaps that need to be filled to inform progress on the UNICEF Strategic Plan 2022–2025. Prioritization and a clear strategy to generate scientifically sound evaluative evidence at scale in key areas of UNICEF engagement are required to catalyze organizational advance to the next development frontier for children.

UNICEF Evaluation, guided by the Evaluation Policy of 2018, and with the oversight of the Director of Evaluation and the UNICEF Executive Board, promotes and supports the culture of better performance, continuous learning and strengthened accountability for sustainable results. It is supported in this by other data and evidence functions. It has accountability to the UNICEF Executive Board to strengthen its work on evaluating the impact of UNICEF interventions up to 2025 and beyond. Supporting the new UNICEF Strategic Plan 2022–2025 and recognizing increased pressures within donor countries for greater accountability on development spending, the member states repeatedly called for the UNICEF evaluation function to *"expand the use of rigorous impact evaluations in the period up to 2025, particularly in thematic areas where new approaches have potential for scaling up or where existing impact evidence on approaches is lacking".*[3]

This is aligned with recommendations of the MOPAN Assessment Report (2021) which underscored the need for "more robust evaluative evidence on efficiency and sustainability of UNICEF programming".[4]
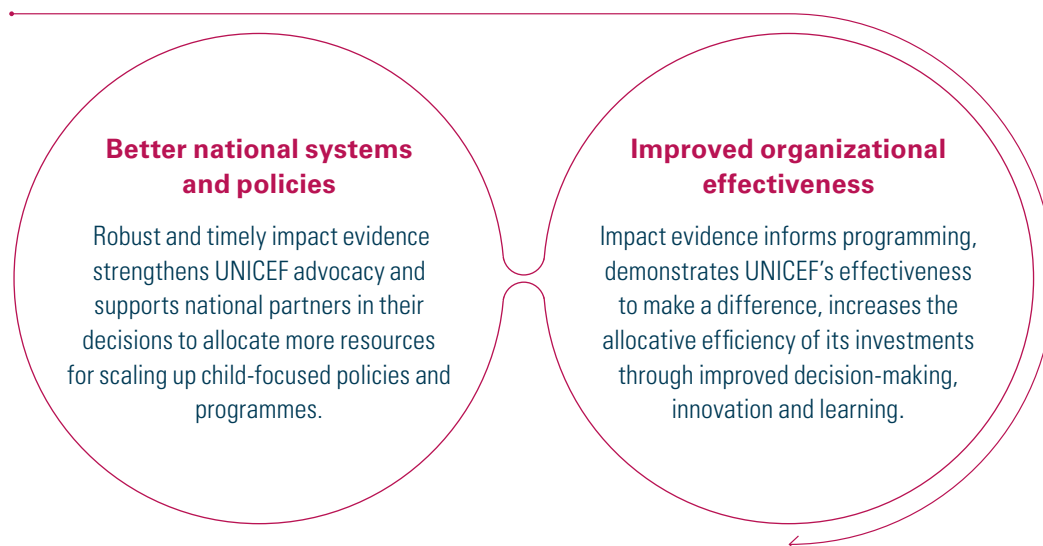
Meeting the evidence needs through rigorous[5] impact and outcome-level evaluative evidence will allow UNICEF to:

▶ allocate efficiently limited resources at the decentralized level where they will make the biggest difference for children as well as mobilize new resources by demonstrating the added value of UNICEF investment to achieve progress for SDGs;

▶ foster institutional learning and enhance the programme implementation process by strengthening the underlying programme logic and testing its assumptions; and

▶ enable UNICEF to stay relevant for the national policy process, strengthen advocacy where it matters, and critically assess UNICEF's contribution to results outlined in the 2022–2025 Strategic Plan.

This Strategy and Action Framework ('the Strategy') is the first step in advancing this agenda within the next quadrennial and beyond. In this it directly supports the new vision for the evaluation function for 2022–2025, which is to consistently leverage rigorous, strategically prioritized evaluative evidence for UNICEF and its partners to realize the rights of every child in the Decade of Action.

With this forward-looking vision this Strategy aspires to foster:

Better national systems and policies

Robust and timely impact evidence strengthens UNICEF advocacy and supports national partners in their decisions to allocate more resources for scaling up child-focused policies and programmes.

Improved organizational effectiveness

Impact evidence informs programming, demonstrates UNICEF's effectiveness to make a difference, increases the allocative efficiency of its investments through improved decision-making, innovation and learning.

The Strategy serves as a practical reference to UNICEF regional and country offices, evaluation and programme staff in their efforts to inform national plans and child-focused policies with robust evaluative evidence. It hopes to present basic technical concepts in an accessible way with the expectation that this document will stimulate further technical enquiries found elsewhere. The document will cover operational definitions, highlight identified challenges and present three strategic pillars of work to respond to these challenges including corresponding actions. It will then discuss UNICEF approaches to methodological and process-related aspects and will outline partnership and resource requirements.

---

1 Ann Mei Chang (2019). Lean Impact. How to Innovate for Radically Greater Social Good. John Wiley and Sons, New Jersey.

2 For example, UNICEF has contributed to global knowledge on the impacts of unconditional cash transfers in Africa on child outcomes through the Transfer Project, implemented jointly by the Carolina Population Centre (CPC) of the University of North Carolina at Chapel Hill (NCU), the UNICEF Office of Research-Innocenti. and the UN Food and Agriculture Organization (FAO).

3 Executive Board Sessions: E/ICEF/2021/L10 (1–4 June 2021); E/ICEF/2022/L.5 (8–11 February 2022).

4 MOPAN Assessment Report (2021). United Nations Children's Fund (UNICEF). 2020 Assessment Cycle. Available at: MOPAN | Multilateral Organisation Performance Assessment Network (mopanonline.org).

5 There is no universally applied definition of 'rigorous' in impact evaluation. For the purpose of this document, it is understood as the application of a combination of methods that aim to isolate the effects of the programme from other factors and potential selection bias, ultimately aiming to achieve internal validity and high credibility of findings.

2

Defining Impact and Scope

This Strategy defines impact as the positive and negative, direct or indirect, primary and secondary, short, medium or long-term change in the lives of children and families produced by an intervention.

causality
pathways

policy
pathways

Policy
actions
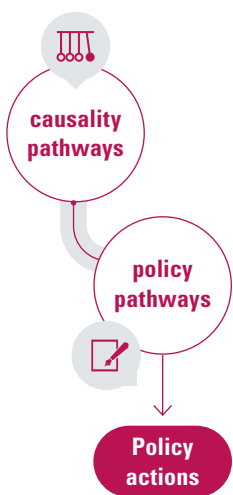
This definition is broadly aligned with the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD) definition of impact as it emphasizes a wide range of outcomes beyond those specified in a theory of change (ToC), implies a measurable change, and reinforces the notion of causal attribution – effects are 'generated' by intervention.[6]

High level, sustainable change can happen through a non-linear and interactive chain of effects along the pathways of the proposed ToC and be influenced by a range of social and cultural norms which are not easily amendable by policy action. UNICEF evaluations which examine impact will have to look at the results through the short-term and intermediary outcomes defining them as causality pathways to sustainable, long-term impacts. This is critical because 'causality pathways' can be translated into 'policy pathways' with concrete policy actions towards desired transformative change. The latter can imply a system-level change alongside tangible and measurable well-being outcomes that can be evaluated through a rigorous counterfactual. Examples include, but are not limited to, SDG target 4.2.1 ensuring children are developmentally on track; SDG 4.6.1 on achieving minimum literacy levels; SDG 2.1.1 and 2.2.2 on the reduction of stunting and wasting; and SDG target 5.4.1 on gender equality and time spent on domestic and care work, among others. The UNICEF definition underscores the importance of a scientific framework offered by rigorous impact evaluation (IE) while also recognizing the value and unique insights provided by qualitative, non-experimental evaluative approaches to examine causal relationships. Given the broad programmatic mandate of UNICEF, the Strategy outlines the Evaluation of Impact as a general scope of evaluative work at the outcome and impact level at UNICEF, that employs two distinctive strands of impact inquiry reflective of evaluative purpose (as opposed to method): examining causal attribution or causal contribution of a specific programme or its components to programme impacts (text box 1).

Text box 1

Establishing a causal relationship: attribution or contribution?[7]

Figure 1. Attribution and contribution in development evaluation
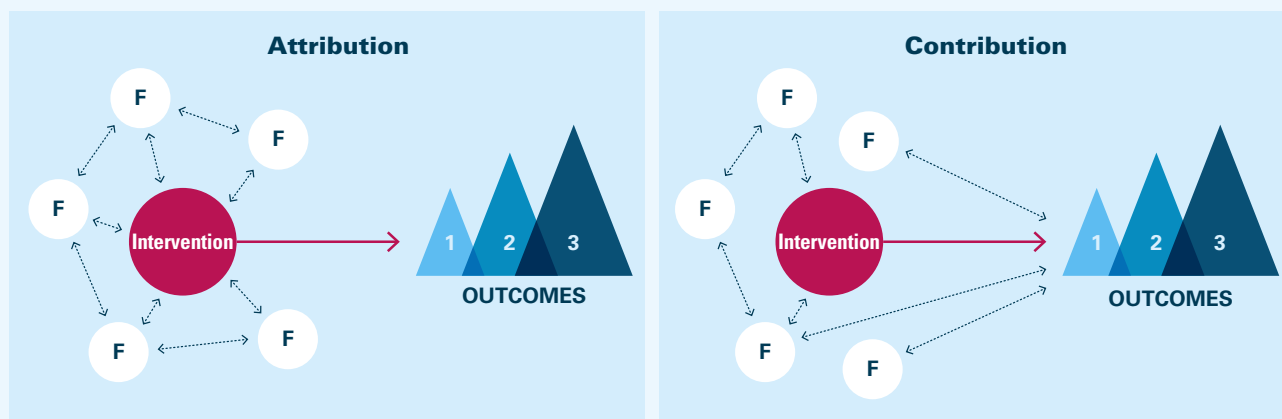
Note: Graphic is adapted from Tamarack Institute, Evaluating Community Impact. Capturing and Making Sense of Community Outcomes, ppt
'F' - factors affecting or interacting with the programme

Attribution

As any programme competes for limited public resources, UNICEF staff, donors and partners want to know if the intervention made a difference as a result of designated inputs and irrespective of underlying biases and contextual factors. Changes should be substantively large to justify resource allocation at scale, so a relevant linked question is 'by how much', referring to the degree of attributed effects. To answer this question requires:

a. **isolating the effect** of the programme from other factors and potential selection bias and

b. **demonstrating a causal link** between observed (expected or unexpected) changes and a specific intervention with high level of internal validity (estimation accuracy).

Analysing attribution requires comparing the observed situation in the presence of an intervention to the situation without it. The most robust way to do this is through constructing a **counterfactual**, namely "**outcomes for participants had they not been exposed to the programme**".[8] As an individual cannot be at the same time with and without a programme we need to find his/her closest comparison.

One can think of identical twins as the closest analogy for a comparison (one receives an information package on nutrition, and the other does not). In reality, even twins will have some unobserved differences that can affect the comparison.

Random selection of individuals into two groups is one validated method to ensure close comparability because it ensures the treatment and control groups are similar in all aspects except that one receives an intervention and another does not (experimental design or randomized controlled trial – RCT).

Other methods may include instrumental variables, regression discontinuity design (RDD), difference-in-difference (DID), and matching, use statistical techniques to find the closest matching groups of individuals (quasi-experimental designs).[9] In addition, natural experiments exploit naturally occurring phenomena to find an appropriate matching case with and without a programme using observational data statistical techniques. They all have the advantage of formally addressing a counterfactual problem and normally are at the core of an impact evaluation design.

Contribution

Often direct attribution is not possible or desirable for a range of reasons. First, the outcomes or impacts of interest might not be easily measurable or quantifiable (e.g. institutional change, policy dialogue, governance, impact of advocacy efforts). Second, the nature of the programme could reflect the contribution of many actors, convergence of different policies and actions interacting with environmental, political and other factors to bring about systemic change. In this case, isolating the net effects of one intervention or a 'package' may be not feasible.[10] Finally, there are operational issues that can make attribution analysis unfeasible: misalignment between evaluation and programme planning often results in the lack of baseline data resulting in retroactive tracing of results.

Selected qualitative methods (causal case study designs within theory-based evaluations / 'non-experimental approaches'[11]) aim to understand how, why and under what circumstances programmes and policies work by utilizing explicit reference to the underlying programme and policy logic. Methods such as process tracing, contribution analysis, the most significant change, qualitative comparative analysis, among others, have the comparative advantage of assembling qualitative evidence from a wide range of sources, and perspectives and explaining the mechanism between the cause and effect. They rely on chains of logical, programme theory-based arguments and assumptions (as described by the ToC) about the causal pathways between the programme and results that can be verified through a systematic and iterative analysis of available evidence and data ('detective work').

a. **collecting evidence** to verify 'causal claim' of the programme contribution to outcomes based on the ToC and its assumptions

b. **understanding and explaining** how, why and under what circumstances programmes and policies work

Both causal attribution and causal contribution are relevant for UNICEF efforts to evaluate the impact-level change. Programme characteristics examined in conjunction with a ToC will help to assess the feasibility of establishing programme attribution versus contribution or a combination of both within specific a context (figure 2). These include:

Programme characteristics

▶ **The nature of intervention** (e.g. level of intervention and its coverage) and its outcomes (e.g. measurable or not?);

▶ **Operational features of the programme** (e.g. eligibility criteria of beneficiaries, scale of the programme, cost, the timeline of the intervention);

▶ **Type of evaluative questions** (e.g. cause and effect? descriptive? normative? explanatory?); and

▶ **The purpose or intended use of impact evidence** (e.g. major scale-up or ongoing improvement?).

