

Comparative study of the WHO evaluation function with selected UN entities

Report



WHO/DGO/EVL/2024

© World Health Organization 2024

All rights reserved.

This health information product is intended for a restricted audience only. It may not be reviewed, abstracted, quoted, reproduced, transmitted, distributed, translated or adapted, in part or in whole, in any form or by any means.

The designations employed and the presentation of the material in this health information product do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted; the names of proprietary products are distinguished by initial capital letters.

The World Health Organization does not warrant that the information contained in this health information product is complete and correct and shall not be liable for any damages incurred as a result of its use.

The purpose of publishing evaluation reports produced by the WHO Evaluation Office is to fulfil a corporate commitment to transparency through the publication of all completed evaluations. The reports are designed to stimulate a free exchange of ideas among those interested in the topic and to assure those supporting the work of WHO that it rigorously examines its strategies, results and overall effectiveness.

This is an independent evaluation published by the WHO Evaluation Office, based on the work done by the independent evaluation team. This publication does not necessarily reflect the views of the World Health Organization. The text has not been edited to official publication standards and WHO accepts no responsibility for error. The designations in this publication do not imply any opinion on the legal status of any country or territory, or of its authorities, or the delimitation of frontiers.

Table of Contents

Acronyms	2
Executive summary	3
Introduction	8
WHO context	13
Key findings	14
Overview of UN agency comparators	26
Best practices from comparator agencies	34
Conclusions and recommendations	36
Annex 1: Terms of Reference	40
Annex 2: Interviews	42
Annex 3: Evaluation functions of UN comparators	44
Annex 4: Bibliography	78
 List of tables	
Table 1: Key Features of Comparator Evaluation Functions	3
Table 2: Enquiry Questions	9
 List of figures	
Figure 1 Summary of key steps in developing the evaluation function	12
Figure 2 Evaluation organigram	18

Acronyms

ADG	Assistant Director-General	IEO	Independent Evaluation Office (UNDP)
AFRO	Africa Regional Office	ILO	International Labour Organization
AIDA	Artificial Intelligence for Development Analytics	IOS	Internal Oversight Services
AMSTG	Agile Member States Task Group	JIU	UN Joint Inspection Unit
CEQAS	Centralized Evaluation Quality Assurance System (WFP)	JPO	Junior Professional Officer
CPD	Country Programme Document	LTA	Long Term Agreement
CPE	Country Programme Evaluation	MOPAN	Multilateral Performance Assessment Network
CRM	Compliance and Risk Management and Ethics	OECD	Organisation for Economic Co-operation and Development
CSP	Country Strategy and Programme	OED	FAO's Office of Evaluation
DAC	Development Assistance Committee	OEV	Office of Evaluation
DEQAS	Decentralized Evaluation Quality Assurance System (WFP)	PAHO	Pan-American Health Organization
DWCP	Decent Work Country Programme (ILO)	PBAC	Programme Budget and Administration Committee (WHO Executive Board)
DE	Decentralized evaluation	PHC	Primary Health Care
EB	Executive Board	PSA	Planning Resource Coordination and Per Monitoring
ECD	Evaluation Capacity Development	QA	Quality Assurance
EISI	Evidence Information Systems Integration	RO/	Regional Office
EMCP	Evaluation Manager Certification Programme	SDG	Sustainable Development Goals
EQAA	Evaluation Quality Assurance and Assessment system (UNFPA)	SEARO	Southeast Asia Regional Office
EQAS	Evaluation Quality Assurance System (WFP)	UN	United Nations
ERC	Evaluation Resource Centre (UNDP)	UNDAF	United Nations Development Effectiveness Framework
ERG	Evaluation Reference Group	UNDP	United Nations Development Programme
EURO	European/Europe Regional Office	UNEG	United Nations Evaluation Group
EVL	Evaluation and Organizational Learning (WHO)	UNESCO	United Nations Educational, Scientific and Cultural Organization
EvO	Evaluation Office	UNFPA	United Nations Population Fund
FAO	Food and Agriculture Organization	UNHCR	United Nations High Commissioner for Refugees
FENSA	Framework for Engaging with Non-State Actors	UNICEF	United Nations Children's Fund
GEHR	Gender Equity and Human Rights	UNSWAP	UN system-wide Action Plan
GEROS	Global Evaluation Reports Oversight System	USD	United States Dollars
GNE	The Global Network on Evaluation	WHA	World Health Assembly
GPW13 / 14	Thirteenth / Fourteenth General Programme of Work	WHE	World Health Emergency
HLE	High Level Evaluation	WHO	World Health Organization
IDP	Internally displaced person / Displaced Person	WPRO	Western Pacific Regional Office
IEE	Independent Evaluation of the Evaluation Function		

Executive summary

This study has compared the WHO evaluation function, both at global and regional level, with eight UN agencies (FAO, ILO, UNDP, UNESCO, UNFPA, UNHCR, UNICEF and WFP) that were selected purposively based on comparability to WHO in size or function, or as agencies recognized (e.g. by UNEG Peer Review/MOPAN) to have strong and mature evaluation functions.

The objectives of the study are to:

- Identify good practices applicable to WHO in terms governance, coverage and resources which contribute to ensuring 1) independence, 2) credibility, and 3) use.
- Compare current policy & systems, practice and resourcing of WHO evaluation function with selected UN entities' organizational setting, mechanisms, processes and tools.
- Make recommendations on the three dimensions (policy & systems, practice, and resourcing) to Member States and the Secretariat to enhance the ability of the WHO evaluation function to stimulate learning and promote accountability, transparency, and effectiveness through independent, credible and useful evaluations.

Whilst there is no direct comparator to WHO in structure, size or mandate, this study has sought to identify lessons learnt and best practices from comparator agencies which may inform WHO's approach to strengthening its evaluation function going forward.

The below table outlines some of the key features of WHO's evaluation function and that of comparator agencies, showing it (alongside UNHCR) has the lowest financial resources as a percentage of organisational budget, and a relatively low number of staff and volume of evaluations.

Table 1: Key Features of Comparator Evaluation Functions

Organization	Evaluation Budget (million USD)	Evaluation spend as a percentage of organization budget	Volume	Number of evaluation staff
WHO	\$7.7	0.1%	36 (11 corporate, 25 decentralized) (2022-23)	12
FAO	\$10.2	0.8%	103 (2021-22)	20
ILO	\$7.3	0.8%	72 (2 Corporate, 70 Decentralized) (2022)	12
UNDP	\$50	0.77%	416 (20 corporate, 396 decentralized) (2022)	35
UNESCO	\$1.1	3%	36 (5 corporate and 31 decentralized) (2023)	6
UNFPA	\$9	0.83%	23 (2022)	12
UNHCR	\$6.5	0.1%	26 (12 corporate and 14 decentralized) (2023)	13
UNICEF	\$65	1%	199 (20 corporate, 179 decentralized) (2023)	40
WFP	\$32.85	0.31	54 (27 corporate and 27 decentralized) 92022)	86

Key points are:

- A well-functioning evaluation system is integral for WHO to achieve its strategic goals, including those of GPW13 and GPW14 and its role in relation to the health-related SDGs and health emergencies; and to function as a knowledge-based,

learning organization with strong accountabilities.

- Over the last 10 years (since the 2014 JIU report), the 8 entities considered for this study have made substantial investments in their evaluation functions, in relation to evaluation policies, systems, practices and resourcing.
- Meanwhile WHO's evaluation function has also evolved in key aspects since the creation of a separate evaluation unit in 2014, and after a pause during the COVID-19 pandemic, there is now more recently additional impetus which has come with new leadership of the function. Strengths include delivery of high quality, relevant centralized evaluation helped by:
 - Bringing in evaluation specialists with relevant professional expertise from other agencies into EVL.
 - Developing more detailed evaluation guidance in key areas.
 - Working with regional leads by relaunching the global evaluation network and developing the partnership between the different levels of WHO's evaluators (global, regional, country).
 - Planning evaluations with the regions and moving to an LTA for procurement of support from evaluation providers.
 - Recent progress at regional level in developing capacity, although this has a long way to go both at regional and country level.
 - Creating a roster of evaluation specialists.
 - Developing a costed evaluation workplan.
- However, despite this good work, in many key respects (such as decentralized evaluation, coverage, financial and human resourcing, and evaluation use) there are still significant gaps in WHO's evaluation function at different levels. In summary, WHO has not in recent years been able to keep pace with the steadily advancing standards for evaluation among comparable UN organizations. The recommendations in this study and the examples of best practice from the comparators are intended to suggest how this can be remedied. There is substantial scope to learn from the innovations that have been introduced in the 8 comparators, see the overview of their functions in the main body of this report, and the agency summaries in Annex 3.

Policy

WHO's 2018 evaluation policy and accompanying implementation frameworks have some of the key elements required for a strong enabling environment for evaluation, but there are also many gaps and awareness of the policy requirements is highly variable. All the comparators have reviewed and updated their policies, informed by UNEG peer reviews or independent evaluations; and then made progress in key areas such as coverage norms, use of evaluation and resourcing. The ILO approach is one interesting example of how results against evaluation policy norms are tracked in detail. The WHO 2018 policy needs updating and fleshing out in line with UNEG norms and standards and best practice in comparators, with a clear roadmap for how it will be delivered.

Recommendations

R1. Update the 2018 evaluation policy, addressing the gaps noted in this study. Keep it regularly updated, informed by UNEG peer reviews/independent evaluation, every 5 years.

R2. Develop explicit coverage norms in key areas of WHO work, and track progress annually, relating them to strategic outcomes for WHO. These should include:

- Balanced coverage of WHO's intended strategic outcomes based on the SDG3 goals, GPW14, key WHA resolutions and policies.
- 100% coverage (through inter-agency evaluations) of global health emergencies and level 3 humanitarian crises.

- 100% coverage of all country strategies in the year prior to their revision, for those countries with off track health SDGs and/or high levels of risk identified in programming and audits. Coverage at least once every 2 programme cycles for all other country strategies.

R3. Set out a fully developed roadmap on how the evaluation policy will be delivered, accompanied by a detailed results reporting arrangement including key performance indicators around quality, credibility, capability, resourcing and use.

Systems and processes

WHO's evaluation systems also include some of the elements one would expect to see, including a consultative approach to work planning with a newly developed costed workplan, guidance and support from EVL, reporting to the governing body via PBAC, management response systems etc. Again, however, there is a need to take this to the next stage and strengthen systems -this is important particularly for decentralized evaluation¹.

Recommendations

R4. Further develop and embed an institutional approach to work planning (including with WHE) and consulting on selecting evaluation topics, informed by the coverage norms as per R2.

R5. By agreement with Management and the Governing body as appropriate, set out an explicit budget line for evaluation so that the costed workplan and the supporting function can be delivered; and to underpin structural independence of the function.

R6. Strengthen evaluation oversight and visibility by:

- informal sessions with Member States where EVL presents the evaluation report and management present their response².
- an internal evaluation committee in which WHO senior management discusses and approves the evaluation and the management response.

R7. Monitor the coverage and quality of decentralized evaluations and take steps to map the existing coverage more reliably and enhance the quality of DE over time.

- This should be supported through a well-developed approach to training and capacity building learning from the approach adopted in ILO, UNDP, WFP and others.
- Ensure that specific guidance is available to inform quality of humanitarian and evaluation of emergencies.

R8. Consolidate existing elements such as QA checklists and templates already developed by EVL, into a fully developed QA system, with annual reporting including ex-post QA scoring for centralized and decentralized evaluations (including WHE evaluations). A good model here is the approach developed in WFP. A staged approach is recommended, starting with a system at central level closely followed within 2 years by expanding to cover decentralized evaluations.

¹ An influential JIU report (JIU/REP/2014/6) on evaluation functions looked at their maturity in the UN systems and covered many aspects including a finding that policies, norms and standards for decentralized evaluation were inadequate. Since then, all the comparator agencies considered here have strengthened their decentralized evaluation functions. WHO's decentralized structure and focus on results at the level of member states means that progress on this aspect is particularly important.

www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/en/reportsnotes/JIU%20Products/JIU_REP_2014_6_English.pdf

² The internal evaluation committee should also receive annual reports from EVL on progress against results outlined in the evaluation roadmap (i.e. evaluation planning, implementation, resourcing, quality, evaluation culture) and ensure necessary actions are required to stay on track.

Financial and human resources

The resourcing of the evaluation function is not clear and is significantly below what is needed given the size, scope and complexity of WHO's work. There have been sharp increases in resourcing for evaluation in many of the comparators, but not so in WHO. A first essential step is (as per recommendation 5 above) to have an explicit budget line for evaluation. The second key step is to identify and implement a target level of resourcing overall. The 2018 policy includes a reference to the norm of 0.5% to 1.0% established by the JIU report, but this has never been implemented and resources are closer to 0.1% currently.

The human resources at the level of EVL and capacity at regional and country level to support evaluation are an order of magnitude below what is needed for a well-functioning evaluation function in an organization like WHO. This requires additional staffing both in EVL and at regional and country level (see R12 below).

Recommendations

R9. Ensure that there are dedicated 'draw down' resources available for humanitarian and evaluation of emergency response, articulated in the costed workplan. This should include identifying an agreed percentage (e.g. 1%) of voluntary contributions for emergencies that is dedicated to evaluating the emergency response.

R10. Commit to a target level of resourcing as a percentage of WHO expenditure. The target should be set out in the revised policy at the level of 1.0%, in line with the norm across the UN, and then related to progress against and resources for delivery of coverage norms³.

Decentralized evaluation

The decentralized evaluation function in WHO is significantly under-developed. This is at odds with WHO's overall strategic focus, which includes a decentralized governance structure, strong emphasis on its regional and country level work and assessing impact at the level of member states in relation to the triple billions.

Recommendations

R11. Regional directors, with advice from the director of EVL, should be asked to develop a clear strategy for WHO to invest in the decentralized evaluation function to bring it up to the standards of comparable UN organizations and address issues highlighted 10 years ago by JIU. Build a common strategic approach for evaluation at all 3 levels, using the helpful collaborative approach already established in the Global Network on Evaluation (GNE) as an entry point⁴.

R12. Build capability at regional and country level on evaluation, including regional evaluation units that are staffed at the appropriate level (P4 level as a minimum) with reporting arrangements to regional directors and with a 'dotted line' reporting to the director of EVL on professional and technical aspects. Develop a network of country level focal points for M&E in the larger WHO offices, who can also support smaller countries as required.

Demand for evaluation –

³ This should also be accompanied by a clear and explicit statement in the revised evaluation policy of responsibilities of Management and Member States to ensure that necessary resources are made available for evaluation to meet the agreed standards.

⁴ This must not be a top-down approach, given WHO's governance structures and culture.

evaluation culture / demonstrating value

Strengthening WHO's evaluation function cannot only be done through supply side measures to build capacity: attention to the demand side and culture of evaluation across the organization is at least as important.

Recommendations

R13. As part of revising and launching the new Evaluation Policy, take steps to build awareness of evaluation through a major communications and training initiative led by EVL and championed by senior managers.

- Through this process communicate a clear vision of evaluation across the organization, how it adds value and how evaluations are intended to be used.
- Spell out the accountabilities set out in the evaluation policy so that management and staff are aware of what is expected and why.

R13b. Ensure there is a high-level forum for discussing evaluations in depth.

- See the recommendation R5 above, which would put in place informal sessions with member states a high-level internal Evaluation Committee with representatives from senior management, chaired at ADG level.
- This would need appropriate representation at regional and country level, who can take delivery of and discuss (decentralized) evaluations, findings and recommendations.

R13c. Build demand, support and buy-in for evaluation and maximise use.

- This needs to start with messages communicated by the governing body and senior management on the importance of evaluation to provide a platform for demand.
- It would then be reinforced by closing the learning loop through consistent production of high quality, credible and useful reports which demonstrate what value can come from the evaluation function.

Use, impact and follow up

Ultimately the test of the quality of WHO's evaluation function is the extent to which evaluations are used and inform decision making. Partly because EVL's resources are so stretched in delivering evaluations, there is limited capacity at present for ensuring use and follow up. Best practice in other entities is to have a strategic and innovative approach to disseminating evaluations accompanied by strong incentives for action.

Recommendations

R14. Modernize and deepen EVL's approach to dissemination of evaluations, using a wider range of tools and formats as part of a strategic investment in making evaluations accessible. Starting at the top, this should include the regular informal sessions with EB representatives of member states.

R15. Strengthen the systems and incentives for follow up on evaluation recommendations by clearly defining the responsibility of senior managers in this respect, moving the focus from EVL to operations and building this into performance management and compliance.

- The internal evaluation committee proposed in R6 would provide an important focus within the organization on this, reinforced by greater interest from MS. This in turn would help in creating incentives for staff to focus on the management response actions following each evaluation.

- Staff should receive guidance and training on what is expected around management response and follow up to evaluations.

Introduction

In May 2023, the Programme Budget and Administration Committee of the WHO Executive Board agreed the following decision, namely that WHO should:

Conduct a comparative study of evaluation functions and coverage across entities of the United Nations system that are comparable to WHO in size or structure and, together with recommendations to strengthen the independence, credibility and use of the evaluation function, submit a report of the findings of the study to the Programme, Budget and Administration Committee at its fortieth meeting in May 2024.

Review Purpose and Use

This study commenced in September 2023 and was completed in February 2024. The objectives of the study are to:

- Identify good practices applicable to WHO in terms governance, coverage and resources which contribute to ensuring 1) independence,
2) credibility
3) use.
- Compare current policy & systems, practice and resourcing of WHO evaluation function with selected UN entities' organizational setting, mechanisms, processes and tools.
- Make recommendations on the three dimensions (policy & systems, practice, and resourcing) to Member States and the Secretariat to enhance the ability of the WHO evaluation function to stimulate learning and promote accountability, transparency and effectiveness through independent, credible and useful evaluations.

Approach

Findings for this comparative study have been developed based on a review of key WHO and comparator agency documents as well as 39 interviews with: WHO staff at HQ, regional and country level, Member States, and representatives from each of the eight comparator agencies to ensure that our findings are well-grounded in WHO's organizational reality and that our recommendations are useful, realistic, contextually relevant and implementable.