Figure 2. Selecting evaluative approaches to measure programme outcomes and impact
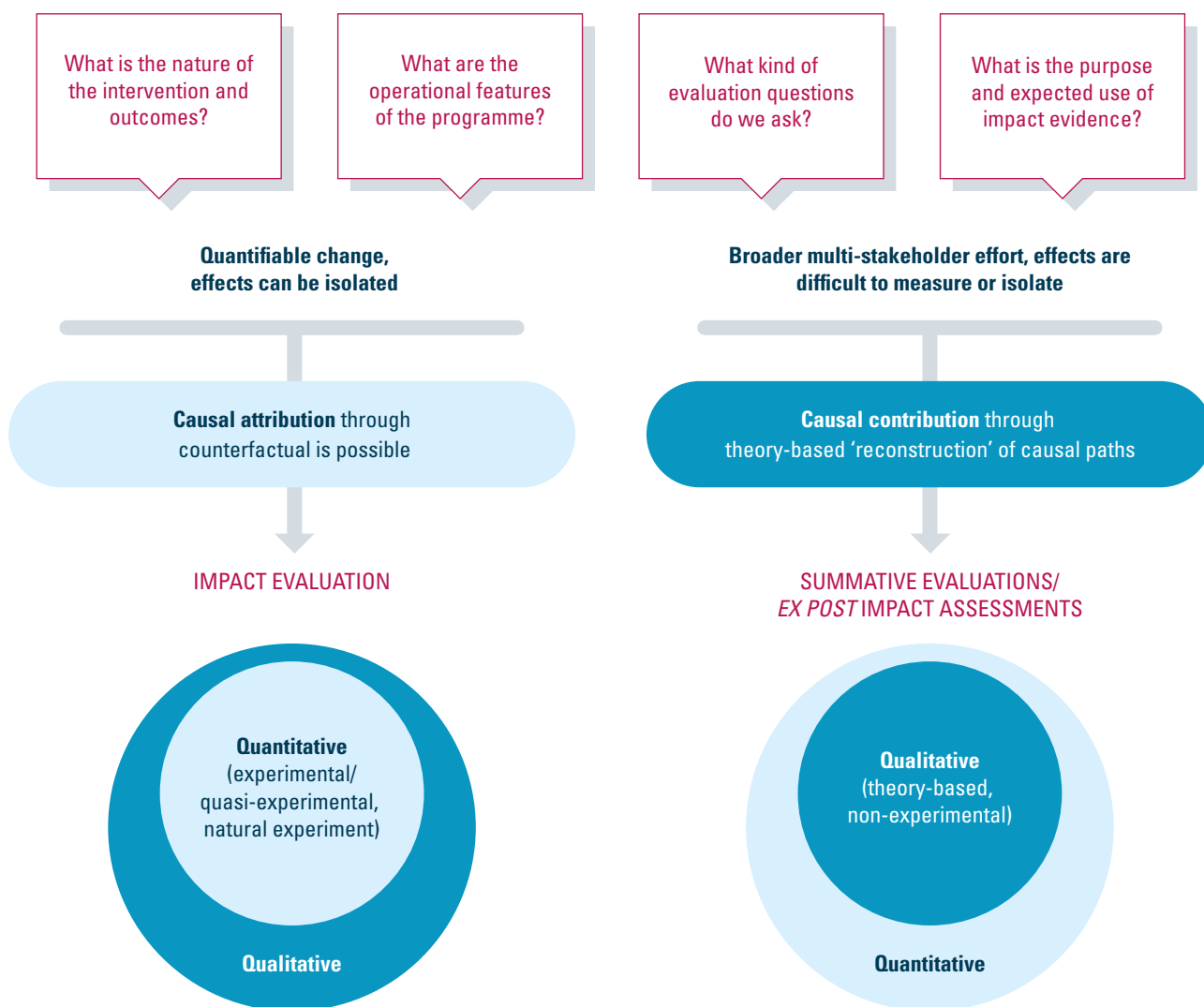
Operational modalities to evaluate impact

Figure 3 presents the proposed operational typology to evaluate impact at UNICEF. It builds on the conceptual distinction between attribution and contribution made earlier and associated methodological requirements to design a counterfactual.

A strong ToC of the programme is a prerequisite for all evaluations examining the impact and outcome-level change.

The typology helps with methodological choices but does not restrict the range of design options under any single evaluation type (see text box 2 for UNICEF examples). A strong ToC of the programme is a prerequisite for all evaluations examining the impact and outcome-level change. While each modality presented in figure 3 has its chosen core method (micro-simulation, quasi-experimental or 'theory-based'), combinations of approaches are highly encouraged as they increase robustness of conclusions and the overall credibility of the evaluation report.

Ex ante evaluation

Formative assessment aims to predict the impact of a policy or programme prior to its implementation. This helps to identify optimal policy designs and their potential distributional effects (e.g. equity), as well as explore potential unintended effects of a policy or programme. Often economic and micro-simulation techniques are used to model expected impacts of future programmes, estimate the effects of subsidies and other cash transfers, and can describe the potential range of anticipated benefits. Contrasting these *ex ante* predictions with those observed from *ex post* impact evaluation can provide insights on the changes due to the real-life programme modifications.

Impact evaluation (IE)

Counterfactual

find the closest comparison for the unit of study

IE attributes changes in development outcomes to a specific programme, project or intervention using a credible counterfactual. This allows us to see what would happen in the absence of the intervention or exposure to it thus allowing a judgment on whether the intervention makes a difference and to what extent. It can serve both formative and summative purposes and respond equally to learning and accountability needs. If aligned with the programme planning stage, it allows feedback to the programme design and theory of change. Experimental, quasiexperimental or natural experiment designs form the core of impact evaluation methods. They estimate outcomes and programme impacts not only in magnitude but also in terms of statistical significance while attempting to minimize the selection biases or unobservable factors or influences using statistical techniques. They are best implemented in combination with appropriate qualitative approaches that answer questions on 'Why?' and 'How' to deepen and extend our understanding and interpretation of observed results.

Impact evaluation (IE) aims to empirically establish attribution by quantifying causal links between the intervention (projects, programmes, policies, networks or capacity development) and outcomes of interest, typically at the targeted beneficiary level.[12] It does so by establishing a counterfactual scenario.

Ex post impact assessment (IA)

A summative, theory-based, empirical analysis will focus on the achieved results in the light of complex and evolving contexts. Similar to the impact evaluation, IA answers causal questions based on the theory of change and underlying behavioural assumptions. Unlike impact evaluation though, it examines a programme's contribution to the observed outcome (no counterfactual) and does not ascertain the causal linkages through statistically significant effects. As a stand-alone analysis it can utilize a wide range of non-experimental approaches as its core methodology. IAs are well suited to evaluate outcomes and impacts of complex and multi-strand interventions where isolation of programme effects through a counterfactual is not possible, or other conditions for impact evaluation cannot be met.

Summative evaluations with integrated causal questions about a long-term change (application of OECD/DAC impact and other criteria)

This method "looks at higher order effects and broader changes to which an intervention may be contributing".[13] It captures the overall significance and potentially transformative effects of the intervention/programme or project. Theory-based, non-experimental (qualitative) analytical approaches can be most appropriate given the summative nature of evaluation. Integration of a counterfactual is possible (see the case of the Nigeria Impact Evaluation in text box 2) but difficult due to operational challenges such as timing or resources allocated for integrating the 'impact component' alongside other evaluative criteria (e.g. relevance, effectiveness, efficiency, sustainability, equity). The main difference between this modality and *ex post* IA is that the 'impact' focus is embedded in a broader evaluative scope.[14]

Figure 3. UNICEF evaluations of impact operational typology and complementarity of methods

PRIMARY FOCUS ON OUTCOMES AND IMPACTS			PART OF A BROADER ANALYSIS
EX ANTE EVALUATION	IMPACT EVALUATION	*EX POST* IMPACT ASSESSMENT	SUMMATIVE EVALUATION WITH OTHER OECD/DAC CRITERIA
Simulating the effect of a programme **(Quantitative)**	Counterfactual design **(Quantitative)** / Complementary **qualitative** approaches	Theory-based contribution techniques **(Qualitative)** / Complementary **quantitative** analysis (monitoring, survey and other data)	Theory-based contribution techniques **(Qualitative)** / **Quantitative** / Relevance, effectiveness, coherence etc.

Key message: Sustainable, long-term[14] change in the lives of children is evaluated through causal pathways which include short-term and intermediary outcomes as defined by a ToC. This will allow the programme to identify immediate programmatic and policy actions that lead to the desired transformative change, which in itself is not immune to environmental, political and economic shifts. Impact evaluation, which aims to empirically establish attribution by quantifying causal links between the programme and outcomes, is the core focus of this Strategy with a due recognition that non-experimental approaches are well positioned to critically assess UNICEF contribution to results in complex programmatic settings. The evaluator's task is to define the evaluative purpose aligned with the desired 'attribution' to programme results or 'contribution'. This will help to define the most appropriate and feasible methodological approaches, including those that allow us to construct a rigorous counterfactual.[15]

---

6   OECD (2019). Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use, OECD/DAC Network on Development Evaluation

7   Some literature uses 'attribution' and 'contribution' interchangeably. As that terminology can be used in a variety of contexts, possible disagreements may arise among scholars. The task of this document is to provide some operational clarity and a conceptual framework for UNICEF and partners that seek evidence on programmatic results through evaluation

8   Khandker, Shahidur R., Koolwal, Gayatri B., Samad, Hussain A. (2010). 'Handbook on Impact Evaluation'. The International Bank for Reconstruction and Development/The World Bank.

9   All quasi-experimental designs, unlike randomized control trials, require further assumptions to make causal attribution, some to a larger degree than others (e.g. matching). Current development practice considers RDD and DiD as the most robust alternatives to RCT. Yet feasibility of specific experimental or quasi-experimental approaches depends on many programmatic and contextual factors as well as data availability and quality.

10   Impact evaluations routinely use 'evaluation arms' to investigate different intervention components or 'packages'. However, with too many activities 'bundled' together it becomes more difficult to explain observed effects.

11   Lemire et al (2020) refer to causal case study methods as a strand of new emerging methods rooted in theory-based evaluation in general (Chen and Rossi, 1983).

12   White, H., and Raitzer, D.A. (2017). 'Impact Evaluation of Development Process'. Asian Development Bank.

13   OECD (2019). 'Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use', OECD/DAC Network on Development Evaluation.

14   The definition of 'long-term' can vary depending on the nature of the programme and outcome in focus. It can range from between 3 and 4 years to between 5 and 6 years. However, it can be argued that from the political and operational point of view, 4 to 5 years is a reasonable term to see sustainable results.

15   Section 5 further elaborates on the operational modalities and methodological options to evaluate impact at UNICEF.

3

Current Challenges

This section draws on the mapping of impact evidence using the UNICEF Evidence Information Systems Integration (EISI) database as well as the bottleneck analysis of responses to the online survey of UNICEF evaluation practitioners.

2018–2021

UNICEF evaluative products

627

36

impact evaluations
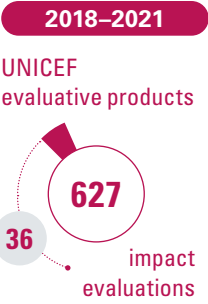
The EISI database contains 627 evaluative products conducted or commissioned by UNICEF between 2017 and 2021. Only 36 of these are rigorous impact evaluations as they employ various methods to construct a counterfactual. The total number is inclusive of products classified in the database as research (3) and studies (7).[16]

The scale of rigorous impact evidence generated is in sharp contrast to the scale of UNICEF programmatic investment over the same period, totaling US$28.5 billion.[18]

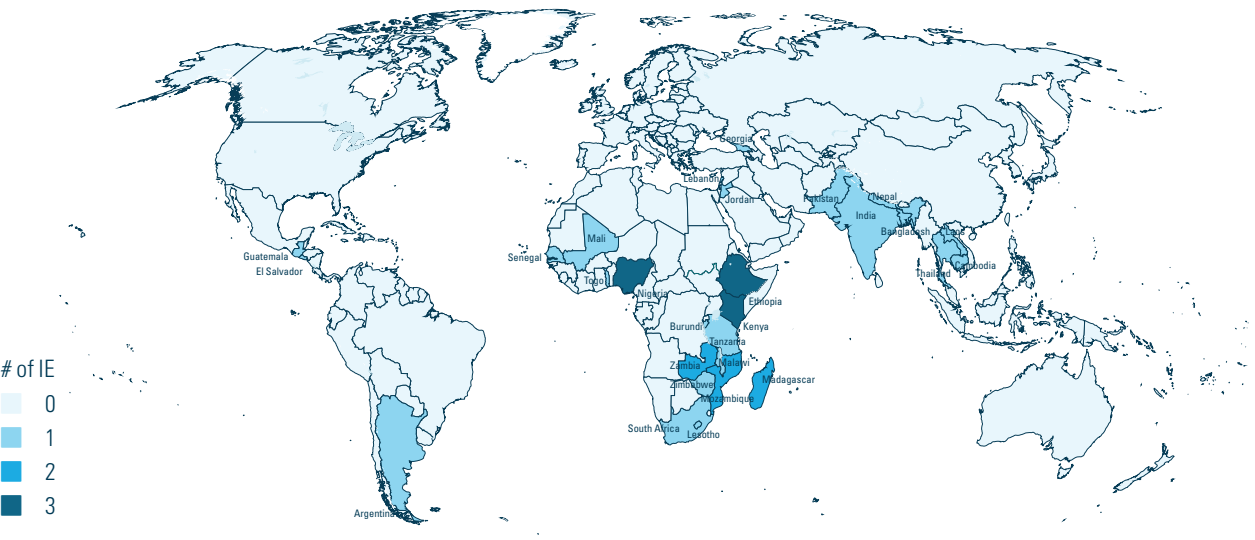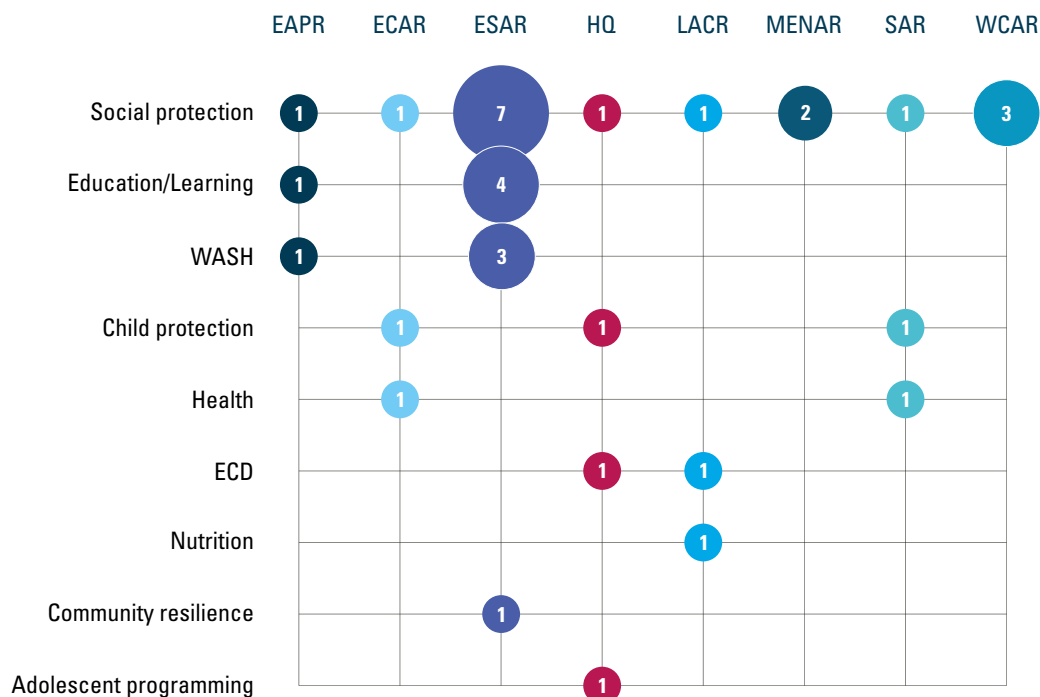**Figure 4.** Country coverage with UNICEF impact evaluations 2017–2021

# of IE

- 0
- 1
- 2
- 3

Figure 5. Absolute number of UNICEF impact evaluations 2017–2021 by thematic areas and regions

	EAPR	ECAR	ESAR	HQ	LACR	MENAR	SAR	WCAR
Social protection	1	1	7	1	1	2	1	3
Education/Learning	1		4					
WASH	1		3					
Child protection		1		1			1	
Health		1					1	
ECD				1	1			
Nutrition					1			
Community resilience			1					
Adolescent programming				1				

Note: EAPR (East Asia and Pacific Region), ECAR (Europe and Central Asia Region), ECD (early childhood development), ESAR (Eastern and Southern Africa Region), HQ (Headquarters), LACR (Latin America and the Caribbean Region), MENAR (Middle East and North Africa Region), SAR (South Asia Region), WASH (Water, Sanitation and Hygiene), WCAR (West and Central Africa Region).

Only 5

impact evaluations conducted in humanitarian or fragile setting

Figure 4 presents the geographic coverage of the 36 impact evaluations conducted. Further, figure 5 indicates a highly skewed thematic focus on the social protection programme area[18] and greater coverage in the region of Eastern and Southern Africa. It appears that Goal Areas 1 (Nutrition/Health) and 3 (Child Protection)[19] have particularly acute gaps in rigorous evidence.

The lowest geographical coverage is found in the regions of Eastern and Central Europe, and Middle East and North Africa. Figure 5 shows that three countries - Ethiopia, Kenya and Nigeria - have the largest number of IEs over that period. Of the few evaluations performed, most (over 70 per cent) received a 'satisfactory' rating in the UNICEF Global Evaluation Report Oversight System (GEROS).

Figure 6 shows that out of the 36 impact evaluations produced, 15 used experimental design (RCTs) and 21 quasi-experimental, with the propensity score matching (PSM) technique being the most common choice (4 evaluations used PSM combined with a difference-in-difference approach and 9 only PSM).

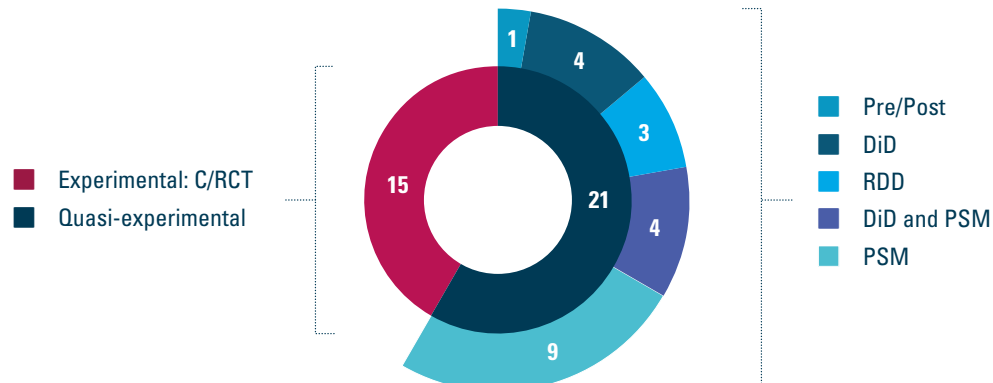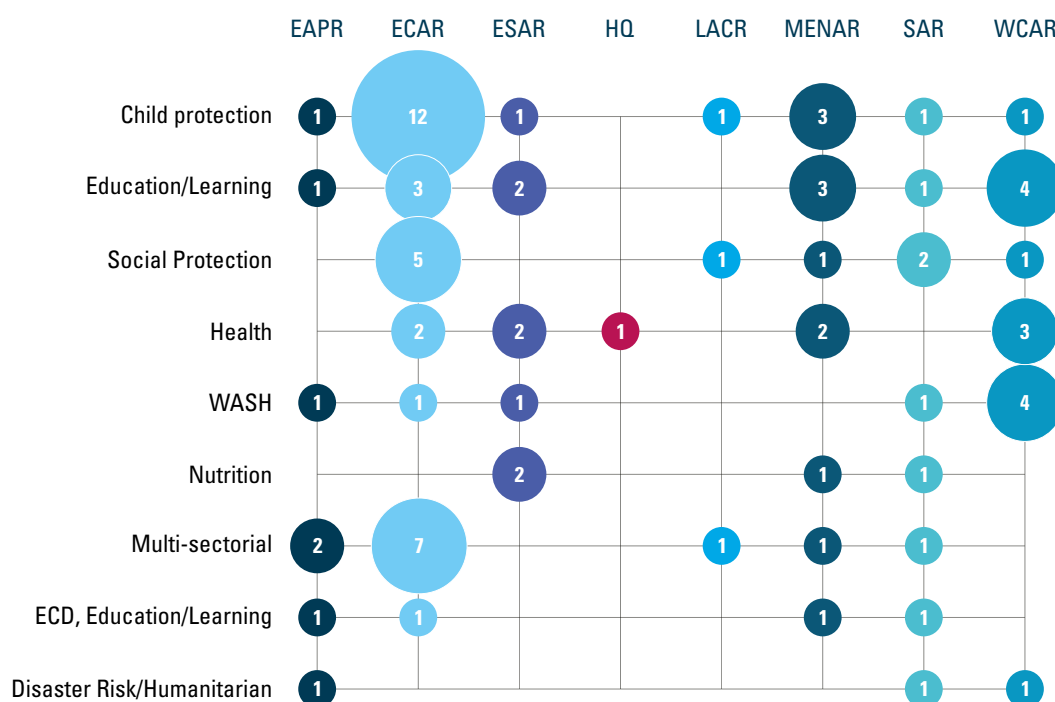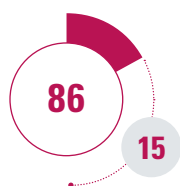**Figure 6.** Methodological choices in UNICEF evaluations with OECD/DAC impact criterion

Legend:
- Experimental: C/RCT
- Quasi-experimental
- Pre/Post
- DiD
- RDD
- DiD and PSM
- PSM

Values: 15, 21, 1, 4, 3, 4, 9

Figure 7. Absolute number of summative evaluations using the OECD/DAC impact criterion between 2017–2021 by thematic area and region

	EAPR	ECAR	ESAR	HQ	LACR	MENAR	SAR	WCAR
Child protection	1	12	1		1	3	1	1
Education/Learning	1	3	2			3	1	4
Social Protection		5			1	1	2	1
Health		2	2	1		2		3
WASH	1	1	1				1	4
Nutrition			2			1	1	
Multi-sectorial	2	7			1	1	1	
ECD, Education/Learning	1	1				1	1	
Disaster Risk/Humanitarian	1						1	1

Only 28/36

impact evaluations included a theory of change

86 **15**

summative evaluations used appropriate causal analysis methodologies

Overall, the use of methods in IEs shows a relatively balanced choice between experimental and quasi-experimental approaches in impact evaluations and very limited utilization of any causality-focused non-experimental/theory-based approaches in the evaluations with the OECD/DAC criterion of impact.

Out of the 86 evaluations with the impact criterion, 63 did not use any methods of causal inference. Only 15 used theory-based approaches, the most common ones being contribution analysis (5 evaluations) and realist evaluations (4 evaluations). Several combined these methods with outcome mapping, system thinking, micro-simulation and comparative case studies. It is worth noting that two evaluations used quasi-experimental designs under the OECD impact criterion and six offered a mix or hybrid of causal approaches.

A similar picture emerges in the use of the OECD/DAC impact criterion in addition to the standard range of evaluative criteria (figure 5). Only 86 out of 627 evaluations (14 per cent) conducted over the same period had an explicit focus on impact and outcome-level results.

There was also a clear decline in the use of the impact criterion during the previous Strategic Plan (from 27 products in 2017 to 6 in 2020). This only changed in 2021 with 19 new evaluations completed.

Thematic distribution of evaluations with the OECD/DAC impact criterion shows a significant thematic and regional imbalance in coverage. Almost a quarter (21 in number, or 24 per cent of the total) of all evaluations with impact criteria were conducted in the thematic area of child protection. Together with evaluations in education these account for 35 evaluations or 41 per cent of the total number of evaluations with OECD/DAC impact criteria. The Europe and Central Asia region leads with 31 evaluations (36 per cent) followed by West and Central Africa (14 evaluations or 16 per cent). There is an emerging trend to use an impact criterion when evaluating cross-cutting themes, but they are still few.

Only 13 evaluations out of all those that focused on impacts (5 in impact evaluations and 8 in those using the OECD/DAC impact criterion) were conducted in an emergency context. However, they cannot all be classified as 'humanitarian' evaluations. This indicates a significant gap of impact evidence in the humanitarian context.

Programme planning

IE planning
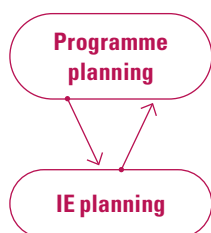
In addition, UNICEF undertook a bottleneck analysis based on the online survey among UNICEF staff[20] and collected expert opinion. Focusing on UNICEF impact evaluations, it highlighted the most common, persistent and interconnected implementation challenges at the decentralized level. Although inevitably contextualized and interlinked, these issues hinder planning, generation and use of robust impact evidence at the different stages of the programme cycle.

Challenges include:

▶ **Misalignment of impact evidence planning with the programme/project cycle and scaling-up efforts**

There is no formal requirement for rigorous evidence on outcome and impact-level results. The idea of evaluating impact-level change often comes too late in the process without a direct link to the strategic advocacy dialogue with partners on scale-up. This results in a lack of ownership by national counterparts, missed 'windows of opportunity' to collect good baseline data and consecutive issues with evidence take-up.

▶ **Weakness or complete lack of a theory of change (ToC)**

This undermines the formulation of intermediary outcomes of interest, pathways and assumptions, resulting in poor quality and low credibility of impact evidence. Diagnostic mapping showed that only 28 out of 36 (78 per cent) impact evaluations conducted between 2017 and 2021 had ToCs, either before the evaluation was conducted or after the evaluation team reconstructed it.