In the inception phase for this study, the review team undertook 9 interviews and a preliminary document review and used this to develop a set of review questions to guide the inquiry; these are informed by both the ToR and the team's

understanding of this exercise, the MOPAN performance indicators on evaluation and the UNEG Evaluation Peer Review criteria. The inception phase culminated in an inception report which was finalized integrating feedback from WHO.

Recognizing the different configurations and structures of WHO and the comparator agencies for this assessment, for the purposes of this study, we have defined ‘centralized’ evaluation as those managed/commissioned by an organization’s central evaluation unit/office (EVL in WHO), whereas ‘decentralized’ refers to those ‘not managed by EVL’ or ‘not managed by an evaluation function’ either at a regional, country or project level.

The eight comparator agencies (FAO, ILO, UNDP, UNESCO, UNFPA, UNHCR, UNICEF and WFP) were selected purposively by the WHO evaluation office based on comparability to WHO in size or function, or as agencies recognized (e.g. by UNEG per review/MOPAN) to have strong and mature evaluation functions which might offer lessons to WHO as its evaluation function matures. We have shared with each agency a summary of findings regarding its evaluation function and have ensured that this data is validated by the agencies themselves before inclusion in the report. We recognize that it is not within the scope of this exercise to ‘judge’ or assess the evaluation functions of other UN entities and as such, we have ensured that evidence on the performance of other evaluation functions is derived from existing, publicly available and independent reviews/evaluations.

We have conducted analysis as a team, analyzing data against the inquiry framework from which we have developed our recommendations. Preliminary findings were presented to WHO in early December 2023 to offer an opportunity for stakeholders to validate or challenge these findings from the varied and valued vantage points that they bring, and to jointly reflect on the initial conclusions presented by the Assessment Team. Recommendations have been codeveloped in a workshop with key WHO stakeholders in January 2024 to ensure that the recommendations are useful and relevant. This draft report has been prepared for comment, setting out the key findings, conclusions, and recommendations. This will then be finalized pending feedback from WHO and then the report will be presented to both the UNEG Heads, and to the PBAC / EB in May 2024.

Table 2: Enquiry Questions

Enquiry Area	Key question
Policy	<p>Is the Evaluation policy up to date, comprehensive, formally approved by governing body and regularly reviewed?</p> <p>Is the evaluation policy underpinned by a clear vision and strategy and fully implemented, overseen by director and evaluation department, and supported by overall oversight by the governing body?</p> <p>Does the policy guide the implementation of the different categories of evaluations, such as strategic, thematic, corporate level evaluations, as well as decentralized evaluations?</p>
Processes and systems	<p>How is evaluation work planning undertaken to ensure relevance, coverage and utility of evaluations? And how is this budgeted? Are there coverage norms in place?</p> <p>How are stakeholders at different levels (Countries, RO, HQ, governing body) consulted on the evaluation work plan?</p> <p>How are evaluation managers selected and trained?</p> <p>Are effective planning arrangements in place for timely procurement, design and delivery of high-quality evaluations?</p> <p>Are effective procurement arrangements in place for external contracting e.g. LTAs, roster of experts etc?</p> <p>Is the evaluation website accessible, well organized and up to date and regularly used by internal and external stakeholders?</p>

	<p>Are evaluation recommendations tracked and monitored (including that follow-up and roles are clearly understood, management responsibility for actioning recommendations, EVL responsibility for tracking and reporting back)?</p> <p>Is the Evaluation function regularly reviewed and assessed (e.g. peer review, external evaluation of function, meta-synthesis reports on quality)?</p>
Enabling environment and culture	<p>To what extent is the Evaluation function effectively led, championed and supported to develop by the organization's leadership?</p> <p>To what extent does the Evaluation Office Engage with the governing body and senior management on evaluation topics e.g., workplan and evaluation findings for major strategic evaluations?</p> <p>Are roles and responsibilities for evaluation clear across the organization?</p> <p>Is there clear understanding of the difference between evaluation and other functions (performance monitoring, research, audit)?</p> <p>Do senior managers champion evaluation as useful for the organization, for learning and improvement?</p>
Independence	<p>To what extent does the evaluation function have operational and financial independence and autonomy of decision-making?</p> <p>How do reporting lines and arrangements for appointing Director of Evaluation protect independence?</p> <p>How are evaluations conducted and what is the role of evaluation staff in this?</p>
Quality	<p>Are there clearly articulated quality standards for evaluations in line with UNEG norms and UN system-wide Action Plan (UNSWAP) requirements at a centralized and decentralized level?</p> <p>Are these standards supported by the provision of appropriate guidance and tools?</p> <p>How is the quality of evaluations monitored (in real time and post-hoc)?</p> <p>Is there an annual synthesis exercise to provide feedback on evaluation quality?</p>
Resourcing and human and financial capacity	<p>How is the central evaluation programme funded? And at a decentralized level (Voluntary contribution vs. regular budget)?</p> <p>Is there clarity on the budget(s) line for evaluation, both in overall size and how it is set, and where the resources come from?</p> <p>Is the evaluation workplan funded and prioritised?</p> <p>Are evaluation resources adequate for delivering the workplan and implementing evaluation policy?</p> <p>Is the quantity and quality of staffing of evaluation function proportionate to policy goals and workplan, both centralized and decentralized?</p> <p>Are evaluation staff adequately trained and supported?</p> <p>Are there effective networks for professional evaluators to support mutual learning?</p>
Evaluation use and impact	<p>Is evaluation perceived as useful and seen to make a measurable and significant difference to organizational learning?</p> <p>How are evaluations used to inform strategic planning? Organizational learning? Accountability?</p> <p>What examples are there of uptake and use of recommendations demonstrating their utility?</p> <p>Is ownership and relevance of evaluation findings supported by effective engagement of evaluation with management, through the different stages of the planning and evaluation cycles?</p> <p>Are evaluations effectively disseminated and shared internally, with tailoring for different audiences including senior management, staff and governing body?</p>

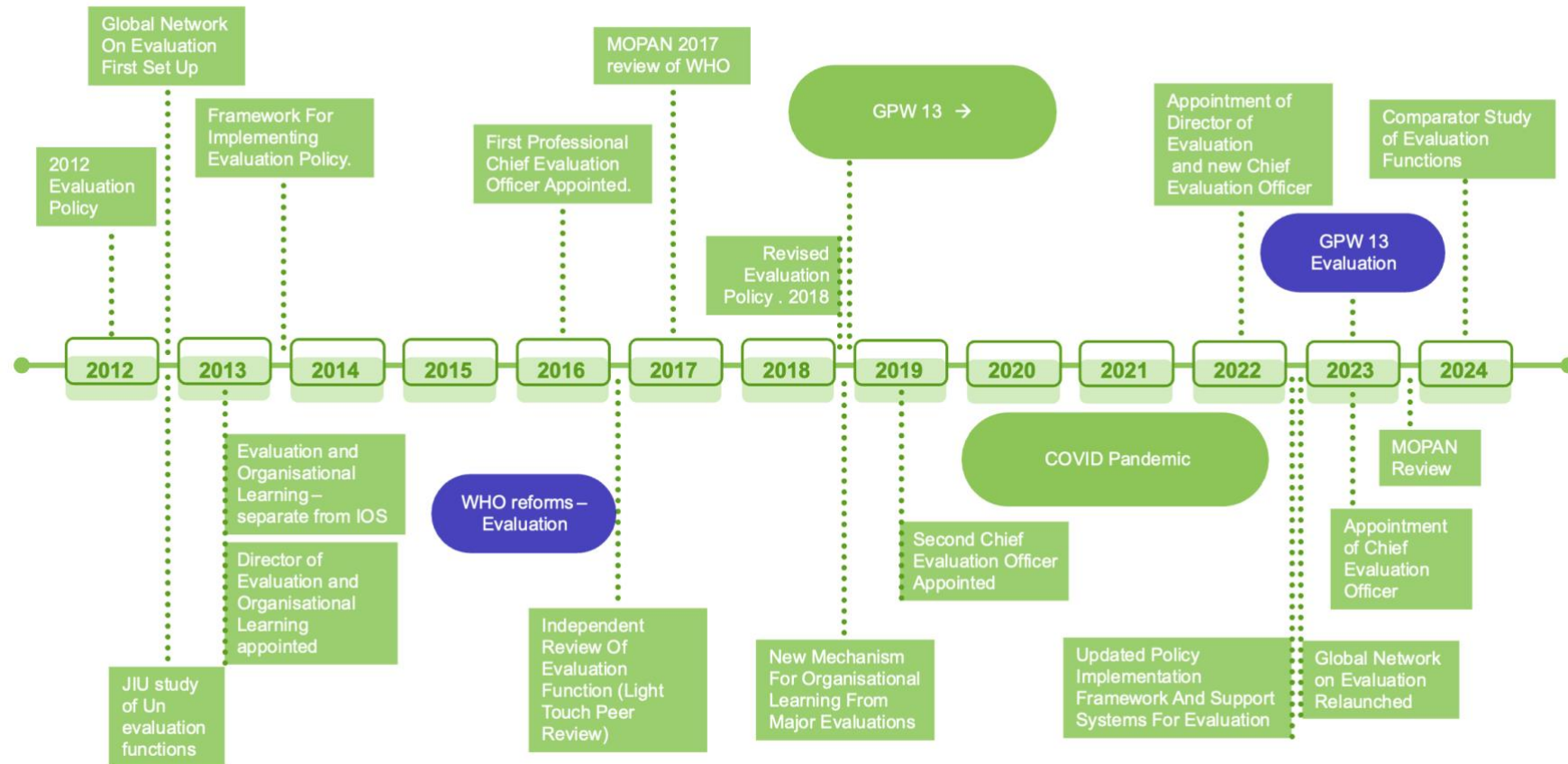
	Are effective systems for knowledge management and curating evaluation evidence in place?
Engagement and follow-up	<p>What mechanisms are in place for the dissemination of evaluations?</p> <p>Are synthesis products produced where appropriate, for cross departmental learning across organization?</p> <p>Are other types of products (beyond evaluation reports) used to drive learning e.g. videos, briefs, visualization?</p>
Relevance and responsiveness	Is the quality, mix and volume of evaluations at a centralized and decentralized level, and at country level, in line with and proportionate to organizational needs?

WHO Evaluation Function Overview

Compared with those in other large UN agencies, WHO's independent evaluation function is relatively young (less than 10 years old in its current format). The Evaluation and Organizational Development Office (EVL) was created in 2014, bringing evaluation out from under the umbrella of internal oversight services with the creation of an evaluation office headed up at director level and with dual reporting to the Executive Board and the Director General's office. This started the process of clearly differentiating evaluation from other related functions including audit, while underpinning independence and increasing the visibility of the function. These decisions were partly responding to a comprehensive and influential Joint Inspection Unit (JIU) report on evaluation in the UN, which looked at the whole range of UN entities and their maturity. Among other organizations it identified evaluation in WHO as relatively under-developed in maturity.

It has been on a journey towards becoming a more mature function, at different levels. That maturation process is by no means complete and was somewhat interrupted by the impact of the demands of the COVID-19 response, for which WHO was in the 'eye of the storm'.

Figure 1 Summary of key steps in developing the evaluation function



Several key elements are in place both on policy, systems and practice at centralized level, as outlined in the findings section. At decentralized level, some elements are also in place in most of the WHO regions, such as the Americas, SEARO and to some extent AFRO and EMRO. In the case of PAHO, the evaluation function at regional level is longstanding and more mature, with PAHO having its own 2021 evaluation policy and senior evaluation adviser.

It is recognized by EVL there further improvements and strengthening are required with regards to policy, systems, resourcing and practice. This comparability study is an opportunity to articulate what those gaps are and allow a process of learning from other UN agencies that are somewhat further down the path of the evolution of their evaluation functions.

Arguably the process of making evaluation more visible across the organization is a work in progress, with some key informants noting overlaps with audit and patchy levels of awareness of the policy both centrally and regionally. In response to the MOPAN report on WHO in 2017-18 and an independent review of the evaluation function shortly before that, further changes were made to strengthen different aspects, notably on reporting lines and independence (including the term and appointment of the Director of Evaluation), organizational learning and tracking management responses.

WHO context

The comparator study is an opportunity to learn UN entities. At the same time, the institutional context for WHO and organizational culture and features also need to be considered – the distinct aspects of how WHO works that need to shape decisions on the evaluation function. Key features which were emphasized in interviews and document review are:

- The **strategic directions set out in GPW13**, including the focus on results, and in particular shifting the centre of gravity towards what WHO achieves through Member States and at the country level. The focus on results and impact underlines the importance of having an effective independent evaluation function that draws lessons learned and accounts for strategic results being achieved by WHO.
- The country level focus of the organization and results systems also places a premium on having effective independent evaluation products at country level. This has started through the **evaluations of WHO contribution at the country level**. These are relatively new products that were initiated in the past five or six years, and so far, there have been around 9 such evaluations and one synthesis. Three are ongoing and nine are planned in the 2024-2025 biennium in the AFRO and EMRO regions.
- WHO's business model and **accountabilities at regional level** are also critically important. While the World Health Assembly provides the overarching governance and the role of HQ is important in leading the organization, each region has its own regional committees and accountabilities and there is significant variation in how the organization works in each region. This means that a single top down and highly integrated model of an evaluation function is not likely to be appropriate or feasible. On the other hand, maximizing the synergies between evaluation functions at the 3 levels is critically important, including ensuring that learning can be aggregated from country, to regional, to global level.
- The nature of WHO's business also shapes the evaluation needs. WHO has a strong **normative and technical focus**. On the health emergencies side, it also has a strong delivery and response requirement, as illustrated during the COVID-19 and Ebola responses. **WHO does not have a project focus** in the field. Instead, it is more programmatic, although it operates as an implementing partner in key areas such as polio, HIV and routine immunization.

- It is important to note that the process of **learning from evaluations needs to include external audiences** that are a feature of WHO's nature as an intergovernmental body. Specifically, evaluations should have a role in informing member states about what WHO is doing and what activities are most relevant and effective. At present, the governing body does not receive or discuss individual evaluations as a matter of course, in the way that happens in many other UN entities, so this channel of learning is not fully exploited.
- A key feature of WHO's resourcing is that it **is highly dependent on voluntary contributions, which make up over 80% of its resources**. While this is by no means unique in a UN context it does have specific implications for evaluation in WHO. The extent to which centralized functions can have clear resource envelopes funded through core resources is highly constrained. Evaluation, like other oversight functions, is funded through the Director General Office budget and the budget line is not separately identified or approved at EB level, as would be the case in many UN or multilateral organizations.

Key findings

Evaluation Policy

WHO's Evaluation Policy was last revised in 2018, with the first iteration approved in 2012. The policy defines the overall framework for evaluation within WHO, to foster culture and use of evaluation across the organization and to facilitate conformity of evaluation at WHO with best practices and with the norms and standards for evaluation of the United Nations Evaluation Group. The policy outlines two kinds of evaluation⁵, corporate and decentralized, defined as:

(a) Corporate evaluations are managed, commissioned or conducted by the Evaluation Office, and include programme evaluations, thematic evaluations and office-specific evaluations.

(b) Decentralized evaluations are managed, commissioned or conducted outside the central Evaluation Office, that is, they are initiated by headquarters clusters, regional offices or country offices and mainly comprise programmatic and thematic evaluations.

The policy outlines the accountability and oversight role to support the implementation of both the policy of the Executive Board and the Evaluation Office, particularly with regards to corporate evaluation but it lacks detail with regards to the roles and responsibilities of managers (heads of regions, divisions, departments and offices, and other programme/project managers) as commissioners of decentralized evaluations, providers of data and information to evaluators, or as users of evaluation both in implementing recommendations and in learning. Furthermore, accountability and oversight for, and ensuring quality of decentralized evaluations is not articulated in great detail and the lack of coverage norms is a gap.

The evaluation policy is accompanied by an implementation framework⁶, developed in 2022, which sets out how the evaluation policy is implemented and presents mechanisms and support systems for the conduct of evaluation across WHO offices: i.e. headquarters (HQ) departments/divisions, regional offices (ROs) and country office. This again sets out some roles and responsibilities with regards to evaluations, in a more granular detail (figure 2) than in the evaluation policy but makes no reference to key aspects of evaluation conduct, for example quality assurance, and monitoring implementation of evaluation recommendations. Further details on roles and responsibilities in the commissioning and conduct and quality

⁵ WHO (2018), Evaluation: Evaluation policy 2018, EB143(9), World Health Organization, https://cdn.who.int/media/docs/default-source/evaluation-office/who-evaluation-policy-2018.pdf?sfvrsn=3448fe8a_3&download=true.

⁶ WHO Implementation framework of the WHO evaluation policy 22 November 2022

assurance of evaluation are detailed in a Standards Operations Procedures and Quality Assurance Checklists draft documents, developed in December 2023, that are being piloted by the Office.

The implementation framework is also missing a results framework or metrics by which progress towards the implementation of the evaluation policy can be measured. In contrast, for example, ILO's evaluation strategy, which operationalizes its evaluation policy, includes clear outcomes, sub-outcomes, milestones and performance indicators to support policy implementation and monitor progress, and EVL submits annual reports to the Governing Body as part of its Annual Evaluation Report, ensuring effective monitoring of outcomes.

The WHO Evaluation Office is required to, and does, report annually to the Executive Board on progress in implementing evaluation activities⁷. This report to the governing body is a key part of ensuring independence and a vehicle for showcasing the work on evaluation in one place, including major strategic evaluations, corporate and country evaluations and some decentralized evaluations. There is a specific section in the report which captures some specific recent examples of how the lessons learned from corporate evaluations are informing policy and decision-making in the Organization. PBAC members interviewed for this study spoke positively about this reporting and were able to reference some of the key strategic evaluations which had taken place recently but stated that they had less visibility on decentralized evaluations taking place. However, the report currently focused on activities rather than at an outcome-level. This is in contrast to the strategies of other agencies; for example the FAO Evaluation Strategy has a series of objectives that are reported against annually (Objective 1.1: Promote strategic use of resources in alignment with FAO information needs by regularly assessing demand to identify priorities for evidence-building activities, Objective 2.2: Expand quality assurance processes to provide supportive and robust quality assurance throughout the life cycle of FAO's Office of Evaluation's work), and the ILO Results-Based Evaluation Strategy outlines strategic enablers, strategy outcomes, sub-outcomes and outputs which are all reported against on an annual basis.