▶ **Data limitations**

In the absence of a credible primary baseline – particularly in a humanitarian context – the use of alternative data sources (e.g. administrative or household survey) can be constrained by poor data quality, access limitations and political sensitivity. This impacts the ability to construct a credible counterfactual. Moreover, the capacity at country level to canvas data sources and diagnose their quality poses further challenges.

▶ **High cost and resources**

The level of expected outreach by the initiative, its importance in the national context and its potential for scaling up can define the need, appropriateness and feasibility of implementing impact evaluations. The high cost associated with it (mostly attributed to the multiple rounds of primary data collection) often becomes a stumbling block in planning any impact evidence at the outset of the programme. Technical expertise required for planning and managing this type of evaluation can also hinder the process.

▶ **Low capacity**

Lack of clarity among UNICEF programme and even monitoring and evaluation staff, management, donors and partners on the benefits of rigorous impact evidence, confusion between 'attribution' and 'contribution' and their methodological and data requirements, limit support, early planning and capable management of impact evaluations at the country level. Low capacity among all stakeholders might also lead to differences in expectations on their value and limitations in answering specific impact evaluation questions.

▶ **Maturity of the impact and long-term (circa 4–5 years) programme planning**

In some programmatic areas intermediary outcomes and impacts might take a longer time to mature (e.g. behavioural change at the community level, change in social norms). The operational context of UNICEF programming, including resourcing, does not necessarily provide incentives for long-term planning and continuity in impact measurement.

UNICEF staff noted an unmet demand from key external stakeholders to generate impact evidence, including a growing interest from government partners, and the need to diversify the thematic coverage of the impact evaluation portfolio into new thematic areas. The survey respondents also stressed that the UNICEF programmatic conditions often require insights not only on 'What works' but also illuminating 'Why' and 'How?' and 'At what cost?' to better understand UNICEF's effectiveness in advancing progress on child-focused SDG areas. The questions on cost-effectiveness are rarely asked but are critically important for any national decisions on scaling up effective interventions. UNICEF staff also noted the lack of clear institutional guidance on the programmatic requirements to conduct impact evaluations including expected financial outlays at different levels of the organization and standardization of practices.

Key message: The diagnostics of rigorous UNICEF impact evidence to date and evaluative work on the OECD/DAC impact criterion showed limited institutional demand for rigorous evaluative evidence at the outcome and impact level and the lack of strategic, at scale, thematic planning for impact evidence on 'what works' within the previous Strategic Plan 2018–2021 with the exception of SDG area 5, social protection. The analysis of the status quo reflects persistent bottlenecks in initiation, planning, implementation and uptake of evidence focused on outcomes and impacts and the lack of institutional guidance on this type of evaluation. The interconnected nature of the problems identified by UNICEF staff and experts requires a more systemic and institutionalized approach, which allows integration of an incentive structure and requirements within a long-term development time frame.

---

16 Methodology of selection was based on selected technical criteria which included meeting evaluative purpose and presence of actionable programme recommendations. Ten impact evaluations classified as research or studies met those criteria and were included in the total number with a corresponding adjustment in the denominator.
17 The number of impact evaluations in social protection includes evaluations conducted under the Transfer Project.
18 UNICEF information database InSight, Management reports: Strategic Plan 2018–2021, 2017 analysis cubes.
19 Under the new UNICEF 2022–2025 Strategic Plan.
20 The online survey focused on impact evaluations was conducted by the UNICEF Evaluation Office in 2020. It is based on a non-representative sample with a mix of semi-structured and open-ended questions. A total of 55 responses were received representing Country Offices (42), Regional Offices (11) and Headquarters (2). Additional consultations with UNICEF evaluation experts contributed to this list.

4

Strategic Pillars and Actions

The identified challenges and bottlenecks in evaluating impacts described in the previous section give us a clear direction on the choice of priority pillars. These are formulated bearing in mind the feasibility of action within the period covering the UNICEF Strategic Plan 2022–2025. Our approach builds on the vision for the evaluation function for 2022–2025 "to consistently leverage rigorous, strategically prioritized evaluative evidence for UNICEF and its partners to realize the rights of every child in the Decade of Action".[21]

STRATEGIC PILLAR I

INCREASE INITIATION AND COVERAGE

STRATEGIC PILLAR II

DIVERSIFY METHODS AND INNOVATE

STRATEGIC PILLAR III

IMPROVE LEARNING AND PROGRAMMATIC SYNERGY

Strategic Pillar I

Increase initiation and coverage

Objective: Expand UNICEF evaluation of impact to new priority themes and new contexts, including humanitarian, and ensure accountability to outcome and impact-level results for major donor-funded projects and programmes.

Actions to support this include:

1   **Launch of the Impact Catalyst Fund** will meet already existing demand for rigorous impact evaluations through technical support and matching grants to country offices.

2   **Stimulate better utilization of the OECD/DAC impact criterion** with appropriate causal analysis designs of outcomes and impacts in global, regional and country-led evaluations, including those conducted in humanitarian and fragile contexts wherever appropriate and feasible.

3   **Promote the inclusion of an impact evaluation requirement in major donor-funded projects** through strategic consultations with key UNICEF donors as well as the set of criteria in Standard Operating Procedures (SOP) at the country office level to ensure its integration into an evidence plan at the planning stage.

4   **Require rigorous impact evaluation planning for any new, strategic pilot interventions during a new country programme development (CPD) process** when appropriate for the size of the country office and its programmatic focus.

The Impact Catalyst Fund aims to incentivize generation of high quality and strategically focused impact evidence that matters for organizational learning as well as accountability. It will identify pivotal dimensions that can give the maximum return on organizational investment. Its 'thematic windows' will be anchored in both the humanitarian and development context and innovative solutions to basic service delivery to ensure scalability but also adaptability of approaches. The UNICEF Strategic Plan 2022–2025 prioritizes the need to "systematically integrate a humanitarian lens into the analysis shaping its theories of change and consequent results framework". Focusing organizational resources on evaluating the outcomes of UNICEF interventions in fragile, protracted conflict and disaster contexts will contribute to the effectiveness of emergency programming as well as to global learning and practice. Given the inherent weaknesses of monitoring data in such settings, and its prevalent focus on activity level indicators, commonly lacking programme theory, great uncertainty and potential security risks, special attention has to be placed on seeking alternative data sources including new digital mobile data and/or natural experimental opportunities without necessarily discarding opportunities for face to face primary data collection when it is ethically sound and appropriate. Focusing on immediate and short-term outcomes (that can be reasonably measured within a one-year

time frame) is a realistic agenda that can support programmatic learning particularly at the critical junction linking humanitarian and development programming and thus contribute evidence on results beyond outputs.

The choice of thematic areas for increased impact coverage will be driven by identified global and regional knowledge gaps with a particular focus on priority areas of UNICEF programming within the next quadrennial. Multi-country impact feasibility assessments (IFAs) led by the UNICEF Evaluation Office (EO) in close consultation with programme group and thematic technical leads at all levels of organization, external experts and partners, could be one way to systematically assess the thematic impact evidence gaps at the global and regional levels. New areas of engagement can include digital services, adolescents' mental health, resilience in fragile contexts, prevention of violence and harmful practices against children, among others.

The choice of thematic areas for increased impact coverage will be driven by identified global and regional knowledge gaps with a particular focus on priority areas of UNICEF programming within the next quadrennial.

To address some of the key bottlenecks identified by UNICEF staff, such as a disconnect between programme planning cycle and impact evaluation planning, the EO will promote the inclusion of an IFA in major global and country donor proposals. IFA can include evidence gap mapping on 'what works' in the thematic area of focus (globally, regionally and at the national level), intervention mapping/assessment based on defined criteria and ToC data diagnostics, and recommendations on scoping, costing and evaluation implementation parameters.

Although a blanket approach of establishing a minimum budget threshold (e.g. US$10 million) requiring an IFA in the first year of implementation (and potentially an impact evaluation) might be hard to enforce and monitor, this is one way to boost systemic and consistent UNICEF effort for greater focus on outcome and impact-level results from the onset of programme planning. Institutionalization of such requirements will also allow greater transparency and accountability for donor investment in the long term.

The second critical entry point is preparation of the Costed Evaluation Plan (CEP) as a part of the country programme development process. This is the stage when strategic discussions on programmatic priorities, new initiatives and policy advocacy need to be supported with long-term planning for outcome and impact-level evidence. Similar to donor proposals, large country programmes can be required to include a rigorous impact evaluation on untested but strategic interventions within their CEPs.

Strategic Pillar II

Diversify methods and innovate

Objective: Demonstrate feasibility and validity of innovative, contextualized and cost-effective methods, data collection tools and design approaches to measure impact.

This stream of work will focus on:

1 **The Methods' Innovation Lab** to promote and expand applications of a wide range of new methods, tools and data sources for evaluations.

2 **The Evaluation Methods Helpdesk** under a UNICEF-wide Evidence Helpdesk will provide support to country and regional offices' needs in evaluation methods, impact evaluation design, data scoping, digital tools, etc.

The COVID-19 pandemic amplified the demand for flexible, rapid and 'light' evaluation approaches that could generate reliable evidence under shortened time frames and with minimum disruption for the programmatic activities. Recognizing some potential risks associated with 'light' or 'rapid' approaches,[22] there is a need to explore methodologically robust evaluative approaches that are tailored to different contexts while allowing improved cost-efficiency and timeliness of outcome and impact-level evidence. In close collaboration with other UNICEF divisions, the Methods' Innovation Lab will test validated but under-utilized sampling and data collection approaches as part of planned impact evaluations. This includes working with secondary data sources: (a) administrative data; (b) multiple indicator cluster surveys (MICS-plus) and other household data; (c) remote high frequency data; (d) spatial data; and (e) big data. The potential of these data sources in counterfactual designs will be explored.

In addition, to support the 'supply side' of impact evidence generation, the EO will provide quality assurance, collect and systematize resources on specialized services and expertise in the priority areas of the Strategy for direct use by UNICEF country offices. This can include setting up a global consultancy roster for non-experimental (theory-based) methods, partnership agreements with academic institutions as well as institutional long-term agreements (LTAs) in impact evaluation.

Launching a methods helpdesk for the use of country and regional offices will strengthen technical exchange and advisory support on all types of impact-related methods and provide feedback mechanisms across the evaluation function on emerging needs as well as successful applications of evaluation practices.

The Method's Innovation Lab will test validated but under-utilized appoaches including the use of secondary data sources and big data for counterfactual designs.

Strategic Pillar III

Improve learning and programmatic synergy

Objective: Strengthen institutional, staff and partners' capacity to plan, manage and use impact evidence through better understanding of the purpose of and requirements for evaluations of impact.

This stream of work will focus on:

1   **Capacity development** of UNICEF staff and national partners has to be integrated into ongoing initiatives such as a global evaluation training programme, United Nations Evaluation Group (UNEG) and partners' learning initiatives, planned evaluability/impact feasibility assessments or annual programme review process.[23] These will focus on formulating the role of evaluation, specific purpose of impact evaluation, choice of relevant causal questions, appropriate and feasible methods.

2   **Methodological briefs and guidance** on a range of non-experimental, theory-based methods to evaluate causal relationships developed in partnerships with academic institutions and UN partners.

3   **Facilitation of institutional and global learning** through active dissemination of UNICEF impact evidence across internal and external platforms including peer-reviewed journals, promotion of 'best practice' exchanges on methods and innovation across the UNICEF evaluation function, the UNEG community and other partners.

4   **Promote upgrading of technical skills** and incentivize ongoing learning (jointly with the Division of Human Resources).

Programme staff's understanding of the benefits of impact evidence, and their awareness of its role to test programmatic assumptions directly, affects staff capacity to develop a solid ToC and initiate evaluation of impact. The former is a pre-condition for the synergy between evaluation of impact and the programme implementation process. The Strategy will support learning integrated into the scoping or evaluability phase of evaluation planning to promote co-design of impact evaluation questions and methodological approaches with programme staff and national partners.

Developing organizational capacity will be a continuous process implemented through the online or face-to-face events which bring UNICEF staff and national partners together using established communication platforms at the national or regional levels. Global dissemination of UNICEF impact evidence is critical to support global learning on the effectiveness of UNICEF interventions and raise the credibility of UNICEF evaluative evidence with national partners, donors and other stakeholders.

Table 1 in annex 1 provides a summary of the proposed Strategy model (ToC). A more detailed action plan for each Pillar with suggested relevant global indicators is presented in the Action Framework (annex 2). These will be further refined at the regional level based on knowledge related to specific gaps, needs and resources within the relevant region. Specific Key Performance Indicators (KPIs) will be developed in due course to monitor progress.

Key message: Three strategic pillars of the evaluation of impact area of work respond to the most persistent and interconnected challenges to initiate, implement and use rigorous impact evidence within the institutional realities of UNICEF work. They target the demand and supply sides of the evidence generation process, promoting a close partnership between evaluation community, academic institutions, national partners and UNICEF programmes.

Figure 8. Linking strategic pillars with the most common challenges for planning, managing and using impact evaluations

---
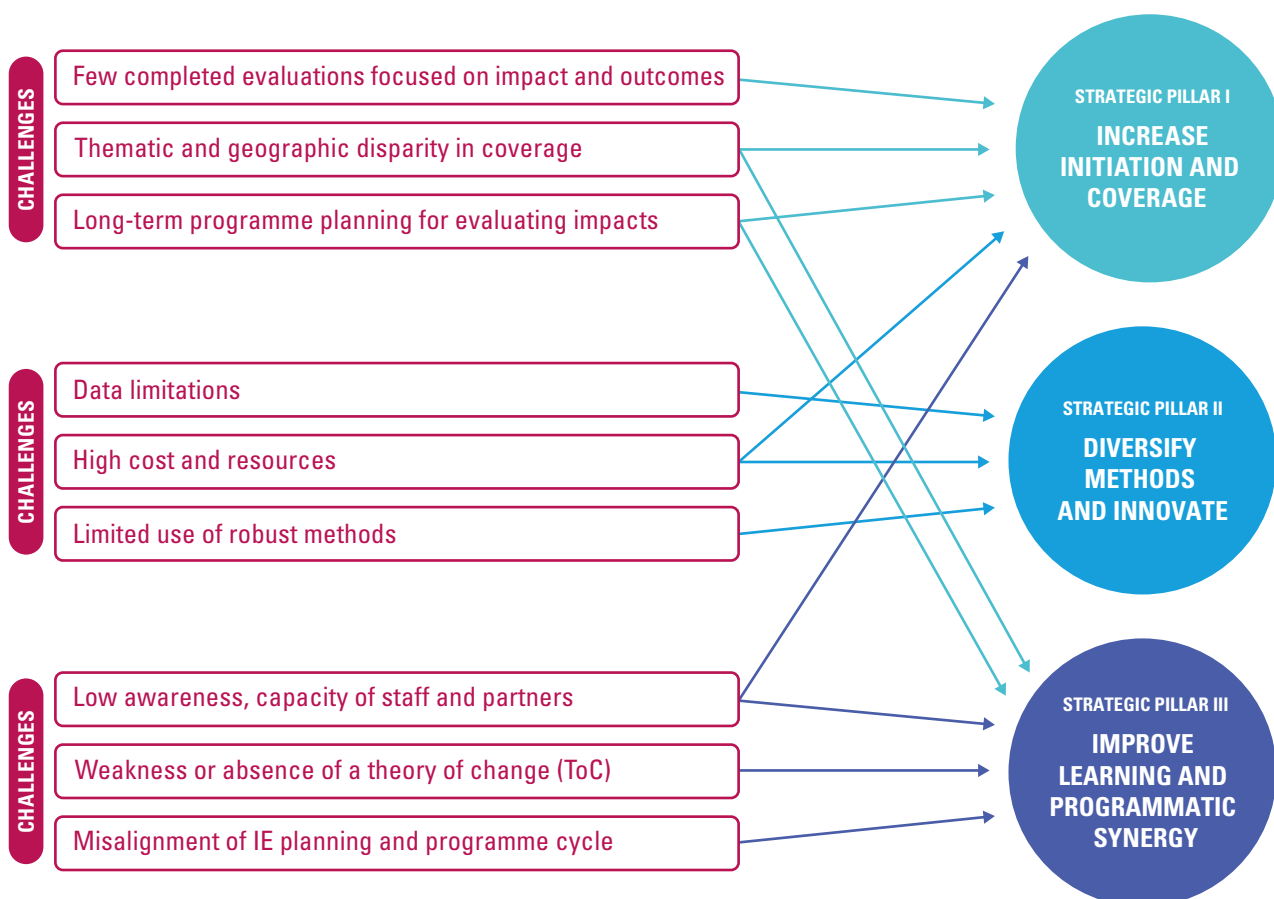
21 UNICEF, internal document 'Vision for the Evaluation Function' on UNICEF SharePoint.

22 Short time frames can be detrimental for participatory approaches, ability to collect data from the most vulnerable and hard to reach populations, and establishing robust ethics protocols. Some strategic and rigorous evaluative exercises, such as longitudinal impact evaluations, can prove to be inappropriate under a shortened programmatic and evidence time frame.

5

Methods and Process

In real-life scenarios of UNICEF programmatic work, getting answers to the questions 'Does it work?' or 'To what extent?' is rarely sufficient. Policymakers, programme implementers and donors are equally interested to know 'Why does it work?', 'Who benefits most?' and 'At what cost?'. Mixed methods and nested impact evaluation designs enable us to generate the most contextualized impact evidence.

Mixed methods, nested evaluation designs

Programme stakeholders ask a range of evaluation questions, including on cost-efficiency of interventions. The latter is particularly salient for any public policy decisions concerning programme scale-up and reflects donors' consistent concern for 'value for money'.[23] Not only do these questions help to contextualize the findings and strengthen analytical narrative, they also support concrete action-oriented recommendations addressing often overlooked unintended results of the programme[24] and its cost-effectiveness.

Complementary evaluation questions that deepen understanding of the mechanisms behind change require applications of different analytical approaches outside traditional experimental and quasi-experimental designs. Today, there is growing interest from development actors and users of impact evidence to build on complementarity of methods, without necessarily sacrificing the rigour and internal validity of each methodological approach.

RCTs remain the most rigorous impact evaluation approach and should be selected in situations when evidence can be generated in an ethical way, randomization is feasible and appropriate for the scale of the programme, and its utilization is clear to all stakeholders.

The Strategy follows principles previously outlined by the Network of Networks on Impact Evaluation (NONIE) guidance[25] on impact evaluation, indicating that:

- ▶ **No single method is best** for addressing the variety of important programmatic questions, but not all questions can be answered under a single type of evaluation. This, for instance, means that impact evaluations complement other types of policy, strategy, system and process evaluations.

- ▶ **Comparative advantage of the method** in application to the specific question should be considered for the choice of the methods or analytical approaches.

- ▶ **Complementarity of methods** ensures a more multi-dimensional picture of impact.

Based on these foundational principles, UNICEF will promote 'purpose-led' or 'question-led' methodology that utilizes the range of methods available to answer relevant causal questions grounded in the dichotomy of attribution vis-à-vis contribution. As discussed in section 2, the choice of core method will be based on comparative advantage of methodology to assessing either attribution or contribution under a specific programmatic context (see figure 2 in section 2).

Experimental (RCT) and quasi-experimental methods are required for the construct of a credible counterfactual to establish an attribution of the results to the intervention.

RCTs remain the most rigorous impact evaluation approach and should be selected in situations when evidence can be generated in an ethical way, randomization is feasible and appropriate for the scale of the programme, and its utilization is clear to all stakeholders. Experimental and quasi-experimental designs can be complemented by qualitative and quantitative methods and techniques that answer 'Why?', 'How?' and cost-effectiveness questions, and that allow contextualization of findings. Such mixed-methods or multi-method designs allow a combination of pre-determined (e.g. RCT, RDD etc.) and emerging methods (e.g. process tracing, geospatial analysis), open- and closed-end questions, statistical and text analysis, interpretations across databases,[26] and have the advantage of generating the most contextualized impact evidence that provides good value for money.

Nested evaluation designs which combine a strategic focus on outcome and long-term change with process-related questions will be promoted. Arguably, this approach helps to meet demand for rapid evidence in a fast-moving context without losing sight of strategic evidence needs that require time and long-term effort. More extensive descriptions of impact evaluation methods and the additional requirements specific to each method can be found elsewhere.[27] Some recent examples of UNICEF impact evaluations are presented in text box 2.

Fostering programme alignment and policy uptake

Early integration of evaluative thinking into programme planning and intervention design is a critical condition to achieve rigour and ensure relevance of impact evidence for policy needs. Early and meaningful engagement with policymakers and programme end users also foster take-up of rigorous impact evidence (e.g. decisions on scaling up) because it builds credibility of the process. Yet currently, these are some of the most persistent institutional challenges this Strategy has to address.

Understanding the end user's (programme staff as well as national partners) needs and interests in impact evidence within a broader evidence package of formative research, monitoring data, lessons learned and other evaluative assessments, provides an entry point for the engagement and helps to build trust with key stakeholders.

This Strategy outlines four specific interconnected elements, which, if reinforced and linked together, can facilitate progress on all three strategic pillars (figure 9):

Figure 9. Engagement with policy stakeholders for credibility and utilization

PLAN	SCOPING	MATCH	BUILD CAPACITY
Refinement of a testable ToC in consultation **with key stakeholders**	**Understanding the evidence** needs of key stakeholders	Evaluation questions must be matched to appropriate design and methods through **impact feasibility assessment** in consultation with stakeholders	Involving **national universities and experts** in the analytical process and implementation

Planning

Refinement of a testable ToC in consultation with key stakeholders is done alongside detailed consideration of the type, scope and characteristics of the intervention package to be evaluated. This should be closely aligned with any discussions on the choice of evaluation questions, design and methods as well as 'suitability' of the intervention for rigorous evaluative testing. Early planning with stakeholders ensures validation of key outcome indicators along the causality 'pathway', verification of relevance of the evaluative exercise to policy needs (through identified evaluation questions), and greater ownership of the whole process.

Scoping

Understanding the evidence needs of key stakeholders can be done through impact scoping or impact feasibility assessment. The latter examines opportunities, limitations and requirements for measuring outcomes and impacts in specific programmatic contexts. This involves assessing feasibility of estimating the observed change and attributing it to the programme through a robust counterfactual informed by ethics norms and considerations. If making an attribution claim is not feasible (e.g. evaluation is summative) or not a priority, we consider availability, quality and comprehensiveness of evidence to support theory-based causal analysis. In both cases, the options are discussed with relevant stakeholders with reference to resource and time constraints, timing of critical decision-making and other relevant factors.

Matching

Evaluation questions must be matched to appropriate design and methods through impact feasibility assessment in close consultation with stakeholders.[28] Consideration is given to the comparative advantages of experimental, quasi-experimental and non-experimental designs to answer different types of selected evaluation questions given their methodological and data requirements. The final choice is likely to be made based on strategic priority given to rigorous impact evidence by different stakeholders at the global, regional and national levels as well as careful weighting of all factors against operational characteristics of the programme.

Building Capacity

Developing capacity of partners has to be integrated into the whole process and supported through stronger ties with national academic and research institutions. Involving national experts in the analytical process and implementation helps to establish trust between evaluation teams and key users such as national governments, and to ensure that findings are contextualized and actionable recommendations are linked to the realities of policy process and fiscal sustainability.

Key message: UNICEF promotes a combination of methods driven by multi-faceted evidence needs of complex development contexts. RCTs remain the most rigorous impact evaluation approach and should be selected in situations when evidence can be generated in an ethical way, randomization is feasible and appropriate for the scale of the programme and its utilization is clear to all stakeholders. Any experimental and quasi-experimental methods can be complemented by other qualitative and quantitative methods and techniques to contextualize findings and help answer 'Why?', 'How?', 'At what cost?' questions. This Strategy promotes nested evaluation designs which combine a strategic focus on outcome and long-term change with process-related evaluative evidence. Any impact-focused evaluation design has to be embedded in continuous and iterative engagement with national partners and key stakeholders to ensure credibility and facilitate policy uptake.

23 For definitions accepted by the United Kingdom government see Department for International Development (2011). DFID's Approach to Value for Money (VfM) (publishing.service.gov.uk).

24 Bamberger, M., Tarsilla, M. and Hesse-Biber, S. (2016). 'Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute', Evaluation and Program Planning, 55, pp. 155–162.

25 The following text is adapted from: Leeuw, F. and Vaessen, J. (2009). 'Impact Evaluations and Development. NONIE Guidance on Impact Evaluation', NONIE, Washington DC.

26 For more in-depth discussion on mixed-methods designs see Creswell, J.W. and Clark, V.P. (2018). Designing and Conducting Mixed Methods Research, 3rd ed. SAGE Publications.

27 For example, Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M.J. (2016). Impact Evaluation in Practice, 2nd ed. International Bank for Reconstruction and Development/The World Bank.

28 See question typology in Befani, B. (2020). 'Choosing Appropriate Evaluation Methods – A Tool for Assessment and Selection' (Version Two), Centre for the Evaluation of Complexity Across the Nexus/CECAN.

Text box 2

Examples of UNICEF evaluations

Impact Evaluation Embedded in OECD/DAC Criteria of Impact and Effectiveness

Impact Evaluation of Volunteer Community Mobilizer Network on Polio Eradication in Nigeria (2021)

Goal Area 1
Every child survives
and thrives

The Volunteer Community Mobilizer (VCM) network programme was launched in 2012 in 12 of the high polio-risk states in the northern part of Nigeria. The impact evaluation focused primarily on immunization-related variables, specifically on polio, routine vaccination and barriers to immunization. Unintended effects of the programme were also examined as secondary indicators. The impact evaluation drew on the OECD/DAC criteria of impact and effectiveness and provided an overview of the cost-effectiveness of the VCM, under the efficiency criterion. Impact and outcome indicators were examined using the propensity score matching technique (PSM). PSM overcomes the selection bias of non-random programme placement by using statistical techniques to construct an artificial comparison group. In addition to using PSM with primary data, the evaluation used secondary historical data to calculate the impact of the programme on the reduction in wild poliovirus (WPV) cases. Here, a panel event study approach was followed. In addition to the primary focus on impacts and outcomes, the evaluation assessed programme relevance and sustainability using qualitative methods.