One of the keys challenges noted with regards to WHO's evaluation policy pertains to its visibility and use; interviews with senior WHO staff indicated little awareness of the policy and its commitments and the roles of different actors in ensuring its effective implementation. Awareness was even less at country-level where there was very limited knowledge of the policy's existence, or even of EVL and its work, indicating the need for increased socialization of the evaluation policy and the need to raise of the profile of EVL more generally across the organization.

Alongside the policy and implementation framework, EVL has been working recently to update and furnish the organization with further, practical guidance on undertaking evaluations. This includes a 'Practical Guide to Evaluation for Programme Managers and Evaluation Staff' (2023), 'Guidance note on integrating health equity, gender equality, disability inclusion and human rights in WHO evaluations' (2023), and a 'Framework for evaluations of WHO's contribution at country level' (2022). However, it is unclear from interviews with WHO staff in a range of across the organization how well socialized or disseminated these documents have been as they were not referred to or recognized by stakeholders in any interviews. Prior to the production of these resources, the last guidance was the 2012 Evaluation Practice Handbook which provided step-by-step practical guidance to evaluation in WHO.

Recognizing the prominence of the World Health Emergencies Programme within WHO, it surprising to note the Evaluation Policy does not outline any specific considerations regarding evaluation in emergencies or humanitarian crises. Furthermore, the consideration of emergencies and humanitarian evaluation is also absent from more recent guidance such as the 'Practical Guide to Evaluation' developed in 2023, although there is guidance on specific evaluation questions relevant for evaluation of emergencies in the 'Guidance note on integrating health equity, gender equality, disability inclusion and human rights in WHO evaluations'.

Relevance and responsiveness

⁷ WHO Evaluation: Annual Report, Executive Board 151st Session

In terms of evaluation coverage, the evaluation policy articulates a series of criteria for the selection of topics for evaluation, including:

- Organizational requirement relevant to: global, international or regional commitments; specific agreements with stakeholders, partners or donors; requests from governing bodies.
- Organizational significance relating to general programme of work priorities and core functions; level of investment; inherent risks; performance issues or concerns in relation to achievements of expected results.
- Organizational utility relating to a cross-cutting issue, theme, programme or policy question; potential for staff or institutional learning (innovation); degree of comparative advantage of WHO.

However, there are no specific coverage norms to inform the selection of centralized or decentralized evaluations and it is not clear if the application of these criteria ensures sufficient coverage of WHO's operations across its three levels. The draft workplan for Evaluation for 2024–2025⁸ outlines its intent to “*cover a wide range of thematic areas and ultimately all the Triple Billion goal areas*”, although it is not clear from the document how this intention has been/will be realized.

Whilst there could be clearer and more robust criteria for the selection of evaluations, there is evidence that the evaluation office is increasingly meeting strategic needs at a centralized level. It has commissioned and managed recent, important evaluations including for example the independent evaluation of the WHO reforms and more recently evaluations of Gender, FENSA, results-based management, normative work, presence in countries, COVID-19 solidarity response fund and many other topics. The ongoing evaluation of GPW13, commissioned to inform ongoing decisions on design of GPW14, is a flagship product for the evaluation function being used in ‘real time’. Most of the decentralized evaluations known about and completed or being planned or underway (e.g. Evaluation of AFRO functional review, PAHO COVID-19 response) are also of relevance. Going forward, there is clear evidence that EVL and regions have been focusing on topics which are relevant and strategically important, for example, GPW13 Evaluation⁹ to feed into GPW14, yellow fever¹⁰ and of WHO's Results-Based Management (RBM) Framework.¹¹

The best practice for evaluation work planning is to have a single, comprehensive, multi-year workplan which relates the strategic objectives and planning cycle of the organization, which is costed and resourced, and which is agreed with the governing body and in consultation with senior management. Typically, these plans are developed through effective and regular consultation with the governing body and managers but informed by coverage norms in the evaluation policy and understanding of evidence requirements and gaps. This has become the norm in the eight comparator agencies.

Historically, work planning for EVL has been undertaken each biennium and led by the EVL office. Several stakeholders interviewed felt that previously, there has been a lack of visibility of how this was undertaken, that the process was not sufficiently consultative and that there was not always a clear strategic rationale apparent for the evaluations selected. There has been a perception that evaluation workplans were somewhat of a ‘Wishlist’ and choices influenced by donors’ interests rather than led by organizational need. A number of WHO staff interviewed felt that the evaluations selected often focused too heavily on WHO reforms and internal processes, rather than focusing on the impact or contribution WHO actually makes. This was felt to exacerbate the sense of evaluation being used and viewed as a compliance tool across the organization rather than for learning.

Whilst there are recognized constraints to WHO undertaking volumes of evaluations of its impact (resources, technical expertise, questions of contribution vs. attribution of health outcomes), there does seem to be an appetite for considering in more detail how WHO might do more in this regard, focusing more on its contribution at country level and its

⁸ WHO Evaluation: update and proposed workplan for 2024–2025, Executive Board 154th Session

⁹ https://cdn.who.int/media/docs/default-source/evaluation-office/evaluation-report-gpw13.pdf?sfvrsn=215b2a79_3&download=true

¹⁰ <https://www.who.int/publications/i/item/WHO-DGO-EVL-2023.1>

¹¹ [https://www.who.int/publications/m/item/independent-evaluation-of-who-s-results-based-management-\(rbm\)-framework-\(2023\)](https://www.who.int/publications/m/item/independent-evaluation-of-who-s-results-based-management-(rbm)-framework-(2023))

accountability to affected populations.¹² Furthermore, the workplans previously were also not felt to give a full view of the evaluation activity taking place across WHO as there are examples of evaluations commissioned at a decentralized level where EVL was not sighted. Attempts to set up and maintain a comprehensive repository of evaluations at the three different levels have also run into challenges, and it seems likely that more evaluations are being undertaken largely in response to donor requirements than are captured in the evaluation workplan.

The process and workplan for 2024–2025 demonstrate a much more consultative and strategic approach to evaluation work planning. The workplan is currently available in draft form and due to be approved in January 2024. EVL has clearly sought in this iteration of the work planning process to develop a more comprehensive approach, working with the regions, and to address some of the noted concerns. There has been greater engagement and consultation with regional colleagues, at country level and with technical departments.

Whilst the forthcoming workplan does set out several of the planned decentralized evaluations, it is not clear if this is the full picture of evaluations. For example, there are only country evaluations listed for the AFRO and EMRO region for the 2024-25 period but not for any of the other regions. The level of coverage and volume of decentralized evaluations is much lower compared to comparator agencies reviewed, including evaluations of country strategies and evaluations of country level contribution. For example, in UNICEF, there are 200 evaluations a year and 95% of these are decentralized. In relation to country programme evaluations, many other agencies have coverage norms in place and a strong track record of evaluating all country programmes at the end of the strategy period, whereas there is no systematic approach for WHO in undertaking these.

Additionally, the forthcoming workplan includes relatively few emergency or humanitarian evaluations listed. The draft workplan outlines that together with the WHO Health Emergencies Programme, the Evaluation Office conducted a separate planning exercise for evaluations of humanitarian actions/emergency interventions due to the need to take a different evaluation approach in humanitarian situations with a high degree of fluidity and the benefits of introducing a concurrent feedback loop to emergency operations. However, interviewees from EVL outlined that they do not have oversight of the planning and selection of evaluations undertaken by WHE and are generally not involved in their management. Although the 2024-25 workplan states that evaluations of emergencies should be undertaken as required, there is not allocation for this within the costed organization-wide workplan.

The workplan also sets out in intent to improve the balance in coverage of the Triple Billion pillars so that Member States and stakeholders are better informed of the achievements under the Thirteenth General Programme of Work in a holistic manner (although it is not clear from the document exactly how this has been established and evaluations are not mapped to GPW13 outcomes). It has also been reviewed by an Independent Expert Oversight Advisory Committee. The 2024-25 workplan is the first costed workplan that WHO has developed and given that the workplan is to be signed-off by the Executive Board, provides an impetus to EVL to advocate for the necessary resources to fulfil the workplan's aspirations. As well as a centralized evaluation workplan, there are also evaluation workplans in a number of regions (E.g., SEARO, PAHO). In AFRO, there is an evaluation registry of past and future evaluations (or evaluative activity) aligned to GPW13 Outcomes.

WHO takes part in Joint Evaluations of Level 3 inter-agency humanitarian evaluations and is currently leading in the Joint Evaluation of the Global Action Plan for Healthy Lives and Well-being for All, with planned participation in a number of other joint evaluation exercises. Stakeholders from some comparator agencies felt though that there remains further opportunity for WHO to engage in joint evaluations, for example on refugee health and the social determinants of health.

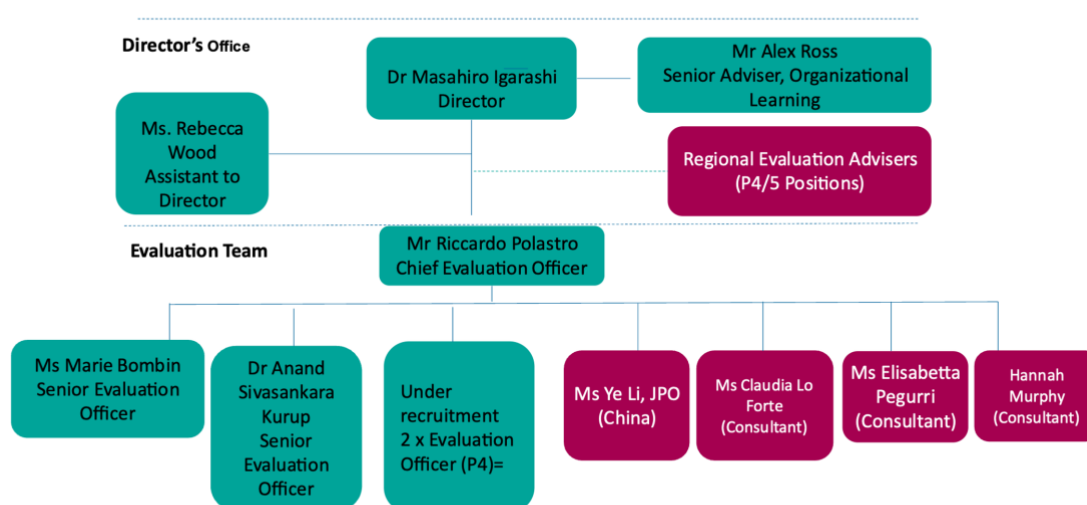
Human Resources

¹² The term 'impact evaluation' has many different meanings. In the WHO context with its strong links to public health and scientific methods might be understood as collecting primary data in the field at household level and using specific methodologies such as experimental (randomised control) and quasi-experimental designs to establish causality. Few UN agencies have yet invested in this type of evaluation – the exceptions are WFP which has a specific strategy for impact evaluation, while UNICEF is also considering how it can make greater use of impact evaluations. Instead, impact here refers more to assessing the causal attribution and of results to WHO's work, requiring the presence of good quality data and sufficient human and financial resources.

In terms of human resources, there have been important steps to bring in external and professional evaluation expertise – starting at the level of the Chief Evaluation Officer (a position established in 2016) and with the appointment of Dr Igarashi as Director at D2 level more recently. This has allowed expertise from other evaluation functions to inform WHO’s work, namely WFP (Anne-Claire Luzot), UNICEF (Robert McCouch and Riccardo Polastro) and FAO (Dr Masa Igarashi).

Nevertheless, the WHO evaluation is significantly under resourced given the size and scope of WHO’s work, both in size and in the number of professional evaluators. The below organigram sets out the staffing as of January 2024:

Figure 2 Evaluation organigram



Currently there are five staff, two of which are evaluation specialists and three consultants. Three positions are vacant, two P4s and a JPO. One of these P4 positions is intended to include a focus on country strategy, programme and humanitarian evaluations. This is not a significant improvement from the 2017 Independent Review of the WHO Evaluation Policy, which found that: “Current resourcing of evaluation at around 0.1% of budget and only 6 staff in EVL is well out of line with accepted UN benchmarks and quite obviously unrealistically low for a large and important organization with ambitious policy and organizational aims”.¹³ In interviews with WHO staff, it was clear that there is a lack of designated staff within EVL to enable it to fulfil its mandate and the obligations of its workplan. This has meant that in some cases, planned evaluations have not been able to go ahead but has also meant that other key roles of the function, such as the work on culture, on capacity building etc. have not received due attention.

In terms of regional staff, WHO is at a relatively early stage compared to other agencies in building this decentralized capacity. In some large agencies, there are P5 (UNICEF) or P4 (WFP) regional evaluation advisers, and in WFP and UNICEF there are teams at regional level. In WHO, there is a P5 in PAHO, and P4s in EMRO, SEARO and EURO (EURO’s is very recently recruited. Recruitment for the P4 position in WPRO is underway. In AFRO, there is a temporary P4 for 6 months and two P3s. With the exception of PAHO, these are all relatively recently in post and up until this, decentralized evaluations were supported by ‘evaluation focal points’ who did the role on top of their ‘day job’ and had a lack of specific, designated time

¹³ Independent review of the implementation of the WHO evaluation policy and the framework for strengthening evaluation and organizational learning, June 2017.

within their workplan for evaluation tasks. The regional evaluation advisors are accountable to and report to the RO rather than EVL; this is different to most other agencies which see the reporting lines of regional staff to evaluation as key to enhance accountability for evaluation, or others where there are dual reporting lines.

Financial resources

The WHO Evaluation Policy sets out that the Director General must ensure that there are adequate resources, within the range recommended by the United Nations Joint Inspection Unit (between 0.5-3%) to implement the biennial Organization-wide evaluation workplan; not only the evaluations to be conducted but all activities required to ensure the strengthening of the evaluation culture and the professionalization of evaluation conduct across the Organization. However, EVL is significantly under resourced to meet the needs of the organization at HQ, regional and country levels.

For 2022-23, the budget approved for EVL by the Executive Board was US\$ 7.7 million (US\$ 5.4 million for staff and US\$ 2.3 million for activity costs to implement the biennial workplan), which represents 0.1% of the organization's \$6.72 billion budget for budget, and therefore significantly below the range recommended by the JIU and no increase from the percentage reported in the 2017 review of the evaluation policy. In contrast, the spend in other agencies varies from between \$10m and \$60m (and in ILO and FAO 0.8% of the organizational budget), including agencies with less of a global mandate/footprint than WHO.

In terms of resources for decentralized evaluations (and supporting human resources), EVL does not currently have a way of tracking this expenditure as it comes out of Regional Office, Country or project budgets, meaning that the total resources spent on evaluation across the organization are unknown. Similarly, as mentioned, the costed workplan does not have a budget with regards to evaluation for emergencies and so it is unclear how much is intended to be spent in this regard.

Whilst the move to a costed work plan for 2024-25 is an important step forward, the evaluations and evaluations activities detailed in this alone equate to \$5.4 million, without staff costs, and whilst costed, it is not clear how this evaluation plan will be resourced and how EVL will advocate for sufficient resources to fulfil these ambitions given that EVL has a budget for 2024-25 of \$2.3 million for activity costs to implement the biennial workplan.

There are also important considerations around the allocation and management of EVL's budget which affect its independence. EVL does not have a separate budget line with the programme budget, but funding is allocated from within the DG's Office budget. The Director of Evaluation must make a request to spending above a certain threshold. There is currently no systematic country level resourcing for evaluation, nor monitoring of funds allocated, although EVL has managed to engage with a small number of WRs to undertake country evaluations (most predominantly in EMRO).

Independence

The 2018 policy provides clear definitions of impartiality and independence and certain provisions which help to support them. These include:

- The policy and work plan are approved at Executive Board level.
- The separation of evaluation as a distinct function from other oversight functions, a decision taken soon after the very first evaluation policy in 2012.
- The role of EVL and its director, separate from management and with a dual reporting line to the Director General and to the Executive Board via PBAC.

- Arrangements for appointing the Director of Evaluation, who can serve for 4 years, renewable once, without re-entry to WHO if appointed internally. The incumbent was appointed externally from FAO.

Based on practice in other comparator organizations and informed by UNEG norms and standards for evaluation functions, the role of the Director General, regional directors and the senior managers should be to help safeguard independence as well as to help to build a culture of evaluation. The senior managers have an important role to advocate for the function so it can develop, to allow space for evaluations to be conducted without undue interference on evaluation reports/findings; and to ensure that sufficient resources are available.

In the case of WHO, the culture and resources aspects are referred to in the policy - but the role in terms of protecting independence is not explicitly stated; this could be made more explicit. Comparing with other organizations, these roles are set out very clearly in the evaluation policies of all comparators. Having said, no problems with behavioural independence in practice have been experienced. The EVL commissioned reports have been able to cover important and challenging/sensitive topics – such as WHO reforms, GPW13, FENSA - without fear or favour.

In one important respect, structural independence is not fully in place. The lack of a clear budget line means that the Director of Evaluation does not know in advance what resources are available for the work plan and the budget for individual evaluations must be approved on a case-by-case basis, when above a certain threshold. Whilst the Director of EVL can in principle select subjects for evaluation, without the associated budgetary control in reality this is not done.

This contrasts with the position in all the comparator organizations. It is clearly undesirable that the Director of Evaluation should have to negotiate for resources during the biennium, to deliver on agreed topics in the workplan already approved by the Executive Board. This could lead to a situation where decisions on what to evaluate and when, would be affected by likelihood of having sufficient resources. It must be noted, however, that even in organizations with a very explicit clear budget line and committed resource budgets this is an issue, as shifts in actual resource positions during the planning period do occur. The level of oversight of the evaluation function is also noted as a challenge; PBAC members interviewed did not feel that there was a sufficient time to provide in depth oversight of the evaluation function given the breadth of their role.

An important aspect of independence is the presence or otherwise of regular, independent assessment. The policy and practice of evaluation in WHO has not yet been subject to a full, independent UNEG peer review. In all other organizations looked at in this study, this has become standard practice on average every 5 years, either through a UNEG peer review, or a full independent evaluation of the function. It is one of the main entry points for strengthening those functions, including when new directors of evaluation are appointed since they can use peer reviews to take stock and identify areas for strategic improvements.

In WHO, the MOPAN assessments have covered some of this ground, a light touch peer review was conducted in 2017; and the present comparability study provides a platform for covering similar issues. In future it would be important to move to the full peer review process.

No concerns were noted with regards to behavioural independence in interviews or in documents reviewed.

Quality

WHO's Evaluation Policy sets out that quality will be ensured through:

- (i) the continuous adherence to WHO evaluation methodology as elaborated in the WHO evaluation practice handbook, the applicable guidelines and the 2016 norms and standards for evaluation of the United Nations Evaluation Group;

(b) an independent quality assurance mechanism for all decentralized evaluations; and

(c) independent quality assessment of corporate and decentralized final evaluation reports. It is intended to cover both the evaluation process and products.

The quality and credibility of EVL commissioned evaluations is generally seen as high, following clear processes based on the policy and handbook. The 2017-18 MOPAN assessment confirmed that WHO corporate evaluations use appropriate methodologies for data-collection, analysis and interpretation and present the evidence in a complete and balanced way. The independent review of the Evaluation Policy in 2017 also found that the quality and credibility of corporate evaluations had a good start and had generally been done to an adequate and sometimes high standard. The 2022 (self-reported) UNSWAP report found that WHO exceeds requirements with regards to the integration of gender and human rights considerations in evaluation.

The quality assurance of centralized evaluations and some decentralized evaluations seems in most cases to be supported by EVL, either as part of the workplan or through more ad hoc backstopping of evaluation managers and therefore is undertaken independently of evaluation teams. Whilst WHO does not yet have a formal QA system, it does have certain arrangements that help to support quality and ensure credibility. **Specifically:**

- The evaluation handbook, guidelines and how to notes provide some guidance on conducting evaluations, for example at country level. A quality assurance checklist was developed in 2023 and is being used by EVL staff and consultants and shared with consultants working on independent evaluations.
- EVL has a roster of evaluation experts, who can support individual evaluation managers and also provide quality assurance of evaluation products.
- EVL staff also help to ensure and backstop quality for a selection of decentralized evaluations. However, this is somewhat ad hoc and is constrained by the limited number of staff to cover the relevant evaluations.
- The Chief Evaluator in PAHO supports quality for evaluations commissioned in the region of the Americas. In SEARO, an evaluation committee has been established. In other regions, the evaluation leads, and M&E focal points help to coordinate evaluation work although their work is not solely about evaluation and capacity is stretched.
- The Global Network on evaluation has been relaunched and is appreciated, which provides a forum for discussing what is planned in terms of decentralized evaluations and supporting the awareness of good practice.
- More recently, procurement has also been enhanced by introduction of LTA for companies.

However, specific aspects missing in WHO compared with other organizations are:

- A systematic independent quality assurance and reporting system with a full set of defined processes within a well-organized and accessible platform.
- A rating process implemented by external evaluators, to track quality of evaluations over time and inform delivery of consistent standards across the organization.
- WHE focal points for evaluation are not yet integrated into the Global Network for Evaluation which presents a missed opportunity for lesson learning, supporting quality standards and for EVL to engage with and support WHE's work.

- There is no specific guidance or tools for humanitarian or evaluation of emergencies.
- An up-to-date set of tools and templates covering the full range of different types of evaluations is not yet fully in place. EVL has recently developed, and is piloting the use of SOPs and checklists, but these are yet to be shared and used more widely throughout WHO.
- An annual synthesis of findings from evaluations and reporting on progress in enhancing evaluation quality.

In terms of decentralized evaluations, there is not yet a comprehensive QA system as seen in other agencies. For example, WFP (CEQAS, DEQAS) and in UNICEF (GEROS) have well-developed systems covering both levels, including assessment during the evaluation process and ex-post scoring and reporting, accompanied by professional quality enhancement support. Similar arrangements are seen in different ways in all the comparator organizations considered for this study, although they are more recent in some cases such as UNHCR.

The role of evaluation professionals and networks in building skills and awareness for evaluation managers and other staff is also key to supporting quality, as in the case of UNDP and ILO who have very well-established and comprehensive learning programmes with this aim. For WHO's decentralized evaluations however, there is no structured process whereby evaluation managers are trained or selected and may be programmatic staff with limited experience or understanding of evaluation.

The quality and credibility of decentralized evaluations is in some cases assessed by Regional Evaluation Officers/focal points, but because in some cases, these roles are overloaded, consistent professional support is not yet sufficient (with EVL being stretched to do this from the global level and only limited regional level staffing), this is likely to be inconsistent and not undertaken systematically. There are not currently checklists, templates, etc. shared across the organization to maintain consistent quality standards (although draft QA checklists for TORs, inception and evaluation reports has been developed in 2023 and currently being piloted). WHO is however, in the process of developing more structured checklists against which quality will be verified, which will be rolled out.

The policy is accompanied by guidance on quality via the implementation framework and recent guidelines e.g., a guidance note on integrating health equity, gender equality, disability inclusion and human rights in WHO evaluation. Recently, training has started on how to use these, via the Global Network on Evaluation. Last year EMRO conducted a Training on Evaluation Management in WHO context, with the UN staff college, aimed at building the capacity of at least one focal point from each division in the region and each EMRO country office.

The provision of further guidance and training to ensure evaluation quality is positive, and templates have been developed in several areas including gender, equity and human rights (GEHR). Going forward, it would be helpful to ensure this guidance is well socialized across potential evaluation commissioners and managers as has happened in other agencies.

Enabling environment and culture

The process of **building a culture of evaluation**¹⁴ across WHO is rather nascent. This includes identifying and encouraging demand for evaluations that are useful for learning, accountability and management decision making in support of performance improvement and impact.

¹⁴ An evaluative culture in WHO would be characterized by strong demand for evaluation to support learning and self-reflection, achieved through leadership, organizational systems and structures and a learning focus in how WHO works. Going beyond this, it can also refer to a willingness to experiment, take risks with the aim of innovation and improvement, and a tolerance for and ability to learn from mistakes.

A significant positive development has been hiring of seasoned evaluation professionals with external experience to lead the function, meaning that it is now well positioned to develop under new leadership and is informed by good practice externally.

Access to the governing body is well established and provides an important entry point in the sense of annual reporting to the EB via PBAC. However, at the moment this is entirely process-focused and in other organizations there is dedicated space for substantive discussion of evaluations through an evaluation committee (e.g., WFP). The depth of engagement with the governing body needs to be enhanced, including substantive discussion of evaluation findings and recommendations. Understanding of Member States in how to use and advocate for evaluation remains to be developed.

This is particularly important at a regional level, where leadership for evaluation could play a very strong role, building on the experience in certain regions (PAHO) where this is already well established. In UNDP (for example) the regions and senior management play a formal role in ensuring that evaluation is resourced and safeguarded.

The level of demand for and understanding of evaluation across the three levels of WHO is variable. There are positive signs in HQ in certain areas such as health emergencies and selected technical departments which have benefitted from evaluations. Examples which have been identified in this study as having been influential include evaluations of GPW13, PHC and Normative Functions, GEHR, yellow fever and polio. However, at the country level there is very limited visibility and awareness of evaluations, with a few positive exceptions where individual WRs are starting to request evaluations (Djibouti, Iraq, Tunisia and Yemen). Stakeholders also outlined that there is often a confusion between evaluation and audit across the organization, with evaluation often being perceived as serving primarily an accountability purpose rather than to fulfil learning needs.

However, several stakeholders outlined that there is a lack of senior-leadership buy-in and support to evaluation and that this has led to it being deprioritized and influenced evaluation use and impact. It will be important going forward to nurture and engage senior staff across the organization to be evaluation champions to strengthen the organizational evaluation culture.

Evaluation use and impact

One key opportunity, consistently emphasized by leaders across the organization, is demonstrating that **evaluation can and does add value, through independence, credibility and how it is used for decision making**. Ultimately, the reason for having an evaluation function is to support organizational learning and improvement on key strategic areas, which depends on the evaluations being used, as well as to support accountability and compliance.

- Ensuring utility and impact depends on success in the following areas:
- Raising awareness and understanding of the evaluation function, positioning it as influential and adding value in the organization.
- Ensuring that it is clearly differentiated from other related areas include monitoring, audit, knowledge management, organizational learning and research.
- Delivering evaluations that are relevant to key decisions, high quality, credible and useful – therefore influential.
- Effective dissemination, using a range of tools and formats for different audiences, beyond the basics of publishing the reports.
- Knowledge management and curation of evidence from previous evaluations, accompanied by synthesis of lessons.

In comparator organizations, substantial progress has been achieved through considerable investment of time and energy to link the evaluation functions to the needs of strategic decision makers from the governing body downwards. Evaluation is rather visible, respected and strategically positioned for impact in organizations such as UNDP, UNHCR, ILO and WFP – indeed in all the comparator organizations there has been progress, noting that historically evaluation functions in international organizations have not always been influential or ‘on the map’. Even so, ensuring that evaluations are used and stay relevant is a constant challenge.

In WHO, there is an urgent need to establish a similar reputation by **closing the learning loop** - to show what evaluations can deliver, show their value added and utility, and build credibility of the function. When interviewed for this study, key stakeholders felt that EVL has generated interesting, relevant and potentially useful reports but the widespread perception is that very little has happened with them.

The experience is also that the same rather generic recommendations emerge across evaluations repeatedly. Developing effective recommendations has been a challenge in most organizations that undertake evaluations, and there is plenty to learn from their experience of improving this aspect e.g. using co-creation processes, reducing the number of recommendations, being specific about who they are aimed at etc.

In terms of raising awareness and understanding of evaluation, the policy and implementation frameworks set out clearly what evaluation is and its unique role. Nevertheless, the understanding of these definitions and the socialization of the statements in the policy across the organization is incomplete.

More positively, in terms of choice of topics, EVL has at the global level delivered on some strategically important evaluations that are clearly highly relevant. The ongoing GPW13 evaluation is a positive example of how evaluation can add value to strategic decision making by focusing on areas of key interest to the organization. This provides a platform to build on and, over time, helps to increase the influence of the function and build its reputation.

One way that utility can be enhanced further is to prioritize evaluations that **focus on WHO contribution and learning**, as opposed to systems and accountability. In line with the focus of GPW13 and WHO’s overall strategy, learning about **WHO’s impact at country level** is clearly a high priority – or needs to be. There is only limited opportunity for countries to learn from global evaluations, except through material contained in case studies. In that respect, the ongoing programme of evaluations of WHO’s contribution at country level is an important step, including the recent guidance on how and why they should be done. As more of these are completed, regular synthesis of lessons will be important, supported by efforts to disseminate those lessons and monitor how countries and ROs are using these in planning and programming.

The approach to dissemination of evaluations is fairly basic at present, which is not surprising given that human resources are stretched with actually doing the evaluations. Evaluations are published on the website, accompanied by an evaluation brief and management response. By contrast, the approach to communication of findings in comparator agencies is now quite sophisticated and strategic. WHO is working to advance its practice in this area though; with a range of new formats and techniques for communicating evaluations, ranging from data visualization and infographics to use of social media, videos and web-based electronic formats. It is also encouraging that EVL has recruited a communications consultant to develop an evaluation dissemination strategy and develop improved evaluation products going forward.

On the knowledge management side, EVL has put in place a dedicated resource for organizational learning, recognizing the importance of this function. This is a senior post, but without additional support and it needs to cover engagement with other processes such as MOPAN or JIU assessments.

Engagement and follow-up

A key aspect here is having an effective system for tracking **Follow up on Evaluation Findings and Recommendations**.

In comparator organizations, the responsibility for this is clearly set out in the policy, including requirements on management responses and who is accountable for actions to what timescales. Importantly, the responsibility rests with management to take ownership of follow up actions.

This is formalized and systematic, linked to a repository of evaluations that includes not only the evaluations but also management responses and action plans, and supported by regular reporting to the governing body. For example, in UNDP it is possible to see what percentage of actions have been addressed across the organization, how this has changed over time and how it varies by region.

In WHO, management responses are produced and published alongside the evaluation on the EVL website; and there is regular reporting to PBAC on evaluations completed. A template and guidance have been developed to ensure management responses are more effective. However, a strong message from interviews for this study is that there is a lack of institutional incentives for action once an evaluation is completed – in some cases they are indeed acted on (as for example with the evaluation of gender and human rights), but the decision largely or even entirely up to the individual manager to judge rather than organizationally driven. The 2023 Evaluation report highlighted that neither the Organization-wide evaluation work plan nor the Organization-wide repository of evaluation plans and reports systematically capture all decentralized evaluations that are conducted across the Organization, thus preventing effective tracking of management responses and implementation of recommendations of completed evaluations.

Of course, a compliance approach on its own would not be sufficient, this needs to link back to how evaluation and learning is regarded (culture) and ownership of the evaluation process. A significant gap is having a dedicated time and space for substantive discussion of evaluation findings with the governing body and with senior management. For example, in WFP there is a committee of the leadership group which takes receipt of strategic evaluations and in UNICEF there are clear channels for discussion on draft findings with directors.

Overview of UN agency comparators

This section summarises key features on the evaluation functions of the eight UN agencies which are the comparators selected for this study.

- FAO
- ILO
- UNDP
- UNESCO
- UNFPA
- UNHCR
- UNICEF
- WFP

This material is based on documentary review focusing on the evaluation strategies, annual reports and policies, as well as on MOPAN reports and UNEG peer reviews where relevant or recently available. Full summaries for each agency can be found in Annex 3.

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
Overall maturity and trajectory	<p>Relatively mature function in most respects, but still evolving on decentralized evaluation. The 2016 independent evaluation found significant progress in the evaluation function...but also noted that the quality of evaluations is uneven and there is a critical gap in decentralized evaluation. More recent work is aimed at addressing these gaps, see below. External evaluation conducted in 2016. Recommendations now being followed up through interim strategy. Last peer review was 2012.</p>	<p>Mature evaluation function. Most recent Independent evaluation of the evaluation function (IEE) (2022) notes high level of effectiveness. Achieved or partially achieved 95 per cent of intended outcomes...and that. evaluation function is meeting the expected norms and standards for credibility, quality and independence. However, it has not yet fully attained expectations for utility. A particular strength to learn from ILO is around professional training of evaluation managers. Regular independent evaluations of the evaluation function are conducted every 5 years, latest in 2022.</p>	<p>Well established evaluation function. UNDP played a leading role in establishing norms and standards by supporting UNEG. Mature and well-resourced function, with strong coverage at all levels. Particularly strong on DE coverage and on evaluation capacity development. Has a particularly rich set of building blocks for synthesis of evaluation. Policy says peer reviews should be conducted every 4 years. Most recent was in 2012/13. Independent review of Evaluation Policy in 2019: this</p>	<p>UNESCO's corporate evaluation function is relatively small co-located with internal oversight services, although it 'operates separately'. 2019 peer review found it was "mature and well-grounded", although the DE function had some way to go to reach maturity. Peer review was conducted in 2020. It found that UNESCO has a high-quality evaluation service. Peer review will be conducted again in 2024.</p>	<p>One of the smaller and younger (2009) evaluation functions but strongly capable and 'increasingly robust'. Highly valued in UNFPA and by the Executive Board for its value added. Though relatively small, it has a high profile via its work on system-wide, inter-agency and joint valuations. Peer review was undertaken in 2023 for the first time. All recommendations were accepted by UNFPA management.</p>	<p>A very young evaluation service established when the dedicated evaluation service was set up in 2016, and the first external head was appointed in 2017. It has quickly progressed and is now performing to a high standard according to the peer review. May offer useful lessons for WHO because of similarity in size of the function and progress already made. Peer review was undertaken in 2021 and considered that the evaluation function in UNHCR has made impressive progress in the last few years</p>	<p>Mature and well-resourced, well-functioning. 2023 Peer review noted: progress towards policy aims is broadly good; high-quality, useful and credible evaluations. Areas to enhance include enabling environment and accountability and monitoring of roles and responsibilities, synergies with other knowledge functions. Peer review conducted in 2023, and regularly every 5 years since 2008, with strong follow up on recommendations in updated policy.</p>	<p>Mature and well-functioning, one of the largest of all UN agencies. Peer review found that all aspects of WFP's evaluation architecture are well-articulated, and the governance structure is effective... It also found that WFP is achieving its policy aims on evaluation. Peer review conducted in May 2021 and also previously in 2007 and 2014. The first UN agency to undergo a UNEG peer review. Good follow up on recommendations.</p>