Impact Assessment Using Theory-Based Approach and Quasi-Experimental Design

Summative Evaluation of the Accelerated Community Actions for Reducing Stunting in Zimbabwe (2021)

Goal Area 1
Every child survives
and thrives

This summative evaluation in Zimbabwe was designed to generate evidence on programme results in improving food and nutrition security and increasing household resilience to repeated shocks. An assessment of all programme activities (at national, subnational, district and community level) was implemented from January 2017 to September 2020. The methodological design took a hybrid approach, combining contribution analysis as a theory-based approach, quasi-experimental design (before and after comparison) and a participatory approach of extensive consultations with programme stakeholders. This was implemented to overcome any method-related deficiencies and gather rich and complementary information for cross validation and triangulation.

Experimental Impact Evaluation in a Large-Scale Project

Effectiveness and impact evaluation of a WASH in schools intervention in Laos, 2014–2017:
A randomized controlled trial

Goal Area 4
Every child lives in
a safe and clean
environment

In Lao PDR, UNICEF contributed to the 'Laos Basic Education Water, Sanitation and Hygiene Programme', with the aim of increasing school attendance by strengthening water and sanitation services. An RCT of the programme was conducted between 2014 and 2017 in randomly selected schools. While the trial found the intervention had minimal impact on diarrhoea prevalence, it did not achieve its main objective of increasing attendance, even in schools with the highest adherence to the programme.

Mixed-Methods, Experimental Impact Evaluation

Impact Evaluation of the Accelerated School Readiness Programme in Mozambique (2019)

Goal Area 2
Every child learns

UNICEF Mozambique and Save the Children, in partnership with the Mozambique Ministry of Education and Development, implemented an Accelerated School Readiness (ASR) pilot programme from 2016 to 2019. Impact evaluation had a mixed-methods, experimental (RCT) design aiming to determine the impact of the ASR programme on children's school readiness, on-time enrolment, and academic achievement in Grade 1 of primary school relative to comparable children in communities with no pre-primary education. An Early Learning Assessment was used to evaluate the children in six outcome areas. The evaluation also measured cost-effectiveness using the community- and child-level costs of providing the ASR pilot programme. Qualitative methods were applied to identify which aspects of community context and implementation process seemed to facilitate or inhibit the success of the ASR pilot.

RCT of a Multi-Component Programme

End of Programme Evaluation IKEA - II - Improving Adolescents Lives in Pakistan

Goal Area 3
Every child is protected from violence and exploitation

The UNICEF Improving Adolescents Lives in Pakistan (IALP) programme, implemented between 2015–2019, represents a clear example of a robustly evaluated, multi-component, and successful child protection programme. Using an RCT which included 446 treated and 207 control communities, evaluators were able to assess impact of both the programme components, a Child Protection intervention aimed at increasing knowledge and education, and the establishment of Non-Formal Basic Education (NFBE) Centers. Results show that IALP has a positive impact on child marriage reduction and the NFBE Centers increase school attendance (especially for girls) and reading skills. Awareness campaigns directed at men and boys also significantly changed attitudes toward child marriage for the better.

Impact Evaluation in a Humanitarian Context

Impact Evaluation of Fiavota Emergency and Recovery Cash Transfer in Madagascar (2018) – quasi-experimental design to assess effects on drought-affected families with children.

Goal Area 5
Every child has a fair chance

Due to the El Nino weather phenomenon causing severe drought in southern Madagascar, in September 2016 the Malagasy government declared a state of emergency. To address acute food insecurity of drought-affected households with children under the age of 5 in southern Madagascar, UNICEF Madagascar partnered with the World Bank and Ministry of Population, Social Protection, and Support of Women (MPPSPF) to implement an emergency cash transfer (Fiavota) in the five southern districts of the Toliara Province. The impact evaluation was intended to help policymakers to understand how best to address the acute needs of the affected populations during the emergency response phase of Fiavota, from December 2016 to September 2017. The evaluation used quasi-experimental design (propensity score matching) techniques to create treatment and comparison groups. It found consistent positive impacts on food security and increased consumption. In contrast, no impacts were found on agricultural investment or children's malnourishment status (anthropometric measurements). The qualitative component of the evaluation showed that overall the programme was implemented as planned.

6

Implementation

Translating strategic pillars into actions relies on adequate resourcing, initiative, persistence, and a long-term vision of evaluation champions (see annex 1 for the proposed action framework). Given the decentralized nature of evaluation work at UNICEF and variations in capacity level and context, some adaptation of approaches and their timelines will help to ensure programmatic relevance within this guiding framework.

The three complementary and mutually reinforcing strategic pillars outlined above aim to improve generation of rigorous evaluative evidence on UNICEF programmes in cost-effective and innovative ways. This requires adequate resourcing at all levels of the evaluation function, supported by effective communication and advocacy.

Human and Financial Resources

The Strategy will require sufficient staff with skills and capacity to initiate and lead the technical aspects of its implementation. The Senior Evaluation Advisor (Methods and Impact), supported by the Innovation Specialist, will be complemented with additional human resources in line with implementation progress. Each strategic pillar will require additional core staff to support key areas of work. Expansion of strategic thematic windows under the Impact Catalyst Fund will require in-house impact evaluation capacity to support multi-country approaches to ensure consistency, sustainability and high quality. The EO will review staffing needs on an annual basis in line with the decentralized nature of the function and will designate additional resources to strengthen regional evaluation staff capacity to implement the Strategy.

The Capacity-Building Team of the EO will lead on implementation of the capacity development component of Strategic Pillar 3.

The EO will review staffing needs on an annual basis in line with the decentralized nature of the function and will designate additional resources to strengthen regional evaluation staff capacity to implement the Strategy.

Additional specialized technical expertise in impact evaluation will be sought to support specific pilot projects and analysis developed either under the Impact Catalyst Fund or Data and Methods Lab. At the regional level, implementation of the Strategy's three pillars will be managed by UNICEF regional evaluation teams led by Regional Evaluation Advisors in collaboration with country office programme staff and evaluation focal points and in close collaboration with programmes.

Implementation of the actions foreseen under the three pillars of the Strategy will require pooling financial resources from three main internal and external sources:

a. **country-led allocations** as part of regular resources or thematic funds;

b. **a percentage of the global evaluation pooled fund** earmarked for the development of new areas of strategic importance; and

c. **donor and government funding** for either EO, regional, or country-led evaluations.

Diversification of funding sources and their complementarity are critical to reduce the risk of limited coverage of impact evidence driven by single regions, the size of country programmes and/or identified donor preferences. Presented below is the overall budget envelope that would be required to achieve tangible outcomes by the end of the proposed 2022–2025 timeline (see table 1).

Under Strategic Pillar 1, the Impact Catalyst Fund (ICF) is the main incentive vehicle to increase the impact evidence base in strategically important thematic areas of the UNICEF Strategic Plan 2022–2025. The ICF expansion will be gradual with around four impact evaluations supported with ICF matching grants in Years 1 and 2, five in Year 3, and six in Year 4 with the total number of ongoing projects reaching 19 by the end of 2025. The scale can be increased with additional donor support.

The cost estimates include fixed (human resources) and variable costs (data collection and associated work), and include, out of the total cost of the planned impact evaluations, only a matching share from the EO. The ICF will support a country office's own IE resources with a matching grant that can cover between 60 per cent and 80 per cent of the total estimated IE budget. The ICF matching share will depend on the country programme context (e.g. small vs large UNICEF country office), availability of complementary sources of funding, country's standing on the 1 per cent assigned to evaluation, and other factors, and will be decided on a case-by-case basis.

Strategic Pillar 2 will require funding to support tests and pilots at the decentralized level with the use of innovative methods and new data to evaluate outcomes under the Methods' Innovation Lab. Strategic Pillar 3 will require funds to scale up integrated capacity modules (external expertise) and support partnerships with academic and research institutions across the global North and South.

Table 1. Budget envelope to implement evaluation of impact strategy

STRATEGIC PILLAR	2022	2023	2024	2025	TOTAL
	USD million	USD million	USD million	USD million	USD million
Pillar 1: Increase Initiation and Coverage	2.8	2.8	3.6	4.4	13.6
Pillar 2: Diversify Methods and Innovate	0.5	0.5	0.5	0.5	2.0
Pillar 3: Improve Learning and Programme Alignment	0.3	0.3	0.3	0.3	1.2

Risk Mitigation

The Strategy may face several implementation risks:

▶ potentially slow improvement in the enabling programmatic structures and institutional processes that aim to support a robust evaluation at the impact and outcome levels;

▶ systemic low quality of monitoring data at the outcome and impact levels; and

▶ inadequate integration of the RBM and ToC mindset in every new project, programme and policy initiative.

The Strategy aims to advocate for improvements in these areas, but this will not happen without a wider organizational effort. To this end, the ongoing initiative across the organization to achieve a focus on the outcome level of the UNICEF Strategic Plan 2022–2025 is encouraging and is generally aligned with the framework of this Strategy. Another potential risk is underfunding and timeliness of funding to meet the Strategy's key resource requirements. UNICEF will have to commit the core resources required to deliver results on an annual basis, while also working to identify new external funding sources to support country offices in scaling up rigorous impact evidence.

Design and implementation of rigorous impact evidence is associated with high human and financial costs. These are due to the requirements for specialized technical expertise and cost of primary data collection over multiple data collection rounds spanning several years.

Partnership

UNICEF evaluation of impact work will build on current global practices and experiences of other UN agencies (e.g. World Food Programme)[29] and other international development institutions (e.g. the World Bank)[30] in institutionalizing impact evaluations to stimulate a results-oriented culture.
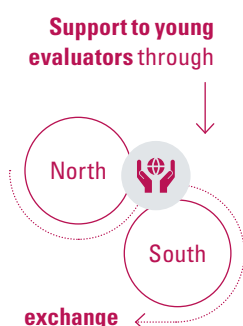
The EO, regional evaluation teams and evaluation focal points at the country level will expand existing, and form new, partnerships to build demand for rigorous impact evidence within the organization, and will purposefully plan the collection, processing and dissemination of findings and recommendations (figure 10).

Collaborations are to be built with other evidence functions including the Data, Analytics, Planning and Monitoring section (DAPM), the UNICEF Innocenti-Global Office of Research and Foresight, and the Office of Innovation teams on more strategic and complementary efforts in testing data and methods innovations, evidence synthesis, learning and capacity-building. Further consultations will be held with other Divisions concerned with evidence to find complementarity of technical expertise and synergies of joint efforts in taking the impact evidence agenda forward.

The UNEG community and regional UN evaluation networks provide excellent platforms for inter-agency learning and exchange. Specifically, the UNEG methods group will be utilized for consultations and promotion of both rigorous impact evaluation methods and qualitative approaches to causal inferences. Evaluations conducted through joint UN management and governance structures are also relevant for building a common inter-agency ground for strengthening impact evidence and its utilization. The work will build on the ongoing partnerships with the United Nations Sustainable Development Solutions Network, the Lancet COVID-19 Commission, and its Task Forces, the Harvard Humanitarian Initiative, and other academic institutions in the North and South focusing on bridging the expertise and supporting capacity of young evaluators in low- and middle-income countries.

Building connections with other organizations active in the area of impact evaluation, including 3ie, J-PAL Poverty Action Lab, Innovation for Poverty Action (IPA), Development Impact Evaluation (DIME), International Growth Centre (IGC), and others will also be actively explored.

Support to young evaluators through

North

South

exchange

Figure 10. UNICEF evaluation of impact partnership ecosystem
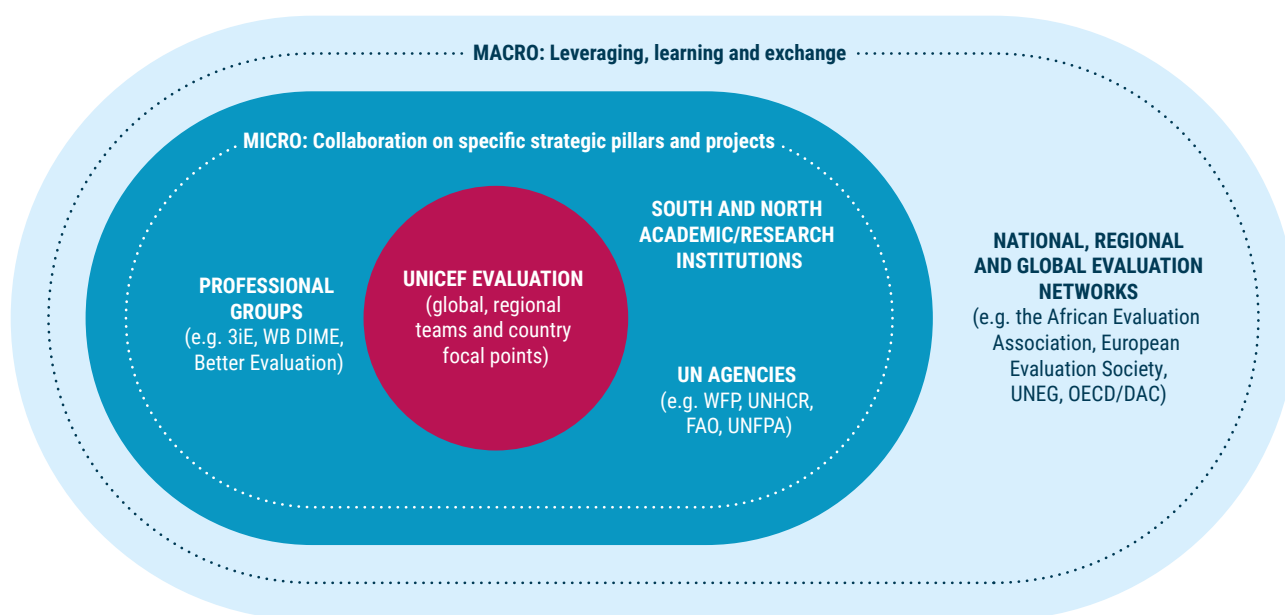
Abbreviations: 3iE = International Initiative for Impact Evaluation, DIME = Development Impact Evaluation, IPA = Innovation for Poverty Action, JPAL = Abdul Latif Jameel Poverty Action Lab, OECD/DAC = Organisation for Economic Co-Operation and Development/Development Assistance Committee, UNDP = United Nations Development Programme, UNEG = United Nations Evaluation Group, UNFPA = United Nations Population Fund, UNHCR = United Nations High Commissioner for Refugees, WB = the World Bank, WFP = World Food Programme

Oversight

The Evaluation of Impact Strategy and Action Framework is formulated in accordance with the UNICEF 2018 Evaluation Policy which sets the principles of the decentralized evaluation function within the organization. A revision of the current Evaluation Policy conducted through the peer-review process will be informed by the current Strategy and its underlying principles.