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
			recommended action on DE quality among other areas.					
Evaluation Policy	The new Interim Strategy 2023-25 approved in 2023 seeks to address gaps identified by independent evaluation. 2010 Charter is now quite dated – intention is to update policy in next couple of years. Handbook (2015) provides guidelines for staff.	2017 Policy and more recently the Strategy 2018-23 have provided a clear and detailed overall framework for evaluation in ILO, including both centralized and decentralized evaluation. A results-based strategy 2023-25 is now in outline form, responding to the latest context and findings of the 2022 IEE.	2016 updated Evaluation Policy provides very clear framework setting out roles and expectations at all levels. Audit and Evaluation Advisory committee provides additional oversight.	Evaluation Office reports through the Director of IOS. The new Evaluation policy 2022-29 includes a stronger focus on strengthening DE, in response to the peer review.	Revised evaluation policy was approved in 2019 and then reviewed and updated in 2024. Peer review notes the strong enabling environment in UNFPA, which provides a firm foundation for further strengthening the evaluation function.	The evaluation policy was revised in 2016 and an Evaluation Strategy was adopted in 2018-22 to operationalize the policy.	2023 Draft Policy is the latest in a series updated every 5 years drawing on recommendations of regular peer reviews. Complemented by a formal Procedure for Implementation of Evaluation Policy.	Policy regularly updated – latest is 2022. It.... sets out a clear vision and purpose. WFP's evaluation strategy sets out a phased implementation plan. Regional evaluation strategies mirror the structured approach at global level.
Coverage	Prioritization process set out in the new interim strategy uses clear selection criteria for deciding what evaluations will be conducted. Coverage norms are not mentioned in the Charter or Interim Strategy (check) although balanced coverage is one of the criteria for selection.	Clear evaluation policy requirements on coverage are being met. Coverage for internal evaluations need to improve further. Dominance of project evaluation: 81% of the portfolio...unsustainable workload, number of evaluations needs to be reduced without compromising accountability	545 decentralized evaluations were planned for 2022, completed 73% of these, nearly 400 in total. Over 50 strategic evaluations (about half were independent country evaluations).	Evaluation policy refers to the aim of coverage of all strategic outcomes, priorities and cross-cutting themes over the 8-year period covered by the UNESCO Medium Term Strategy.	Policy sets out coverage norms, including for country programmes, policies and strategies. Around 5 centralized evaluations and 20 regional and country evaluations per year. A CPE is required at least once in every two programme cycles.	In terms of coverage, and in line with the new Evaluation Policy, all major policies, themes, strategic results areas, and geographies of the organization's operational work will now be evaluated at least once over a 5 to 10-year period.	Diverse and growing...has almost doubled from 107 in 2018 to 199 in 2022. Policy sets out very clear and detailed coverage norms including for frequency of country programme and country thematic evaluations and norms for various corporate, humanitarian and joint evaluations.	Has increased centralized and decentralized evaluations, achieving the coverage norms. 16 out of 26 policies have been evaluated, OEV on track with the roll out of country strategic plan evaluations. 40 percent of country offices have conducted at least one decentralized evaluation
Quality	QA framework in place - but until recently did not apply criteria and checklists – QA	The evaluation function benefits from multiple layers of quality assurance, including ex -post assessment of quality, supported by up-to-date, comprehensive	All IEO evaluations are peer reviewed by thematic and	Quality of CE and DE is reported on in annual synthesis of	The evaluation office supports quality through technical guidance and	External assessments have shown that the quality of evaluations has	GEROS system in place and well established – provides the focus for a highly	CEQAS (centralized) and DEQAS (decentralized) evaluation quality systems are well established with rich data. They provide a

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
	approach is now being expanded	quality guidelines. The ILO evaluation function has quality guidelines for the conduct and quality assurance for centralized and decentralized evaluations. ILO's guidance and approach on assuring the quality of decentralized evaluation has evolved and expanded over time.	country-level experts, hired externally. Evaluation-related standards are included in the quality assurance checklist for new CPD formulation and operational guidance. IEO operates the Evaluation Excellence Awards	evaluation reports. Quality Assessment Template based on UNEG guidelines with UNESCO-specific checklist with criteria for Evaluation managers.	tools, capacity building, internal quality assurance and external quality assessments. However, the peer review recommended to strengthen the quality assurance and assessment system and expand its coverage to all decentralized evaluations.	improved from 58% in 2018-19 to 70% in 2021. The Evaluation Office has revised and expanded its external quality assurance guidance and tools.	systematic quality assessment of all evaluations with annual reporting. GEROs is seen as a key enabler for quality of evaluations.	highly systematic approach to quality with tools, templates checklists and annual reporting. Includes templates, tools and guidance notes.
Relevance and responsive-ness	Work plan of evaluations 2022-25 update was presented to EB in Nov 2022, having been approved in the previous year. Seeks to reposition FAO's Office of Evaluation's (OED) work. Does not appear to have resource plan included (check if it is a separate document).	Centralized evaluations (decent work country programme evaluations, high level evaluations (HLE), and thematic evaluations) form a rolling four-year work programme. The work programme for HLE topics is the product of annual informal consultations.	IEO has a multi-year workplan (2022-25) approved by the Board with evaluation budget. It focuses on evaluations by IEO. Many DEs are conducted (over 350 in 2021). Need to check if there is a DE workplan.	'Rolling plan; agreed with ADGs and responding to requests from EB...however, self-assessment is that this is not an optimal approach, needs a more strategic approach.	Planning of strategic and centralized evaluations is based around a Quadrennial budgeted evaluation plan, using clear criteria for prioritizing evaluations.	Biennial workplan covers both global independent (centralized evaluations) and management commissioned global, regional and country evaluations (decentralized) evaluations. Does not show costs, as these are separately covered in strategy.	Plan for global evaluations is agreed every 4 years, covering corporate evaluations. This is approved by the EB and includes detailed resources. For DE, costed evaluation plans are annexed to CPDs.	4-year work programme is presented to the EB for approval, aligned to management plan. Includes considerable detail on resourcing of the evaluation function. Clear process of consulting on workplan.
Enabling Environment and Culture	OED's interim evaluation strategy sets out an initial architecture for excellence in	Considered to be a conducive culture within ILO to use evaluations to assess impact and a shared commitment to compliance with evaluation responsibilities. Volume of		UNESCO is perceived as having a relatively strong culture with regards to		The evaluation culture is nascent given its infancy and there is a lack of familiarity with evaluation. The		

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
	supporting FAO Members and Management's culture of evidence in decision-making and outlines a shared evaluation culture as one of the conditions for its success. There is considered to be good receptivity to evaluation across the organization.	evaluations, and time constraints of staff though mean that there was scope to improve the use of evaluation for learning purposes. The Evaluation Manager Certification Training Programme is thought to have promoted an evaluation culture in ILO. Good engagement from the governing body and management with evaluation findings.		evaluation. The IOS-EV Office delivers monthly webinars on evaluation and targeted and in-depth training events.		evaluation function is working to promote the use of evaluation as a management tool by trying to build relationships with UNHCR senior management.		
Independence	Evaluation Charter highlights independence as one of the primary principles underpinning evaluation. Evaluations. A number of areas have been noted where OED's independence could be strengthened (reporting lines, a lack of guidance).	Independence of the ILO evaluation rated as 'highly satisfactory' in its most recent MOPAN report with strong operational and financial independence. Direct reporting lines to the Director-General and the Governing Body, with structural independence.	The Independent Evaluation Office is a functionally independent unit within UNDP that supports the oversight and accountability functions. The UNDP Administrator safeguards the integrity of the evaluation function, ensuring its independence from operational management and activities	The 2017-18 MOPAN report highlighted the independence of UNESCO's evaluation function as 'Highly Satisfactory'. Structurally, the Evaluation Office is positioned independently from UNESCO management and programmatic functions within IOS, alongside audit and investigation functions. The Evaluation Office has operational independence	The last MOPAN assessment (2017-18) had already found that evaluation independence was well established, accompanied by good quality assurance mechanisms and planning of evaluations to cover all strategic areas.	The Head of the evaluation function reports to the High Commissioner; and the High Commissioner, in principle, approves the budget and workplan; whilst this falls short of full functional independence as envisaged by UNEG norms and standards, it considered that in practice it is adequate as the Head of Service has sufficient autonomy to manage a pre-agreed budget, select evaluands, and issue	The MOPAN assessment 2020 had also found (similarly) that UNICEF has a strong evaluation function with a high degree of independence.	The UNEG Peer Review (2021) concluded that at WFP a "highly strategic independent corporate evaluation function oversees the production of high-quality centralized and decentralized evaluations".

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
				and a discreet budget allocation.		evaluation reports.		
Evaluation Use and Impact	Strategy refers to promoting use of evaluations through a wider range of evaluation products, including briefs. Various dissemination activities including workshops, events, online presence, briefs and infographics, blogs etc.	ILO has an Evaluation Advisory Committee which was established to provide a mechanism for overseeing the use and implementation of follow-up to lessons learned and recommendations resulting from the ILO's evaluation activities. Its objective is to promote institutional follow-up on independent evaluation findings and recommendations.	Evaluation Resource Centre (ERC) provides a platform for knowledge management of all UNDP evaluations. Recently revamped. Now includes methods and tools and guidance as well as evaluations. Dedicated section of IEO focuses on synthesis and lessons from evaluation.	Tailored communication strategies are a requirement in evaluation design. All evaluations, require a Management Response and Action Plan. The IOS EO monitors implementation of recs. An Evaluation Reference Group is a key mechanism for stakeholder engagement to ensure use.	The Evaluation Office is respected across the organization for its professionalism and its added value for decision making. There is a global management response tracking system, which sits with the policy and strategy division as custodians on behalf of management.	New policy (2022) requires management to report on implementation of accepted recommendations. UNHCR's Global Learning and Development Center is a key part of the learning system. Evaluation office has invested in strategic communications.	Peer review suggests areas for strengthening learning. There is a high level of formal compliance with management response processes. A more user-friendly digital repository of evidence is suggested by the peer review.	The updated 2022 Evaluation Policy includes a new outcome on use of evaluation evidence explicitly monitored. Specific actions around tailoring of evaluation products, processes for integrating evidence into programmes and policies. The EB considers all OEV commissioned evaluations and their management responses.
Resourcing - financial	Charter states that evaluation budget will attain at least 0.8% of the total Regular Programme Budget.	The operational budget for Oversight and Evaluation (Evaluation Office EVAL only) in 20-21 amounted to \$8.4m. Of this, roughly \$3m was allocated to the Evaluation Office (EVAL), \$3.4m went to Internal. This does not include resources/staffing for evaluation at regional level).	Policy norm for resourcing is 1% of combined programmatic spend. In 2022 he IEO spent \$11.83 m on evaluations and other activities. Funded both from regular and other resources. Expenditure on	Budget for evaluation had been modest and falling. Now \$2.1m for staff costs and \$1m for corporate evaluations. New focus on policy commitment of 3% of programme expenditure set aside for evaluation. Need for	\$9m pa spent on evaluation, of which \$4m centralized. Overall evaluation spending, as a percentage of UNFPA total programme expenses, increased from 0.45% in 2014, to 0.83% in 2021. This is below the 1.4-3.0 per cent target range (as recommended by	Evaluation budget is submitted as a separate line to the Executive Committee and then approved by High Commissioner. Budget has increased from \$1.58m in 2013, to \$4.2m in 2015 to \$6.2m in 2021.	The credibility of the evaluation function has been "strengthened by investments in human/ financial resources." Between 2018 and 2021 evaluation spend increased from \$50m to \$65m. Overall target for evaluation resources is 1% of program	Very well resourced both for centralized and decentralized and impact evaluations, through a range of sources. Since the adoption of the policy the financial resources available for the evaluation function have almost tripled. In 2020 USD 26.02 million, or 0.31 percent of WFP's contribution income.

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
			decentralized evaluations was a further \$24m. Since 2017, resources have increased from 0.48 % to 0.77 % of UNDP program.	adequate resourcing and fungibility.	JIU) and remains an issue of debate within UNFPA.			
Human resources	Following the independent evaluation, OED is strengthening evaluation capacity by addressing gaps in technical expertise including at regional level, by the creation of regional evaluation teams.	With all quality assurance taking place 'in house' by a relatively small evaluation office, staffing is quite stretched. From 2018 to 2021, progressively higher numbers of ILO staff were trained as evaluation managers or internal evaluators - an important part of EVAL's hybrid, decentralized evaluation system. A suite of training programme was developed.	Human resources. In 2022, the IEO maintained 35 posts, including 27 professionals and 8 General Service staff, with strong arrangements for continuing professional development. In addition, there are significant staffing resources at regional and country level for M&E. Roadmap includes plans for 6 regional evaluation advisers (check with UNDP).	professional posts. DE are conducted by external consultants with support from EO if requested. Network of Evaluation Focal Points.	The Evaluation Office has strong technical capacities in evaluation, including 11 staff positions headed at D1 level. More capacity is needed on humanitarian Evaluation. There are 5 P5 regional evaluation adviser posts, reporting to the Regional Director/Deputy.	Has increased to 12 full time positions including 3 outposted as regional senior evaluation officers, to better support decentralized evaluations. Peer review found that Strong professional team. Global M&E community of practice provides mentoring support for over 200 staff.	Additional funding allowed regions to recruit multi-country evaluation specialists and all regions passed the 0.5 per cent funding level by 2021.	Peer review notes that WFP has increased the number of evaluation professionals in OEV and the regional bureau, enabling it to deliver its work programme to the required standard
Decentralized evaluation function	Started process of strengthening DE in 2020. Regional	Network of staffing at regional level to help provide management and ensure quality	UNDP has a Road Map for Strengthening	Decentralized evaluation policy	Still developing. Peer review found there is still	Continuing to expand regional presence to	Very strong decentralized function, with key	Well-resourced and structured system of decentralized evaluation,

Criteria	FAO	ILO	UNDP	UNESCO	UNFPA	UNHCR	UNICEF	WFP
	staff have been hired after delaying due to COVID-19.	of DE. Departmental focal points, regional evaluation officers and M&E officers. 144 certified evaluation managers among ILO staff.	Decentralized Evaluations (2022), may be useful learning resource. 10 regional evaluation focal points. Strong process of evaluation capacity development (ECD) for country office and programme staff. Evaluation Scorecard is a dashboard on DE performance.	framework is still being developed, according to the 2020 peer review. The IOS annual report notes that DE is a priority, with support to 45 DEs ongoing. Network of evaluation focal points.	room to improve the relevance quality and learning from decentralized evaluations. Regional M&E Advisers appear somewhat stretched given other responsibilities alongside support on evaluation.	support DE. 5 senior regional evaluation officer positions now established. DE workplan. 6 DE were conducted in 2022, with support from evaluation office.	role for regional evaluation units. Nine out of 10 of evaluation products are now managed at the country or regional level within a decentralized evaluation system (similar to other UN agencies such as UNDP and WFP).	supported by systems, regional evaluation units and coverage norms. Systems include the DEQAS tools and monitoring of quality.

Best practices from comparator agencies

	Good practices in comparator agencies for WHO to consider	Agency
Evaluation Policy	Up to date and comprehensive evaluation policy is in place, informed by external peer review or independent evaluation, and with a clear vision.	UNICEF, WFP, UNDP, UNFPA, UNESCO, UNHCR, ILO
	Clear accompanying strategy to operationalize implementation of the evaluation policy, with detailed results framework which is tracked and reported on.	UNHCR, FAO, ILO, WFP
	Lessons have informed policy on how to strengthen independence of evaluation function.	ILO
	Strategic approach to strengthening decentralized evaluations has been followed with strong regional ownership.	UNDP, WFP, UNICEF, ILO
	Well-established approach to country evaluations, linked to country planning cycles.	UNDP, WFP, UNICEF
	Clear coverage norms (including of major policies, themes, strategic results areas and geographies).	UNHCR, UNICEF, WFP, UNFPA, ILO, FAO
Enabling Environment and Culture	High level of institutionalization of evaluation function, supported by engagement with and from governing body and senior management, strong culture of evaluation.	All
	Clear governance structure and roles for evaluation, supported by board and senior mgt.	UNDP, WFP, UNICEF
	High level of value added and respect from across the organization, well supported by interactions with governing body.	UNFPA, WFP
	Certified Evaluation Manager Training programme to increase evaluation capacity and embed evaluation culture.	ILO
Human and Financial Resources	Investment in financial resources for evaluation and clear percentage target for funding, wholly or partly achieved.	UNICEF, UNESCO, UNDP, UNFPA
	Balanced approach to resourcing using a range of sources.	WFP
	Extensive programme of evaluation training for staff and country M&E focal points.	UNDP
	Established an evaluation trust fund to generate evaluation funding.	FAO, WFP
	Top up fund to add to evaluation for projects to improve quality.	UNHCR, FAO
	Strong capacity both centralized and in regional evaluation units.	WFP, UNDP, UNICEF
Quality	Fully developed system for quality assurance and enhancement.	UNICEF, WFP, UNDP and others
	Built high level of capacity in human resources and skills of IEO.	UNDP, UNICEF
	Clear criteria to guide the selection of evaluations.	FAO, UNHCR, ILO, WFP, UNICEF, UNDP, UNFPA
	Comprehensive system of evaluation quality assurance.	UNFPA, WFP, UNICEF, UNHCR
	Comprehensive guidelines and handbook to support quality.	UNDP
	Annual synthesis of evaluation finding (mapped to the SDGs) and of evaluation quality.	UNESCO
	Evaluation excellence awards to incentivize quality.	UNDP
Relevance and Responsiveness	Strong approach to evaluation planning, informed by strategic plan of the organization.	UNFPA, WFP
	High level of evaluation coverage including at DE level, with a diverse range of products aimed at learning.	UNICEF
	Use of evidence and meta-evaluation as building blocks for synthesis and strategic learning.	UNDP, ILO, UNICEF, WFP, UNESCO
	Evaluation resource centre repository of evidence and methods.	UNDP
	Wider contributions to evaluation capacity building and partnerships globally and nationally, through UNEG and ECD work.	All
Independence	Established dual reporting lines for Regional Evaluation Advisers.	FAO

	Good practices in comparator agencies for WHO to consider	Agency
Evaluation use and impact	Ensure designated budget time for evaluation and structural independence.	All
	System for tracking use of evidence and management responses.	UNICEF
	A strategy is in place to enhance evaluation use through communications and knowledge management was put in place includes a range of approaches such as infographics, videos and podcasts.	UNFPA
	Roll out of structured training on evaluation policy/standards to build capacity and understanding of evaluation and its use across the organization.	UNESCO
Engagement and Follow Up	Use of I-EVAL, web resources to showcase evaluation work.	ILO
	Ensure all evaluations include a communication plan which outlines how findings will be disseminated and targeted with user-friendly modalities.	UNESCO
	Mandatory 'action plan' to facilitate evaluation use and uptake of recommendations.	UNESCO

Conclusions and recommendations

Conclusions

WHO's evaluation function has evolved in key aspects since it was established, various steps having been taken to enhance the function in the period from 2012 – 2018. That process is ongoing and has been given new impetus recently. Notable milestones have included the first evaluation policy in 2012, setting up EVL as a separate function in 2014, updating the evaluation policy in 2018, bringing in evaluation professionals from outside from 2016 onwards, developing implementation frameworks including for country level evaluations and drafting a fully costed workplan in discussion with regions.

More recently the main areas where progress is visible are particularly around delivering high quality, relevant centralized evaluation helped by bringing in more professional staff, developing more detailed guidance in key areas, relaunching the global evaluation network, planning evaluations with the regions and moving to an LTA for procurement of support from evaluation providers.

However, in certain other respects such as financial and human resourcing, there has been little to no improvement. The scope to build systems for decentralized evaluation has been limited, not least because of the need to put work on hold during the COVID-19 pandemic when WHO was hugely stretched given its important role in supporting the response.

Over the last 10 years (since the JIU report), other UN agencies including those used as comparators for this study have made substantial investments in evaluation and progress in systems and practices. Consequently, WHO has now clearly fallen behind what is the norm among comparable UN organizations of the stature and importance of WHO. A well-functioning evaluation system is integral for the WHO to achieve the goals of GPW13 and the forthcoming GPW14, concerning to WHO's increasing emphasis on "delivery for impact" and country-focus.

The recommendations below set out several of ways that WHO can bring its evaluation function up to the standard of current UN practice in terms of policy, systems, and practice.

Recommendations

	WHO Finding	Comparator agency practices	WHO Recommendation
Policy	The 2018 evaluation policy was a step forward at the time but still has gaps in key areas (e.g., on DE, resourcing, coverage, emergencies and humanitarian evaluation) and needs updating.	Evaluation policies are regularly updated every 5 years, informed by UNEG peer reviews or independent evaluations of the function.	R1a. <u>Update the evaluation policy</u> , informed by this comparator study, addressing the gaps noted in this study. R1b. Keep the evaluation policy regularly updated, informed by UNEG <u>peer reviews or independent evaluations, every 5 years</u> .
	Coverage norms are absent from the policy and implementation frameworks.	Detailed and explicit coverage norms are typically embedded in evaluation policies as the norm.	R2. Develop explicit <u>coverage norms</u> in key areas of WHO work, and track progress annually ¹⁵ , relating them to strategic outcomes for WHO. These should include: (a) balanced coverage of goals in GPW14 (b) balanced coverage of WHA resolutions (c) balanced coverage of WHO policies (d) 100% coverage of global health emergencies where WHO is leading the international response (e) active WHO participation in inter-agency evaluations of level 3 humanitarian crises. These should be updated every 3 years in the case of protracted crises (e) 100% coverage of all country strategies in the year prior to their revision, for those countries with off track health SDGs and/or high levels of risk identified in programming and audits. Coverage at least once every 2 programme cycles for all other country strategies.
	The 2022 implementation framework is useful and detailed in many respects but is missing a results framework and inclusion of considerations for humanitarian and evaluation of emergencies.	Evaluation policies are typically accompanied by formally approved strategies with detailed planning and results monitoring to track delivery.	R3. Set out a fully developed roadmap on how the evaluation policy will be delivered. This should be accompanied by a detailed <u>results reporting arrangement including key performance indicators</u> to monitor how the function is improving in terms of quality, credibility, capability, resourcing and use.
Relevance and Responsiveness	The evaluation work planning process is consultative but not fully developed. Evaluations are typically process focused. WHE processes for planning and implementing evaluation are not well linked or integrated with EVL.	Comparator agencies typically have a rolling 3- or 4-year work programme based on extensive consultation and clear criteria, aligned with organizational strategies.	R4. Further develop and embed an institutional approach to work planning (including with WHE) and consulting on <u>selecting evaluation topics</u> , informed by the coverage norms as per R2.

¹⁵ See for example Table 2 in WFP's 2022 evaluation policy: https://executiveboard.wfp.org/document_download/WFP-0000135899 or table 2 in UNICEF's 2023 evaluation policy, <https://www.unicef.org/media/54816/file>

	WHO Finding	Comparator agency practices	WHO Recommendation
Independence	Structural independence is not fully in place as there is a lack of a clear budget line for the evaluation function.	Other agencies report a discreet budget line for evaluation with independence regarding its use.	R5. By agreement with Management and the Governing body as appropriate, set out an explicit budget line for evaluation so that the costed workplan and the supporting function can be delivered; and to underpin structural independence of the function.
	The level of oversight of the evaluation function from WHO's governance body/PBAC is noted as a challenge	Other agencies report more engagement and oversight from governing bodies.	R6. Given that PBAC and Regional Committees are being restructured, there is an opportunity to strengthen the prominence of evaluation and increase evaluation oversight <u>by (a) having informal sessions with Member States where EVL presents the evaluation report and management present their response (b) an internal evaluation committee in which WHO senior management discusses and approves the evaluation and the management response.</u> <u>The internal evaluation committee should also receive annual reports from EVL on progress against results outlined in the (recommended above) evaluation roadmap</u> (i.e. evaluation planning, implementation, resourcing, quality, evaluation culture) and then make recommendations to PBAC on necessary actions to deliver the policy.
Quality	Coverage of DE is unclear due to lack of complete data, quality is not yet monitored. There are no specific tools or guidance to inform humanitarian evaluation.	Quality and coverage of decentralized evaluations has steadily improved, and in some agencies are the largest component of evaluation.	R7. Monitor the coverage and quality of decentralized evaluations and take steps to map the existing coverage more reliably and enhance the quality of DE over time. This should be supported through a well-developed approach to training and capacity building learning from the approach adopted in ILO, UNDP, WFP and others. Ensure that specific guidance is available to inform quality of humanitarian and evaluation of emergencies.
	Elements of a quality assurance and enhancement system are in place, but not fully systematized.	Comparator agencies such as WFP, UNICEF and UNDP have well-established systems for quality enhancement and reporting.	R8. Consolidate existing elements such as QA checklists and templates already developed by EVL, into a fully developed QA system , with annual reporting including ex-post QA scoring for centralized and decentralized evaluations (including WHE evaluations). A good model here is the approach developed in WFP. A staged approach is recommended, starting with a system at central level closely followed within 2 years by expanding to cover decentralized evaluations.
Financial and Human Resources	The resourcing of the evaluation function is not clear and is significantly below what is needed given the size, scope and complexity of WHO's work.	An explicit level of resourcing is agreed in advance when the workplan is approved, to ensure it can be delivered.	See R5 above under independence, recommending establishing an explicit budget line for evaluation. R9. Ensure that there are dedicated 'draw down' resources available for humanitarian and evaluation of emergency and that these are articulated in the costed workplan. This should include identifying an agreed percentage (e.g. 1%) of voluntary contributions for emergencies that is dedicated to evaluating the emergency response.
	A policy target on resourcing was implied in the 2018 policy but has not had traction. Resourcing is of the order of 0.1% of WHO resources.	Explicit targets are in place on resourcing the evaluation function and progress is tracked, typically of the order of 1% of overall resources.	R10. Commit to a target level of resourcing as a percentage of WHO expenditure – at least 0.5% as was intended in the 2018 policy and not yet been implemented. The target should be set out in the revised policy at the level of 1.0%, which is more in line with the norm across the UN, and tracking of this should be related to progress against coverage norms and the resources required for each norm. This should also be accompanied by a clear and explicit statement in the revised evaluation policy of responsibilities of Management and Member States to ensure that necessary resources are made available for evaluation to meet the agreed standards.
	The decentralized evaluation function in WHO is under-developed in key aspects, despite WHO's organizational focus on the 3 levels. WHO has fallen behind on this key function.	Following the JIU report, the agencies have invested heavily in strengthening their decentralized functions.	R11. Regional directors, with advice from the director of EVL, to develop a clear strategy for WHO to invest in the decentralized evaluation function to bring it up to the standards of comparable UN organizations and address issues highlighted 10 years ago by JIU. Build a common strategic approach for evaluation at all 3 levels, using the helpful collaborative approach already established in the GNE as an entry point. This must not be a top-down approach, given WHO's governance structures and culture.

	WHO Finding	Comparator agency practices	WHO Recommendation
	Capacity at regional and country level is very weak.	Several agencies have developed regional evaluation units staffed at P4 or P5 level, with country level focal points.	R12. Build capability at regional and country level on evaluation, including regional evaluation units that are staffed at the appropriate level (P4 level as a minimum) with reporting arrangements to regional directors and with a 'dotted line' reporting to the director of EVL on professional and technical aspects. Develop a network of country level focal points for M&E in the larger WHO offices, who can also support smaller countries as required.
Enabling Environment and Culture	The level of visibility and awareness of the evaluation function and the evaluation policy is low. The vision for how evaluation adds value is also not yet fully understood and there is no vehicle for high level discussion of evaluation findings.	Evaluation policies are well understood and supported by senior management and staff, partly through discussion of major evaluations, but also from communication and training. Evaluation findings are discussed in depth at the level of governing bodies and by senior managers.	<p>R13. As part of revising and launching the new Evaluation Policy, take steps to build awareness of evaluation through a major communications and training initiative led by EVL and championed by senior managers. Through this process communicate a clear vision of evaluation across the organization, how it adds value and how evaluations are intended to be used. Spell out the accountabilities set out in the evaluation policy so that management and staff are aware of what is expected and why.</p> <p>R13b. Ensure there is a high-level forum for discussing evaluations in depth. See the recommendation R5 above, which would put in place informal sessions with member states a high-level internal Evaluation Committee with representatives from senior management, chaired at ADG level, and with appropriate representation at regional and country level, who can take delivery of and discuss evaluations, findings and recommendations. Having a similar high-level committee at RO level, given WHO's governance structure, will likely also add value, especially on resource allocation, building demand, and fostering accountability for evaluation from budget centre managers. Such committees at the regional level could be chaired by RD/Director Of Programme Management level with representation from CO & RO BC managers, RO/HQ evaluation focal points, etc.</p> <p>R13c. Build demand, support and buy-in for evaluation and maximize use. This needs to start with messages communicated by the governing body and senior management on the importance of evaluation to provide a platform for demand. This would then be reinforced by closing the learning loop through consistent production of high quality, credible and useful reports which demonstrate what value can come from the evaluation function.</p>
Evaluation use and impact	The approach to dissemination of evaluations is based around publishing reports and summaries.	Several agencies have fully developed communication strategies and platforms for dissemination, using innovative approaches.	R14. Modernize and deepen EVL's approach to dissemination of evaluations , using a wider range of tools and formats as part of a strategic investment in making evaluations accessible. Starting at the top, this should include the regular informal sessions with EB representatives of member states.
Engagement and Follow up	Monitoring of evaluation recommendations does happen but in practice incentives to act on recommendations are weak, and it is largely up to the relevant manager or policy lead.	Follow up on evaluation recommendations and actions is tracked in detail, at central level and by region, with discussion by senior managers. There are significant consequences if recommendations are not followed up.	<p>R15. Strengthen the systems and incentives for follow up on evaluation recommendations by clearly defining the responsibility of senior managers in this respect, moving the focus from EVL to operations and building this into performance management and compliance.</p> <p>The internal evaluation committee proposed in R6 would provide an important focus within the organization on this, reinforced by greater interest from MS. This in turn would help in creating incentives for staff to focus on the management response actions following each evaluation. Staff should receive guidance on what is expected around management response and follow up to evaluations.</p>

Annex 1: Terms of Reference

Comparative study of WHO evaluation function with selected UN entities

The WHO Evaluation Office (“EVL”) is mandated by member States to conduct a comparative study of the Organization’s evaluation function with selected UN entities:

EB/PBAC May 2023: “Conduct a comparative study of evaluation functions and coverage across entities of the United Nations system that are comparable to WHO in size or structure and, together with recommendations to strengthen the independence, credibility and use of the evaluation function, submit a report of the findings of the study to the Programme, Budget and Administration Committee at its fortieth meeting in May 2024”.

Purpose of the comparative study

The purpose of the study is to provide recommendations to strengthen the independence, credibility and use of the WHO evaluation function by:

- Reviewing a representative sample of UN system evaluation functions which on the one hand are comparable to WHO in terms of governance, structure and resources and, on the other, have developed decentralized models which could be emulated.
- Identifying emerging good practices of UN agencies which have strengthened their evaluation function at the decentralized level, through their evaluation policies, coverage norms, human and financial resource allocation, systems and practices.

Scope

In terms of coverage, the scope of the study includes:

- WHO evaluation function: 1) EVL/HQ, 2) regional offices (RO) evaluation focal points (roles/responsibilities/diversity of maturity levels), and 3) decentralized evaluations.
- Categories of evaluations: 1) Evaluations conducted or commissioned by EVL, including the organization-wide thematic evaluations, 2) Strategic evaluations of WHO contributions at the country level, 3) Decentralized evaluations commissioned by HQ departments, regional offices and country offices, and 4) Evaluations of humanitarian actions or emergency programmes for crisis situations.
- Agencies: compare with 1) specialized agencies (governance structures, reporting mechanisms, decentralized eval models) UNESCO, ILO, FAO and 2) agencies/funds/progs coverage norms, resources, decentralized eval models): UNHCR, UNICEF, WFP, UNDP, UNFPA.

The study will focus on the following three dimensions:

- Policy and systems: Adequacy of 1) WHO Evaluation Policy and structures at different levels, against UNEG Norms and Standards, including roles/responsibilities, and technical oversight for decentralized evaluations, 2) linkages with governing bodies (e.g. reporting and reviewing of evaluations, and use of evaluations by Member States), 3) linkage between the central evaluation office and decentralized evaluation functions, and 4) systems/criteria to ensure adequate evaluation coverage of programmes/projects.
- Practice: Adequacy of 1) evaluation planning and the implementation of workplans, 2) external evaluators selection process to ensure technical quality, 3) quality control and assurance processes, 4) mechanisms to safeguard independence, 5) stakeholder engagement practices and promotion of the use of evaluations (incl. management responses, knowledge management & organizational learning), and 7) EVL ways of working with ROs and support to decentralized evaluations.
- Resourcing: Appropriateness of 1) human resources to manage evaluations at different levels, and 2) the system to finance evaluation functions from adequate sources (incl. different evaluation types, portion of budget for humanitarian/country office/project evals).

Objectives

The comparative study will have the threefold objective to:

- Compare current policy & systems, practice and resourcing of WHO evaluation function (i.e., EVL/HQ, ROs and decentralized incl. humanitarian evals) with selected UN entities' organizational setting, mechanisms, processes and tools.
- Identify good practices applicable to WHO in terms governance, coverage and resources which contribute to ensuring 1) independence, 2) credibility, and 3) use.
- Make recommendations on the three dimensions (policy & systems, practice and resourcing) to Member States and the Secretariat to enhance the ability of the WHO evaluation function to stimulate learning and promote accountability, transparency and effectiveness through independent, credible and useful evaluations.

Methods

Mixed methods will be used to identify good practice applicable to WHO for enhanced governance, structure and resources, including:

- Literature review: determine relevant docs (selected UN entities' evaluation policies/peer reviews and reports); identify data gaps, snowball sampling, collection of data from selected UN entities including through interviews e.g. 2014 JIU report, MOPAN reviews, Policies, UNEG DAC Peer reviews).
- Collection of quantitative data from selected UN entities (e.g. budget, expenditure, human resources, number of different types of evaluations...) possibly using a questionnaire and/or interviews to obtain information.
- Individual/stakeholder key informant interviews and focus group discussions 1) at HQ with EVL staff /leadership and departments/offices (incl. decentralized evaluation commissioners and managers); 2) ROs evaluation focal points and senior management as well as country offices.

Outputs/deliverables

-
- Interim report and presentation to WHO evaluation staff with preliminary findings/conclusions on identified good practices that can be applicable to WHO.
 - Comparative study report with recommendations on how to strengthen the WHO evaluation function in terms of governance, structure and resources.
 - Presentation of the comparative study and policy brief to WHO Member States (EB/PBAC) in May 2024.

Tentative timeline

End of November 2023: interim report

End of February 2024: final report

Consultants profile

EVL is seeking to hire one or two consultants to undertake the comparative study. The consultants should be experienced evaluators (at least 15 years) and have, individually or together, expert knowledge of the functioning of evaluation functions, and evaluation approaches and methodologies. Relevant experience in performing similar assignments, or other peer reviews or analysis of UN system evaluation functions, and of WHO and/or the UN system is key.