The Director of the EO will provide general oversight of the Strategy implementation and will report annually to Executive Board members on the implementation progress.

Within the EO the newly formed Methods, Impact and Learning section, led by a Senior Evaluation Specialist, will oversee the Strategy implementation at global level and provide technical assistance to the regional and country offices.

The team will work with HQ thematic managers to integrate rigorous outcome and impact components into global or multi-country evaluations. It will collaborate with the EO capacity-building team to validate global training content and approaches. Specific roles and responsibilities are defined in Action Framework presented in annex 2.

The EO team will continue to consult and engage the Evaluation of Impact internal Task Force members, who represent the global, regional and country levels of the evaluation function, on the implementation aspects of the Action Framework. Regional programme events, senior management meetings and international communities of practice will serve as important communication platforms to disseminate knowledge and build demand for UNICEF impact evaluations.

This Strategy is the result of UNICEF's commitment to better demonstrate the impact of resources allocated on behalf of governments and donors. That necessarily includes the resources applied to implement the Strategy itself.

We expect to perform an independent review at the end of the defined implementation timeline (2024–2025). To facilitate this exercise, the EO has conducted diagnostic mapping of the status quo of the impact evidence produced by UNICEF within the current Strategic Plan period of 2018–2021. It is expected that this analysis will serve as the baseline for review of the programme in the future.

Given the technical complexity of this evaluation area and the focus on methodological innovations, the EO will form the Technical Advisory Group, which will be nested in the broader inter-divisional evidence network of UNICEF and include experts from partner UN organizations, academic institutions and global impact evaluation network. The technical advisory board will meet annually providing feedback on the ongoing global or multi-country projects, advising on emerging methodological approaches and innovations while also critically assessing the EO approaches. It will help UNICEF EO to stay abreast of technical knowledge and keep connected with the global community of practice in impact evidence to achieve the SDGs for children.

---

29   WFP Impact Evaluation Strategy (2019–2026) | World Food Programme
30   Implementing Impact Evaluations at the World Bank: Guidance Note

---

7

Conclusions

The success of this Strategy is a vital stepping stone for the ambitious UNICEF agenda 2022–2025 to increase accountability for outcome-level results and learn about effectiveness of UNICEF strategic interventions.

The Strategy is aligned with the set vision of the UNICEF evaluation function for the new quadrennium and is supported by its highly decentralized structure and strong programmatic links at all levels of the organization.

Organizational and social return on investment in rigorous impact evidence for UNICEF will be high. Not only has impact evidence potential to influence political decisions to scale up the most effective interventions for children, but it can also save millions of dollars by correcting the course or abandoning approaches that do not work or could work better.

In the long term, it will contribute to an evidence culture across the organization that is focused on results, rather than aspirations. But the value of any evidence is in its use. We should not underestimate the effort, passion, persistence and patience required to build trust and credibility with national counterparts, beneficiaries and national partners during the evidence generation process.

The Theory of Change for the Evaluation of Impact Strategy

ACTIVITIES	OUTPUTS	OUTCOMES	VISION

STRATEGIC PILLAR I — INCREASE INITIATION AND COVERAGE

ACTIVITIES	OUTPUTS	OUTCOMES
Launch of Impact Evaluation Catalyst Fund (2022–2025)	Increased number of rigorous impact evaluations	Multi-country impact evaluations are planned and implemented based on identified strategic evidence needs
Technical assistance on the use of OECD/DAC impact criterion	Impact Feasibility Assessments are implemented for strategic global initiatives	Coverage expands to new thematic areas and geographic regions
Impact evaluation component is integrated into SOPs for major donor proposals	Increased number of summative evaluations with integrated OECD/DAC impact criterion using appropriate methods	Institutional documents reflect the requirements to evaluate impact
Impact evaluation is better integrated into CPD process (CEP)		

STRATEGIC PILLAR II — DIVERSIFY METHODS AND INNOVATE

ACTIVITIES	OUTPUTS	OUTCOMES
New tools and methods are promoted and piloted in several countries	New methods and tools to measure outcomes are piloted in humanitarian and fragile context	Quality of UNICEF evaluations of impact is increased through better utilization of robust and appropriate methods, cost-effective data collection tools and approaches
Evaluation Methods Helpdesk provides support on IE design, data scoping and innovative approaches	UNICEF evaluations utilize a wider range of robust methodologies and data sources	

STRATEGIC PILLAR III — IMPROVE LEARNING AND PROGRAMMATIC SYNERGY

ACTIVITIES	OUTPUTS	OUTCOMES
Learning on IE and methods are integrated into ongoing capacity initiatives and evaluation processes	National partners and UNICEF staff from seven regions receive trainings which include impact evaluation component	Strengthened capacity to plan, design, implement, manage and use impact evidence through better understanding of its role, purpose and technical requirements
Strengthened RBM training content	Knowledge and good practices are consolidated and shared regularly on digital platforms, internally and externally	Synthesis of impact evidence shows the unique contribution of UNICEF to results and effectiveness of its programmes
Contribute to the community of practices, internal and external exchanges on 'what works for children'		

VISION

BETTER NATIONAL POLICIES
Robust, timely and relevant impact evidence strengthens UNICEF advocacy and supports national partners' decisions to allocate more resources for child-focused policies and programmes

IMPROVED ORGANIZATIONAL EFFECTIVENESS
Impact evidence enables improved programming, demonstrates effectiveness and increases allocative efficiency of its child-focused investments

Action Framework

The non-exhaustive and non-prescriptive list of actions, expected results, roles and responsibilities below aims to stimulate actions at different levels of the evaluation function, providing opportunities for further initiatives at the decentralized organizational level.

Strategic Pillar 1 Actions	Expected result	Responsibility
1.1. Launch Impact Catalyst Fund (2022–2025) with an annual Request for Proposals (RFP) call to UNICEF country offices. The RFP will specify selection criteria, conditions for matching financial contribution and technical support package. The following criteria for selection of proposals might apply: (a) thematic relevance for global, regional and national strategic priorities; (b) application of innovative, untested approaches and intervention modalities; (c) knowledge gap on the topic and ability of the IE to address it in the specified outcome area; (d) potential for utilization (e.g. government commitment and interest); (e) technical feasibility for the use of rigorous impact evaluation methods; and (f) compliance with ethical standards. It is expected that each year at least four impact evaluation proposals will be chosen.	(a) By the end of the Strategic Plan period (2025), at least 14 rigorous IEs are supported at the country level as part of the Impact Catalyst Fund. (b) Results and recommendations of all completed IEs are discussed with national partners and disseminated within the global community of practice.	**Evaluation Office (EO):** Preparation of the concept note, RFP and fundraising; strategic dialog with global programme teams; identification of potential partners; launching the 'thematic window', coordination of the selection process, technical support and capacity development; global dissemination of results. **Regional Office (RO):** Consultation with regional programme teams to define programmatic priorities and knowledge gaps. Review and dissemination of the RFP through the regional evaluation network; assistance to CO on proposals with a view on regional priorities; technical support; and, wherever applicable, development of multi-country IE proposals. **Country Office (CO):** Response to the RFP. If successful, the CO evaluation focal point (with the technical support of the multi-country evaluation specialists or EO evaluation specialist) will manage the IE in close coordination with the programme staff at all stages of implementation and dissemination. If required, the evaluation focal point will have to receive technical induction on IE modalities.

1.2. Stimulate consistent focus on outcome and impact level results (wider use of the OECD/DAC impact criterion) in global, regional and country-led summative evaluations **including** in level 3 and level 2 humanitarian evaluations. (a) Methodological guidance and learning series on non-experimental (theory-based) methods to evaluate causal links; and (b) Revision of the GEROS quality assurance system to better capture the methodological rigour and robustness of causal claims for different evaluative products (impact evaluations and those that look at contribution or include 'impact' and 'effectiveness' criteria).	(a) The use of the impact criterion in global, regional and country-led summative evaluations is tracked in the UNICEF integrated evidence information systems for planning, monitoring, reporting and archiving (EISI); (b) At the country level, at least one thematic/goal area will include evaluation focusing on contribution to impact over a five-year period (as reflected in the Costed Evaluation Plan); and (c) Each inception report has a data analysis plan with specification of secondary data sources and methodology to establish causal contribution of the programme to the outcomes /impacts of interest.	**EO:** Technical guidance, capacity support and quality assurance in methodological approaches to evaluate impact. **RO:** Selection of appropriate topics and types of multi-country evaluations for the analysis under the impact criterion. Integration of the impact criterion into country programme evaluations in consultations with CO management. **CO:** Under the guidance of and with support from the RO, defining the ToC, results framework, and setting up robust monitoring systems.
1.3. Include impact evaluation requirement in the SOPs on donor proposals wherever appropriate.	All major global, regional and country-led programme donor proposals with a pilot programme of untested intervention should include rigorous impact evaluation/ impact assessment plans.	**EO:** Standardization of the IE donor proposal clause applicable to global and country-led programmes. Consultations with Programme Division and the audit office on global guidance and implementation. **RO:** Consultation with programme teams at the regional level, communication with CO management. Identifying strategically important pilots and interventions with major donor investments. **CO:** Programme coordination/programme effectiveness units of COs revise SOPs on evaluation and on donor proposals. Identifying IE opportunities and requirements across all programmatic areas at the Research, Evaluation and Studies Committees.
1.4. Strengthen results-based management (RBM) training content in terms of impact and outcomes accountability.	(a) The global guidance on RBM is updated; and (b) At least 70 per cent of CEPs include an evaluation exclusively focused on outcomes and impacts.	**EO:** Collaborative effort with the Programme Policy and Procedure Team on updating the content of RBM training. **RO:** In consultation with CO management, recommendations on appropriate representation of evaluation of impact in the CEPs.

Strategic Pillar 2 Actions	Expected result	Responsibility
2.1. Methods' Innovation Lab is set up at the EO level and supports regional evaluation teams to develop, test proof of concept projects and guide the diversification of sampling methods, data sources and approaches for use in evaluations of impact.	(a) By the end of the Strategic Plan period (2025), at least three impact evaluations or evaluative impact assessments commissioned or implemented by UNICEF at either global, regional or country level, generate robust impact evidence using secondary sources of data and test 'lean approaches'. (b) Long-term agreement for impact evaluations and consultancy roster in theory-based methods set up for global use to incentivize the wider use of methods for causal inference.	**EO:** Identify and test alternative approaches to baseline data and methods for evaluations of impact in close consultation with other UNICEF divisions, particularly Data, Analytics, Planning and Monitoring and Office of Research-Innocenti; develop technical guidance on the good practices and lessons learned. Select and set up a long-term agreement and consultancy roster. **RO:** Identify opportunities for testing and piloting proof of concept approaches at the regional or country level; facilitate communication with the COs; provide information and feedback on the regional priorities and areas of policy interest in innovative methods. **CO:** With the support of the RO, collaborate with national partners including National Statistical Offices and sectoral ministers on building support for strengthening data capacity and identifying opportunities for national data use in evaluation.
2.2. Evaluation Methods Helpdesk under a UNICEF-wide Evidence Helpdesk delivers on-demand support to country and regional offices on impact evaluations, new methods and tools.	At least 10 COs annually benefit from technical support provided through the online Evaluation Data and Methods helpline.	**EO:** Defines the scope and dedicates resources for the provision of remote on-demand support to COs; coordinates with ROs on potential areas of innovation in evaluation. **RO:** Stimulates interest in and demand for alternative and innovative methods for the evaluation of impact through regional evaluation network; facilitates regional learning and exchange promoting diversification of evaluation methods; encourages CO use of the helpline facility. **COs:** Use the helpline facility and participate in regional learning events sharing the country practices and innovations as well as providing support to other COs in the region.