Annex 2: Interviews

Name	Role
Bernard Tomas	Head, Country Strategy and Transformation

Darshak Shah	Chair of Independent Expert Oversight Advisory Committee
David Webb	Special Adviser, Business Operations
Altaf Musani	Director, Health Emergencies Interventions
Leen Meulenberg	Unit Head, HQ/EXT/CRM/EEG, Engagement Established Government Contributors
Erin Kenney	Acting Coordinator for Gender, Equity and Human Rights
Imre Hollo	Director, Planning Resource Coordination and Per Monitoring
Dr Indrajit Hazarika	Senior Public Health Officer, Country Strategy and Support
Mubashar Sheikh	Director, Quality Norms and Standards
Srdan Matic	Head, Program and Resources Management (PRM), EURO
Anuruddhe Thushara	Planning Officer- Program Planning and Coordination, SEARO
Lloyd Masomera	M&E Officer, AFRO
Francisco Katayama	Coordinator, Planning, Budgeting, Monitoring and Evaluation, AFRO
Mika Yamai	Consultant (Evaluation), WPRO
Roberto La Rovere	Chief, Evaluation, PAHO
Atul Dahal	Coordinator - Programme Development and Operations, EMRO
Amr El Tarek	Regional Advisor, Programme Planning, Budget, Monitoring and Evaluation, EMRO
Calistus Wanjala	Regional Evaluation Officer, EMRO
Maria Joana Crespo	Organization Development Officer, EURO
Dr Wael Hatahit Acting	WR Iraq
Dr Ibrahim El-Ziq	WR Tunisia
Dr Reinhilde Therese Van De Weerd	WR Djibouti
Dr Rajesh Pandav	WR Nepal
Dr Cherian Varghese	Dir. Healthier Populations and Noncommunicable Diseases/ Regional Office SEARO
Dr Vimlesh Purohit,	WCO IND Planning Focal Point SEARO
Ibadat Dhillon	RA Human Resource for Health
Dr Sevil	Papua New Guinea WR
Dr Caroline Clarinval	Romania WR
Dr Ying-Ru Jacqueline	Laos WR
Anne-Claire Luzot	WFP, Director of Evaluation
Rob McCouch	UNICEF, Director of Evaluation
Kerry Albright	UNICEF, Senior Evaluation Officer
Alan Fox	UNDP, Acting Director of Evaluation
Bernadine Assiene	UNESCO, Director of Evaluation
Lori Bell	UNHCR, Director of Evaluation
Serdar Bayryyev	FAO, Senior Evaluation Officer
Deborah McWhinney	UNFPA, Senior Evaluation Officer
Guy Thijs	ILO, Director of Evaluation
Roger Drew	RBM Evaluation Lead
Elil Renganathan	Former Director of Evaluation
Nigel Thornton	MOPAN Assessment Lead
Jose Acacio	AMSTG co-facilitator, Australia
Yong Feng	PBAC Chair, China
Björn Kümmel	Former PBAC Chair, Germany
Aishath Rishmee	Immediate past PBAC Chair, Maldives

Annex 3: Evaluation functions of UN comparators

FAO evaluation function

The following draft summary is based on a review of key documents and an interview with Serdar Bayryyev, Senior Evaluation Officer.

Documents reviewed include:

1. 2010 Charter for the FAO Office of Evaluation
2. 2016 Evaluation of FAO's Evaluation Function
3. Interim Evaluation Strategy 2023-2025
4. 2023 Programme Evaluation Report
5. 2017-18 FAO MOPAN Assessment

Overview

FAO's evaluation function is considered to be relatively mature function in most respects, but still evolving on decentralized evaluation. The 2016 independent evaluation found significant progress in the evaluation function but also noted that the quality of evaluations is uneven and there is a critical gap in decentralized evaluation. More recent work is aimed at addressing these gaps. An external evaluation of the evaluation function was conducted in 2016 and its recommendations are now being followed up through the interim strategy. The function was also peer reviewed in 2012.

Policy

FAO's evaluation work is currently guided by an interim strategy, approved in 2023, that will guide the work of the office over the next two years with the goal of increasing its effectiveness and efficiency. The strategy seeks to address gaps identified by the independent evaluation of the function which took place in 2016. Specifically, the strategy outlines that the Office of Evaluation (OED) will focus on strengthening its capacity to: i. Generate robust and useful evidence that supports learning and accountability. ii. Make potentially transformative recommendations to FAO stakeholders. iii. Contribute knowledge to the field(s) in which FAO operates.

Rather than an evaluation policy, FAO's evaluation function is guided by a Charter from 2010 which sets out a description of the roles and responsibilities for evaluation, which the Office of Evaluation is currently working to update with a new policy that will be issued in spring 2024.

Whilst FAO does not have explicit coverage norms in its charter/strategy, the 2018 MOPAN assessment found that evaluation coverage is reasonable. There are clear criteria to guide the selection of evaluations:

1. Respond to learning priorities from Governing Bodies or the Director-General.
2. Fill a gap in evidence for FAO.
3. Have the potential to support decisions.
4. Can be adequately resourced given the learning objectives.
5. Have potential to yield knowledge of broad use.

Relevance and responsiveness

FAO's 3 year rolling workplan is subject to review by its Programme Committee acting on behalf of its Governing body, which is ultimately responsible for approving the workplan. This work plan proposes several types of evaluations, including strategic, thematic and country evaluations. The workplan is reviewed on a yearly basis and modified as required. Whilst the evaluation workplan does not appear to be fully costed, there are clear selection criteria for deciding what evaluations will be conducted.

The 2023 Programme Evaluation report outlines that in the 2021–2022 biennium, OED carried out a total of 103 evaluations and 3 syntheses of evaluations, including 74 project and programme evaluations in 53 countries, 21 regional and country-level evaluations, and eight thematic evaluations that were mostly global. In terms of coverage, this has meant that OED covered all regions in which FAO operates and nearly USD 2 billion of FAO's portfolio of work. Thematic and strategic evaluations accounted for 58 percent of this funding, whereas project/programme and country programme evaluations accounted for 29 percent and 13 percent respectively. Despite disruptions caused by the COVID-19 pandemic, the number of evaluations completed by OED continued to increase over time, from 66 in 2017-2018 to 95 in 2019-2020 and 106 in 2021–2022.

Within FAO, any project above \$6.2 million (previously \$4 but was adjusted for inflation in June 2023) needs to be independently evaluated. Below this, evaluations are not mandatory unless outlined as a donor requirement. Because FAO has an increasing number of projects, there is an increasing demand for evaluations across the organization which OED must meet, creating increased pressure of OED staff to respond to this demand. To respond to this, OED is trying to increasingly cluster evaluation thematically/geographically, although there are challenges in this regard in terms of sequencing.

Enabling environment and culture

The Evaluation of FAO's evaluation function in 2016 found that by reducing the evaluation policy to a charter for OED it had transformed OED into an "evaluation silo" and limited the growth of an evaluation culture. In addressing this, OED's interim evaluation strategy sets out an initial architecture for excellence in supporting FAO Members and Management's culture of evidence in decision-making and outlines a shared evaluation culture as one of the conditions for its success. From OED's perspective, there is a good receptivity to evaluation across the organization and good engagement from the governing body and management with evaluation findings.

Resourcing - human and financial

In terms of human resources, the FAO evaluation function has 20 staff (1 Director – at D2 level, 15 professional staff (at P3, P4 and P5 levels), 4 general service staff), as well as a number of consultants throughout the year working on evaluation/standalone projects.

Of the 20 staff, 5 of these are Regional Evaluation Advisors (P4s). These posts were established based on the recommendations of the 2016 Evaluation of FAO's evaluation function which highlighted the need for increased decentralization. Evaluation personnel in the regions report to headquarters– and closely collaborate with FAO management and personnel in these regions in conducting or managing evaluations. The aim is that this keeps OED connected while close to the ground, ensures a deeper understanding of stakeholder needs and FAO's work; and builds resilient, adaptive strategies across contexts that promote work continuity and success in conducting evaluations.

Relative to other UN agencies, OED does not consider itself to be lacking in resources to undertake its evaluation work. The evaluation budget for the 2021-22 biennium was \$10.2 million (\$8 million in the current biennium). Funding for evaluation comes from the regular programme budget, as well as an Evaluation Trust Fund which was established in 2010 (where 0.8% of the funding of each project is directed to the trust fund). Thematic, strategic and corporate evaluations are funded from the regular budget and most project evaluations are funded from the trust fund.

Evaluation use and impact

The OED Strategy refers to promoting use of evaluations through a wider range of evaluation products, including briefs and various dissemination activities including workshops, events, online presence, briefs and infographics, blogs etc. Evaluation reports, management reports and follow-up reports are all available publicly on the FAO website.

For global evaluations, both the evaluation and associated management response are submitted to the FAO's governing body at the same time and are publicly shared. After 2 years, there is a requirement for FAO management to submit a follow up on how recommendations have been implemented to the governing bodies for them to review what actions have been undertaken. There is system to track follow up of evaluation recommendations. For country and project evaluations, there is a similar process (but not at the governing body level). Management responses for country programme and project evaluations must be issued within one month of an evaluation being completed, with follow-up report on the evaluation recommendations submitted by evaluands to OED after one year. The 2017/18 MOPAN assessment found good evidence of evaluation findings being used at a strategic level, particularly in terms of Climate Change Evaluation findings and adoption of recommendations.

Independence

The Evaluation Charter highlights independence as one of the primary principles underpinning evaluation and that it "should be protected throughout the evaluation process: policy, institutional framework, management of the evaluation function, conduct of evaluations and follow-up". However, the 2016 evaluation noted a number of areas where OED's independence could be strengthened in terms of its reporting lines as although it remains separate from those directly responsible for the design and implementation of the policies and operations that are evaluated, it reports to the ultimately responsible for the design and implementation.

The evaluation and the MOPAN assessment found no issues with behavioural independence and that FAO treats the independence of OED seriously and has recently introduced institutional changes to strengthen the OED and reduce the risk of a conflict of interest. however, there remains a lack of guidance. The OED Evaluation Manual does not provide guidance on how to safeguard the values of independence and impartiality. This is a potential gap, given that OED evaluation managers work closely with commissioned evaluation teams and the units being evaluated over the course of each evaluation.

Quality

In terms of quality, the 2016 Independent Evaluation of FAO's Evaluation Function rated the overall quality of FAO's evaluation methodology as only medium to high, lower than other assessment criteria. The MOPAN assessment found that FAO has a quality assurance framework in place to support the quality of evaluations, but that the system did not apply specific criteria or checklists, which from a quality improvement perspective makes it difficult to examine systematically where strengths and weaknesses lie. Currently QA is largely post-hoc, undertaken by a consultancy firm who QA a sample of evaluation products every year. There is currently no annual synthesis of evaluation quality.

To address these weaknesses, support the ambitions of the new interim strategy and allow more opportunity for 'real-time' improvement in the quality of evaluations, OED is working to standardize and expand existing quality assurance processes to include ex ante and ex post components systematically. Ex ante quality assurance will be provided mostly by internal experts in several critical stages of an evaluation before it is completed, including design, data collection and analysis, and development of evaluation findings and recommendations. Ex post quality assessments will be conducted by external experts reviewing completed evaluations; it will be used to diagnose areas for improvements, inform yearly OED strategic planning, and promote credibility and accountability.

Areas of best practice relevant to WHO's evaluation function:

1. The development of an aspirational strategy to operationalize and support the implementation of the evaluation policy.
2. The development of clear criteria to guide the selection of evaluations.
3. Establishment of an evaluation trust fund to generate evaluation funding.
4. Use of clustering mechanism to group evaluations at thematic/regional level etc.
5. Dual reporting lines of Regional Evaluation Advisors

ILO's evaluation function

The following draft summary is based on a review of key documents and an interview with Guy Thijs, Director, ILO Evaluation Office (EVAL).

Documents reviewed include:

1. 2017 Evaluation Policy
2. ILO Evaluation Strategy 2023-25
3. Annual Evaluation Report 2022-23
4. Independent Evaluation of the ILO's Evaluation Function 2017-21
5. 2021 MOPAN Assessment

Overview

ILO's evaluation function is considered to be a mature function in most respects, with the quality, professionalism, and independence of evaluation that is well-recognized and endorsed by ILO staff and stakeholders and behavioural and organizational independence of the central evaluation function ensured.¹⁶ However, current requirements set for conducting a large number of evaluations place significant burdens on evaluation staff and key stakeholders, which in turn may impact their capacity to make use of findings in strategic decision-making and lesson-learning. Whilst EVAL has supported extensive learning exercises, there used to be more primacy given to the accountability purposes of evaluation rather than learning and improvement which has impacted the evaluation culture. An independent evaluation of the evaluation function was completed in 2022.

Policy

ILO's evaluation work is guided by its 2017 Evaluation policy and its 2023-25 Results-Based Evaluation Strategy which provides a clear and detailed overall framework for evaluation in ILO, including both centralized and decentralized evaluation. The strategy outlines outcomes, sub-outcomes, targets, and measures used to support the Policy's implementation and monitor progress and EVAL reports on this annually to its governing body. The strategy seeks to address gaps identified by the independent evaluation of the function, which was completed in 2022, specifically, regarding the volume of Evaluations and associated workload, and the need to balance better learning versus accountability requirements and to introduce office-wide, structural improvements aimed at fostering a more enabling environment.

The ILO Policy Guidelines for Results-Based Evaluation provide an operational framework for evaluation and describe the principles, rationale, and guidance on planning and managing evaluations. The Evaluation Office was adaptive and developed and published guidance (Implications of COVID-19 on evaluations in the ILO: Practical tips on adapting to the situation) in March 2020 to adapt evaluation policy to the changing context.

¹⁶ The JIU comparative analysis of the evaluation function in the United Nations system of 2014 placed ILO in the top three Level I of development of the evaluation function by size of the organizations' overall annual budgets and the location of the central evaluation function (page 22)
https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/en/reports-notes/JIU%20Products/JIU_REP_2014_6_English.pdf

The policy also outlines coverage norms which are aligned to budget; ILO interventions over USD 1 million are subject to an independent evaluation. All interventions over USD 5 million are also required to undergo an initial monitoring and evaluation appraisal by the Evaluation Office. However, the new evaluation strategy outlines a move to more criteria-based evaluation planning to reduce the large volume of mandatory, decentralized evaluations and allow for more strategically oriented evaluation evidence and learning opportunities.

Relevance and responsiveness

High-level evaluations are planned in a four-year rolling work plan. Decentralized evaluations are planned on an annual basis, and evaluation focal points in each department and region develop rolling work plans to implement their respective evaluation plans reflected in the i-eval discovery platform. A total of 70 decentralized evaluation reports were completed in 2022, alongside two corporate evaluations. This represents an increase of 11 independent evaluations (16 per cent) from the previous reporting period. More independent evaluations were completed in 2022 than in any other year.

In terms of coverage, there is good coverage of ILO's evaluation activities. Guidance for the implementation of different types of evaluations is mapped out in the Policy Guidelines. This includes evaluations of strategies and policies, DWCPs, projects, as well as thematic evaluations, synthesis reviews and meta-studies, impact evaluations, and high-level evaluations. The policy delineates centralized management of corporate evaluations and a hybrid approach for project evaluations. In order to improve the efficiency and relevance of evaluation planning, the ILO is since 2019 increasingly focusing on more strategic and clustered evaluations as the default modality rather than a piecemeal, project-based approach but this can be constrained by donor requirements as donors often demand evaluation reports for their projects.

Enabling environment and culture

The independent evaluation of ILO's Evaluation Function completed in 2022 found that there was a conducive culture within ILO to use evaluations to assess impact and a shared commitment to compliance with evaluation responsibilities, and that this was supported by a structure that embeds senior management and leaders in the response to and actions arising from evaluations as well as tripartite constituents in the design and review of evaluations. However, it noted that given the volume of evaluations, and time constraints of staff, there was scope to improve the use of evaluation for learning purposes. Other functions in the ILO have since also been identified as custodians as key enablers of the new evaluation strategy. Integrating evaluative thinking into ILO's capacity building measures leads to improved ILO's evaluation culture' and includes initiatives for enhancing the enabling environment, covering both institutional and cultural measures and procedures, as well as capacity development. It notes however that the evaluation function remains underfunded. Whilst the Evaluation Manager Certification Training Programme that has been rolled out across the organization has allowed to leverage capacity throughout the organization to conduct an impressive number of evaluations and is thought to have promoted an evaluation culture in ILO, the strategy outlines that institutional support and incentives (culture) for their voluntary contribution remains inadequate. From EVAL's perspective and validated by the independent evaluation of the evaluation function, there is a good receptivity to evaluation across the organization and there is also considered to be good engagement from the governing body and management with evaluation findings, but it remains inadequate.

Resourcing - human and financial

In terms of human resources, EVAL has 7 staff (1 Director, 3 Senior Evaluation Officers, 1 Evaluation Officer, 1 Communications & Knowledge Management Officer and an Administrative Assistant). This is considered to be

insufficient given the volume of evaluations EVAL conducts on an annual basis, and the existing policy requirements, quality assurance mechanisms and knowledge management ambitions further place a strain on the staff of EVAL in terms of volume and pressure to deliver on time.

At a decentralized level, there is a network of five Regional Evaluation Officers (REOs). REOs have a technical reporting line to the Director of EVAL but are not EVAL staff, which has been highlighted as a limitation to the independence of the evaluation function and places a large burden on EVAL staff given the volume of evaluations as EVAL staff are required to quality assure all evaluation products delivered by regional offices. The role of REO's is also recognized to be overloaded with the majority of their evaluation management skills directed to overseeing evaluations for projects less than \$5 million – accounting for nearly 56 per cent of their evaluation portfolio.

There are also non-evaluation staff acting as 14 Departmental Evaluation Focal Points (DEFPs) and over 150 certified Evaluation Managers amongst ILO staff, but these are on a voluntary basis and performed in addition to existing roles. As of the end of 2022, the number of ILO staff who became certified evaluation managers had reached 150. The continued demand for EMCP training has led to a significant expansion of the pool of certified evaluation professionals, with a 12.6 per cent increase. As of June 2023, there were 169 certified staff members, surpassing the established milestone. A cost-recovery scenario to compensate for staff time devoted by evaluation managers to evaluation-related tasks is being considered to incentivize staff (and their managers) to take on the Evaluation Manager role going forward but is facing internal resistance.

In terms of financial resources, the budget allocation for both centralized and regional evaluation accounted for just over 0.8 per cent of the total ILO budget in 2022. The resources for evaluation as a proportion of total programming spend is on par with other UN entities,¹⁷ along with evaluation resources as a proportion of total organizational budget. However, the ILO evaluation policy sets a goal of approaching a combined evaluation expenditure of 1.5–2 per cent of total expenditures, as recommended in international evaluation standards so this goal has not yet been reached, nor is the current budget considered to be commensurate in terms of the volume of evaluations undertaken across ILO. The budget for the core evaluation function (from regular budget funding) in 22-23 was estimated at \$3.2 million with an additional \$2 million from extra-budgetary resources, and regional was in the range of \$2.1 million.