Strategic Pillar 3 Actions	Expected result	Responsibility
3.1 Capacity development delivered through: (a) Interactive workshop will target key national stakeholders and UNICEF staff at the CO and RO levels during the scoping/ inception phase of the programme; and (b) Formal evaluation training includes impact methods module.	At least seven COs (one from each region) and four ROs annually benefit from the workshop on the selection of impact questions and methods and formal training delivered in partnership with academic institutions.	**EO:** Develops the content, structure, and delivery mode of impact trainings in consultation with partners and evaluation network members; provides targeted learning support to COs upon request. **RO:** Identifies capacity gaps at the country and regional levels; advises on regionally relevant topics, case studies and mode of delivery for the capacity development programmes; and disseminates relevant practices across the region. **COs:** respond to needs/capacity assessments implemented by ROs and EO, identifies capacity needs of national partners and facilitates their participation in the organized learning events.
3.2 Building a national cadre of evaluators with skills in causal analysis and impact evaluations jointly with other UN and national academic partners.	(a) At least 20 national evaluators benefit from the formal training delivered by academic partners; and (b) At least one national evaluator is included as co-investigator in the evaluation project focused on impact.	**CO/RO/HQ:** TOR includes a requirement to have a national evaluator as a co-Principal Investigator.
3.3 Learning and dissemination of impact evidence and methodological lessons generated through the Impact Catalyst Fund and other evaluations of impact.	Impact evidence and methodological lessons are regularly disseminated and promoted through internal and external platforms (webinars, evaluation briefs, synthesis of 'what works', professional conferences, peer reviewed journals, etc.) to influence policy and UNICEF development effectiveness.	**EO:** Utilizes global evaluation and other professional networks, academic partnerships and forums to promote the use of UNICEF impact evidence, and its policy/programmatic use. Promote academic publications and feature the most influential impact evidence. **RO:** Promotes and uses the most appropriate internal and external communication and professional network forums to share experiences of COs on generating and using impact evidence. **COs:** Share regularly best practices and lessons learned on policy uptake of impact evidence with a wider evaluation community.

Key Findings from the Diagnostics Mapping of UNICEF Evaluative Products 2017–2021

Impact evaluations
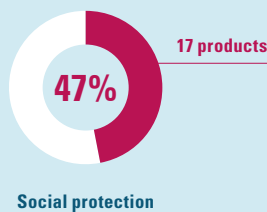
Scale and coverage

Evaluative products

Only

36/627

were identified as impact evaluations

6%

2017–2021

Geography of impact evaluations

16 products

44%

Eastern and Southern Africa (ESA)

1 product

3%

Europe and Central Asia (ECA)

Great disparity in thematic coverage

Highest coverage

17 products

47%

Social protection

Lowest coverage

1 product each

3%

Nutrition – Community resilience Adolescent programming

Trends

Downward trend since the peak in 2018

Total: 36

2017	2018	2019	2020	2021
9	14	2	7	4

Methods

Methods used

42% RCTs

58% Quasi-experimental

19% PSM+DiD

14% RDD

10% DiD

43% PSM

Quasi-experimental
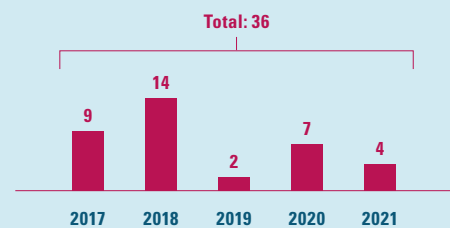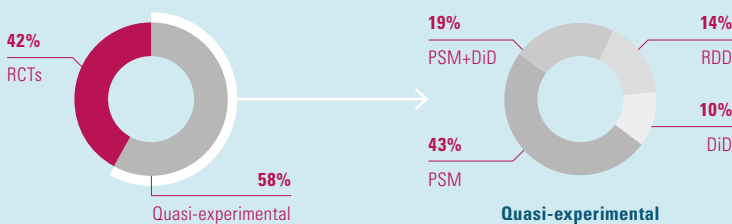
Mixed methods

Quantitative and qualitative methodologies used

6 products Non-mixed

30 products Mixed methods

Quality aspects

Theories of change

Only

28/36

impact evaluations with theories of change (ToC)

78%
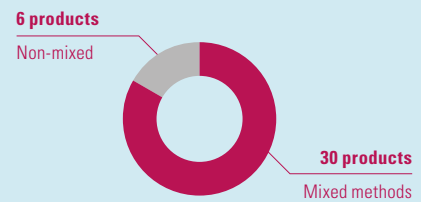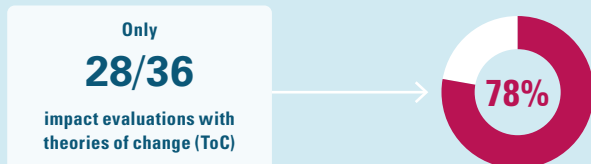
Missing ToC is correlated with the classification of products in EISI (**7 products being either research or studies** and **only one evaluation**)

Data are based on January 2022. The total number of evaluations (627) includes both evaluations and other evaluative products which were classified in the EISI database as research and studies' but match the operational definition of impact evaluation.

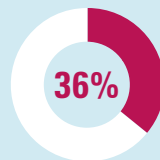# Evaluative products using OECD/DAC impact criterion

Scale and coverage

OECD/DAC impact criterion

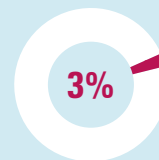Only
86/627
utilized the OECD/DAC impact criterion

→ **14%**

2017–2021

Geography of OECD/DAC impact criterion

36% — 31 products
ECA region
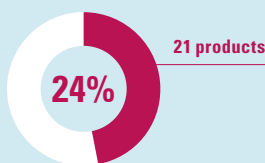
3% — 3 products each
HQ and the LAC region

Great disparity in thematic coverage

Child Protection

24% — 21 products

Systems strengthening, peacebuilding, humanitarian response, sports and culture, national strategy for children's rights, child-sensitive planning, monitoring and evaluation, child rights monitoring, combination of health, WASH, and child protection

Country offices

Only **seven** evaluations with the impact criterion were conducted in emergency settings

Level 2   ESA

Level 3   1 SA, 1 HQ, 2 WCA

Methods

Methods used

63 products
did not apply methods suitable to explore causality questions

15 products
Theory-based methods

6 products mixed approaches
2 products quasi-experimental designs

Non-experimental theory-based designs

5 contribution analysis

4 realist evaluation

2 micro-simulation modelling

1 qualitative comparative analysis, most significant change, outcome mapping, outcome harvesting, systems thinking approach, comparative case study approach

Mixed methods

Several evaluations utilized hybrid approaches which combined quasi-experimental design (DiD or before-after design) with non-experimental (contribution analysis, outcome harvesting, theory-based and systems approach) or qualitative methods.

Quality aspects

Theories of change

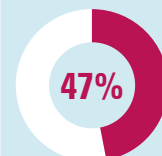**81/86**
evaluative products applying impact criteria have a ToC

→ **94%**
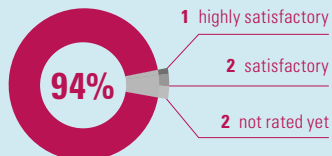
1 highly satisfactory
2 satisfactory
2 not rated yet

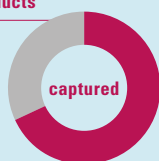Evaluations missing ToC rating

47%
47% of evaluations are consistent in their questions, analysis, discussion of findings and recommendations with regard to identified 'impact', or 'outcome', or 'output'-level change
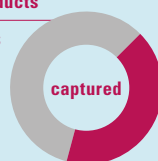
'Impact' or 'Partial Impact'

22 products
findings

captured

47 products
evaluative questions

Outcomes

35 products
findings

captured

25 products
evaluative questions

3 tried to capture 'outputs' rather than 'outcomes' or 'impact' under the OECD/DAC criterion of 'impact'.

18 discussed 'outputs' although the evaluation questions aimed to capture 'outcome' or 'impact' level, based on the ToC.

3 merged 'impact' with 'sustainability' criteria in the findings (only one report merged questions for both criteria).
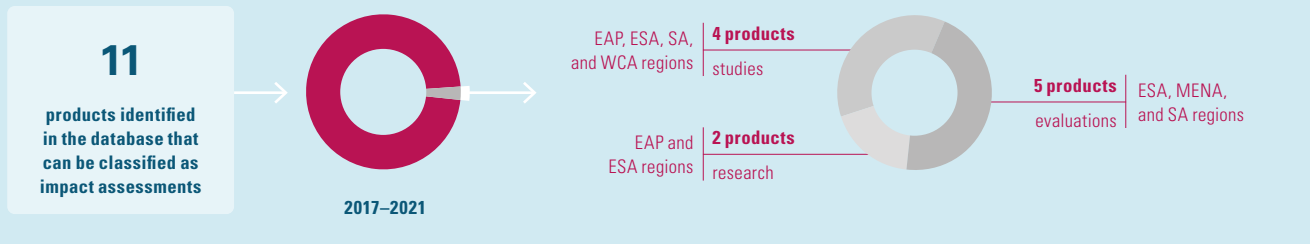
1 merged 'impact' with 'effectiveness' criteria both in questions and findings.

Ex post impact assessment
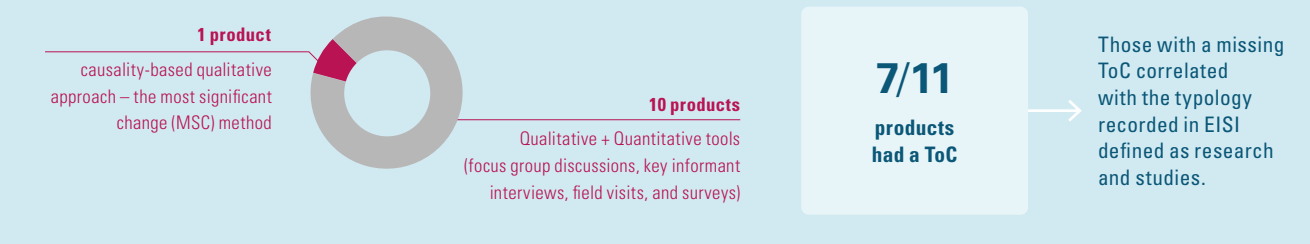
Scale and thematic coverage

Evaluative products

11

products identified in the database that can be classified as impact assessments

2017–2021

EAP, ESA, SA, and WCA regions | **4 products** studies

EAP and ESA regions | **2 products** research

5 products evaluations | ESA, MENA, and SA regions

Great disparity in thematic coverage

Highest coverage

WASH 3 products

Child protection 2 products

Lowest coverage

Toursim, parenting, nutrition, education/ learning, agriculture, adolescent programming

1 product each

1/11

conducted in a level 2 emergency setting (in ESAR) in 2017.

Methods

Methods used

1 product

causality-based qualitative approach – the most significant change (MSC) method

10 products

Qualitative + Quantitative tools (focus group discussions, key informant interviews, field visits, and surveys)

7/11

products had a ToC

Those with a missing ToC correlated with the typology recorded in EISI defined as research and studies.

Further Resources

Experimental and quasi-experimental designs

▶ Gertler, P., Martinez, S., Rawlings, L.B., Premand, P., Vermeersch, C.M.J. (2016). *Impact Evaluation in Practice*, 2nd ed., Washington, DC: Inter-American Development Bank and World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=y (Covers: RCT, instrumental variables, regression discontinuity design, difference-in-difference, propensity score matching.)

▶ Leeuw, F, Vaessen, J., (2009). *Impact Evaluations and Development. NONIE Guidance on Impact Evaluation, The Network of Networks on Impact Evaluation*. https://www.oecd.org/development/evaluation/dcdndep/47466906.pdf (RCT, pipeline approach, propensity score matching, difference-in-difference, judgmental matching regression-based techniques, regression discontinuity design; non-quantitative techniques – participatory approaches; mixed-methods approach (mostly focusing on qualitative and quantative data))

▶ UNICEF (2014). 'Quasi-Experimental Design and Methods', *Methodological Briefs*, Impact Evaluation No. 8. https://www.unicef-irc.org/publications/753-quasi-experimental-design-and-methods-methodological-briefs-impact-evaluation-no.html

▶ UNICEF (2014). 'Randomized Controlled Trials (RCTs)', *Methodological Briefs*, Impact Evaluation No. 7. https://www.unicef-irc.org/publications/752-randomized-controlled-trials-rcts-methodological-briefs-impact-evaluation-no-7.html

Non-experimental design

▶ Bath Social Development Research Ltd: Qualitative Impact Assessment Protocol QuiP [selection of papers and reports].

▶ Mayne, J. (2008). 'Contribution Analysis: An Approach to Exploring Cause and Effect'. The Institutional Learning and Change (ILAC) Initiative. https://nonprofitbuilder.org/storage/377/Contribution-analysis-An-approach-to-exploring-cause-and-effect-ILAC.pdf

▶ Serrat, O. (2017). 'Knowledge Solutions Tools, Methods, and Approaches to Drive Organizational Performance', Asian Development Bank. https://www.adb.org/publications/knowledge-solutions-organizational-performance (Outcome mapping, most significant change, social network analysis)

▶ Centre for Development Impact (2015). 'Applying Process Tracing in Five Steps'. https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/5997/CDIPracticePaper_10_Annex.pdf;jsessionid=CE0D6DBFAD41B99D3E24335484813FBA?sequence=2

▶ UNICEF (2014). 'Comparative Case Studies', *Methodological Briefs*, Impact Evaluation No. 9. https://www.unicef-irc.org/publications/754-comparative-case-studies-methodological-briefs-impact-evaluation-no-9.html

unicef

for every child

For further information, please contact:

UNICEF
Evaluation Office
3 United Nations Plaza
New York, NY 10017
USA

www.unicef.org/evaluation

UNICEF-Evaluation

evalhelp@unicef.org