For projects, resources are earmarked for monitoring and evaluation with a minimum of 2% of total project resources reserved for evaluations and an additional 3% reserved for monitoring and reporting (at a total minimum of 5%), according to the ILO Finance Manual and the ILO Policy Guidelines for Evaluation. These dedicated resources are considered key to ensure adequate evaluation coverage. This earmarking of project funds is not compulsory for smaller projects below \$500,000 that only require a final progress report with self-evaluation components. ILO policies also recommend that resources be set aside for monitoring, collecting baseline data, and reporting and conducting evaluability assessments – a minimum of 3 per cent is recommended. In 2021, around \$1.5 million was spent on project evaluations.

As noted, the volume of evaluations conducted by ILO is thought to be challenging for the organization to manage with existing resources and the independent evaluation of the ILO evaluation function found that the existing financial system and funding mechanisms for evaluations are too rigid to allow for pooled funding for smooth implementation of portfolio or cluster evaluations, which would increase the use of cluster evaluations as a way of generating evaluative evidence in a more strategic and efficient manner. The 2023-25 Evaluation Strategy therefore includes the establishment of an Evaluation Trust Fund to pool resources and maintains the threshold for triggering for evaluations, but regions and departments can now request waivers based on justifying criteria, which could result in fewer, more strategic evaluations.

¹⁷ Based on this comparative assessment, FAO and UNFPA are also both spending around 0.8% on evaluation as a proportion of their total programming spend.

Evaluation use and impact

The independent evaluation of ILO's evaluation strategy found there to be a strong appetite in the ILO to use evaluations to assess impact and a shared commitment to compliance with evaluation responsibilities, although to a lesser degree with regards to learning. The EVAL Strategy highlights expanded knowledge base of evaluation findings and recommendations that effectively contributes to organizational learning and enhances organizational effectiveness as one of its key outcomes and EVAL is currently developing a communication plan informed by the new strategy. Key to this is improvements in functionality and use of the ILO's evaluation dashboard (i-eval Discovery) which publicly displays for over a decade all planned evaluations and completed evaluations, along with their related summaries, lessons learned, good practices, recommendations and management responses to evaluation recommendations. The enhancements are intended to improve evaluation use and the user's experience when accessing the new the i-eval Discovery dashboard.

EVAL is recognized as producing high-quality knowledge products, including i-eval THINK Pieces, synthesis reviews and meta-studies using evaluation reports, as well as learning series to facilitate discussions on findings and recommendations. However, despite these efforts, the use and uptake of evaluation products by stakeholders remains somewhat limited. To address this, ILO has introduced a Criteria-based Integrated Evaluation Planning System (CIEPS) to enhance the use of evaluation results. The system identifies evaluations that meet the ILO's knowledge requirements and learning needs, while maintaining accountability. It prioritizes strategic evaluations and considers different evaluation types, recognizing that the current large volume of project evaluations can jeopardize the utility and learning generated from the evaluations and that there is limited time and opportunity to use evaluations and engage in evaluative practice as a vehicle for critical analysis and continuous improvement.

ILO has had an automated management response system since 2018 for decentralized evaluations, where management responses to independent evaluations are required to be submitted by line managers outlining whether: (a) action taken in response to a recommendation has been completed or partially completed; (b) no action is planned; (c) action has not yet been taken; or (d) the recommendation has been rejected. Management follow-up to these recommendations from decentralized evaluations have been publicly accessible since 2018 on EVAL's online dashboard, called i-eval Discovery. Management follow-up to recommendations from high-level evaluations (strategy, policy, institutional and decent work country programmes) have been submitted and discussed by members of the ILO's Evaluation Advisory Committee (chaired by the Deputy Director-General) and will be integrated into i-eval Discovery) in the near future. This will ensure full transparency and accessibility of all management responses to the evaluation recommendations.

EVAL closely monitors the uptake of evaluation findings in key strategic documents such as the ILO's programme and budget. A study conducted from 2021 to 2023 revealed 115 references to EVAL outputs in 34 strategic documents, including Governing Body reports, programme and budget and programme implementation reports.

Independence

The independence of the ILO evaluation was rated as 'highly satisfactory' in its most recent MOPAN report with strong operational and financial independence. Direct reporting lines to the Director-General and the Governing Body have been set up since 2012, and the Evaluation Office has been given full discretion in deciding on the evaluations to be conducted. Evaluation in the ILO is structurally independent, with the Director of the Evaluation Office reporting directly to the Governing Body. The ILO evaluation policy outlines the importance of independence of evaluation processes and that it will ensure separation of evaluation management and implementation responsibility from line management functions for policies, programmes and projects.

Quality

ILO's guidance and approach on assuring the quality of centralized and decentralized evaluation has evolved and expanded over time. The evaluation function benefits from multiple layers of quality assurance, including real-time and ex-post assessment of quality, supported by up-to-date, comprehensive quality guidelines. Evaluation managers are equipped with tools and information to ensure that appropriate methodologies are used in evaluations. The ILO evaluation function also has quality guidelines for the conduct and quality assurance for centralized and decentralized evaluations. The quality control process for decentralized evaluations is carried out through Regional Evaluation Officers and Departmental Evaluation Focal Points. Evaluation reports are quality controlled by the Evaluation Office, using checklists, prior to publication. In addition, the ex-post quality control of evaluations is conducted externally, and an annual synthesis of evaluation quality is undertaken.

Areas of best practice relevant to WHO's evaluation function:

1. The development of an aspirational policy and time-bound strategies to operationalize and support the implementation of the evaluation policy.
2. The development of clear criteria to guide the selection of evaluations based on knowledge and evidence gaps.
3. Lessons regarding how ILO has increased the independence of its evaluation function overtime.
4. The use of a hybrid decentralized evaluations system with a pool of Certified Evaluation Managers to increase evaluation capacity and embed evaluation culture.

UNDP evaluation function

Overview

The evaluation function is UNDP one of the most well-established of all the UN agencies. Its first evaluation policy dates back to 2006. Since then, it has been reviewed and updated 3 times (board approved 2011, 2016 and 2019). It played a leading role in establishing good practices for evaluation functions by example through its own approach and its support to UNEG, particularly in its early years but also more recently. UNDP has also played a leading role on evaluation capacity development in developing countries for many years.

The 2020 MOPAN assessment of UNDP noted that it has “a well-established, credible independent evaluation function, an up-to-date evaluation policy, robust systems and a growing evaluation budget. Notwithstanding efforts to improve their oversight and quality, decentralized evaluations remain relatively weak.” It also noted that the independent evaluation office (IEO) produces high-quality, highly credible independent evaluations and supports the evaluation function across the organization through clear guidance and robust quality assurance. However, audits and quality assessments have noted persistent issues with the quality of decentralized evaluations – see the roadmap to address this below.

Policy and enabling environment

The evaluation policy and enabling environment for UNDP is very well embedded: the policy sets out particularly clearly the roles of different parts of the organization in relation to centralized and decentralized evaluations, the governance structure for evaluation, the role of management and its responsibilities for decentralized evaluations and engaging on evaluations, and how independence is safeguarded. It exemplifies how a well-functioning evaluation function works as a system, with each level contributing to the other.

The specific governance structure is set out, including that the UNDP Executive Board: “is the custodian of the evaluation policy; annually considers its implementation, and periodically commissions independent reviews of the policy.” It approves the biennial financial appropriation to IEO, as well as its annual programme of work. The IEO submits independent thematic and programmatic evaluations to the Executive Board, which approves or notes the management responses as appropriate.

The Independent Evaluation Office: is a functionally independent unit with UNDP that supports the oversight and accountability functions. As custodian of the evaluation function, the IEO conducts independent evaluations, sets standards and guidelines, manages the systems for quality assessment and evaluation planning and use through the Evaluation Resource Centre, and develops products to support organizational learning, knowledge management and evaluation capacity development.

The UNDP Administrator safeguards the integrity of the evaluation function, ensuring its independence from operational management and activities; ensures that adequate financial resources are allocated to the evaluation function across the organization, in accordance with the Executive Board-approved financial appropriation for

Independent Evaluation Office, and reports to the Board annually on the volume of resources that the organization has invested in evaluation; ensures that the Office has unfettered access to data and information required for the evaluation of UNDP performance; and appoints the Director of the Office in consultation with the Executive Board, taking into account the advice of the Audit and Evaluation Advisory Committee.” The UNDP Administrator safeguards the integrity of the evaluation function, ensuring its independence from operational management and activities; ensures that adequate financial resources are allocated to the evaluation function across the organization, in accordance with the Executive Board-approved financial appropriation for Independent Evaluation Office, and reports to the Board annually on the volume of resources that the organization has invested in evaluation; ensures that the Office has unfettered access to data and information required for the evaluation of UNDP performance; and appoints the Director of the Office in consultation with the Executive Board, taking into account the advice of the Audit and Evaluation Advisory Committee.” Other aspects of roles on evaluation are clearly set out in more detail in the policy.

Systems and practice

An important aspect of UNDP’s evaluation work is the approach it has taken on strengthening decentralized evaluation, which is set out in a strategy/road map developed in 2021/22. This was in response to concerns raised at earlier stages on the quality of decentralized evaluation. The strategy covers areas such as quarterly assessment of evaluation quality, training at different levels including for regional bureaux and national capacity development, evaluation awards to incentivize performance, and enhanced data (ERC revamp) and accountabilities (performance ratings). To support and inform the conduct of evaluation at different levels, UNDP has a comprehensive set of guidelines.

However, the quality of decentralized evaluations remains mixed. In 2022, 307 evaluations were quality assessed by the IEO and 41% were rated as satisfactory, 48% as moderately satisfactory and 11% were moderately unsatisfactory or unsatisfactory.

Another notable feature of UNDP’s evaluation function is the richness of its evaluation knowledge base – which in turn reflects high levels of coverage of evaluation over many years - and its approach to knowledge management to support use of evaluation. The Evaluation Resource Centre is a single large repository of UNDP evaluations summarizes and provides access to over six thousand evaluations conducted over the last 25 years at all levels. It is comprehensive and well organized, having recently been revamped and includes a range of different centralized and decentralized evaluation products including:

- Over 100 thematic evaluations (including periodic evaluations of UNDP’s Strategic Plan)
- 3 evaluations of UNDP’s Global Programme for Development Results
- 171 UNDAF evaluations
- Over 200 independent country evaluations and 64 decentralized country programme evaluations
- Over 4,000 project evaluations

More recently, in support of evaluation use and organizational learning, the IEO has created a dedicated section that focuses on synthesis and lessons from its evaluations. It has started to develop a range of products that distil and capture evaluation evidence in an accessible form.

The Independent Evaluation Office has a clear and systematic approach to engaging with its stakeholders both for planning and selecting evaluation topics and also for learning lessons. This is important for ensuring an effective dialogue with the governing body and senior management on evaluation. It includes the use of retreats with the Board and regular meetings with the UNDP Executive Group on thematic and strategic evaluations. There are evaluation focal points that mirror the areas in UNDP’s Strategic Plan, as a vehicle closely follow the evolution of

UNDP work in different areas and regularly engage with colleagues in UNDP's Global Policy Network. This is one way that evaluation utility is supported.

Meanwhile there is evidence that IEO recommendations are found to be clear and useful by stakeholders surveyed in 2022 and were rated highly for impartiality. Satisfaction with IEO products is also high (between 76% and 90% depending on the type of product) and increased in the last 3 years. Institutionalized follow-up to evaluation recommendations through management responses functions well. Most management responses were actionable and of good quality.

Another example of how evaluation evidence is made accessible and useful is the Reflections series - a set of 22 short publications produced by IEO in different thematic areas such as COVID-19, health, governance and water. These synthesize lessons from past evaluations to support organizational learning about what works and what does not in different development contexts. The aim of the series is to provide relevant, useful, and accessible lessons to country offices of UNDP, as well as to the wider community of development practitioners. It is a rapid evaluation synthesis from material issued by UNDP between 2013 and 2023, as well as from external evaluative evidence.

Resources – human and financial

As of 2022, the Independent Evaluation Office had 35 posts, including 27 professionals and 8 General Service staff, with additional support from 9 long-term consultants. The IEO budget was \$13m in 2022 and the total resources covering IEO, and decentralized evaluation was nearly \$37m. This represented 0.77% of UNDP programme resources. Resourcing of evaluation has risen very significantly in the last 5 years, from \$21.8m (0.48% of programme resources) in 2017. However, there is a gap compared with the target set out in the policy, which is 1% of programme resources. There are 10 regional focal points working to support decentralized evaluations in the regions and who in turn work with country level M&E focal points.

Enabling environment

The evaluation function is UNDP one of the most well-established of all the UN agencies. Not surprisingly, therefore, the evaluation policy and enabling environment for UNDP is very well embedded and sets out particularly clearly the roles of different parts of the organization in relation to centralized and decentralized evaluations, the governance structure for evaluation, the role of management and its responsibilities for decentralized evaluations and engaging on evaluations, and how independence is safeguarded.

This may serve as a useful reference point for WHO in providing an example of how a well-functioning evaluation function works as a system, with each level contributing to the other – as opposed to it being mainly focused on the work of the independent evaluation office.

A second area to learn from in UNDP is its approach to strengthening decentralized evaluation, which is set out in a strategy/road map developed in 2021/22. This covers areas such as quarterly assessment of evaluation quality, training at different levels including for regional bureaux and national capacity development, evaluation awards to incentivize performance, and enhanced data (ERC revamp) and accountabilities (performance ratings).

Other notable areas, see below, include its rich evidence base and repository for evaluation, its strengths in depth and coverage of country programme evaluations, its synthesis and lesson learning products, its work on evaluation capacity development and its leadership and partnership work on joint SDG evaluation.

The specific governance structure and roles for evaluation in UNDP are set out in the UNDP policy and accompanying guidance as follows:

-
- The UNDP Executive Board: “is the custodian of the evaluation policy; annually considers its implementation, and periodically commissions independent reviews of the policy.”
 - The Board approves the biennial financial appropriation to IEO, as well as its annual programme of work. The IEO submits independent thematic and programmatic evaluations to the Executive Board, which approves or notes the management responses as appropriate.
 - The Independent Evaluation Office: “is a functionally independent unit with UNDP that supports the oversight and accountability functions of the Executive Board and the management of UNDP, the United Nations Capital Development Fund and United Nations Volunteers programme. The structural independence of the Office underpins and guarantees its freedom to conduct evaluations and report evaluation results to the Executive Board.”
 - As custodian of the evaluation function, the IEO conducts independent evaluations, sets standards and guidelines, manages the systems for quality assessment and evaluation planning and use through the Evaluation Resource Centre, and develops products to support organizational learning, knowledge management and evaluation capacity development. The IEO also participates in UNEG, which works to strengthen the objectivity, effectiveness and visibility of the evaluation function across the United Nations system.
 - The UNDP Administrator “(a) safeguards the integrity of the evaluation function, ensuring its independence from operational management and activities; (b) ensures that adequate financial resources are allocated to the evaluation function across the organization, in accordance with the Executive Board-approved financial appropriation for Independent Evaluation Office, and reports to the Board annually on the volume of resources that the organization has invested in evaluation; (c) ensures that the Office has unfettered access to data and information required for the evaluation of UNDP performance; and (d) appoints the Director of the Office in consultation with the Executive Board, taking into account the advice of the Audit and Evaluation Advisory Committee.”
 - UNDP programme and policy units (headquarters, regional and country offices) “commission decentralized evaluations according to evaluation plans that coincide with relevant programmes (regional and country) and global projects. The evaluations are to be carried out by independent external consultants, and UNDP management shall take all necessary actions to ensure the objectivity and impartiality of the process and persons hired.”
 - The Bureau for Policy and Programme Support, in addition to undertaking its own evaluations, “coordinates communication between UNDP management and the Independent Evaluation Office and advises regional bureaux on the decentralized evaluation function for UNDP. The Bureau works with the monitoring and evaluation staff of UNDP units to ensure that evaluation plans are properly implemented. Together with the Office, the Bureau provides guidance to UNDP units on the use of evaluation findings and lessons to improve organizational decision-making and accountability and synthesizes evaluation lessons for institutional learning. It also monitors implementation of the management responses to independent evaluations and decentralized evaluations in UNDP.
 - Regional bureaux, in addition to implementing their own evaluations, support country offices in the development of evaluation plans and implementation of evaluations and oversee implementation of evaluation plans through their appointed evaluation focal points.

-
- Bureau and country office senior management (bureau directors, resident representatives and country directors) are responsible and accountable for the development of units' evaluation plans and ensuring their timely implementation.
 - The Audit and Evaluation Advisory Committee, expanded to include evaluation oversight functions, advises the UNDP Administrator on:
 - The Evaluation Policy
 - Appointment and dismissal of the IEO Director
 - IEO multi-year and annual workplans, budgets and periodic reports
 - Thematic and programmatic evaluation reports and management responses
 - The UNDP decentralized evaluation function and national evaluation capacity programming
 - The Committee also periodically receives, and comments on, the IEO programme of work, and appraises the performance of the IEO Director annually. It further helps to safeguard the Evaluation Policy.

A notable feature of UNDP's evaluation function is **the richness of its evaluation knowledge base** – which in turn reflects high levels of coverage of evaluation over many years - and its approach to knowledge management to support use of evaluation.

The **Evaluation Resource Centre** is a single large repository of UNDP evaluations summaries and provides access to over six thousand evaluations conducted over the last 25 years at all levels. It is comprehensive and well organized, having recently been revamped and includes a range of different centralized and decentralized evaluation products including:

- Over 100 thematic evaluations (including periodic evaluations of UNDP's Strategic Plan)
- 3 evaluations of UNDP's Global Programme for Development Results
- 171 UNDAF evaluations
- Over 200 independent country evaluations and 64 decentralized country programme evaluations
- Over 4,000 project evaluations

More recently, in support of evaluation use and organizational learning, the IEO has created a dedicated section that focuses on **synthesis and lessons from its evaluations**. It has started to develop a range of products that distil and capture evaluation evidence in an accessible form.

An interesting innovation is the multi- year effort currently underway to utilize artificial intelligence as a mechanism to scan and provide insights on the over 6,000 reports in the UNDP Evaluation Resource Centre. This effort: Artificial Intelligence for Development Analytics (AIDA), (see <https://aida.undp.org/about>) has gained attention from other UN evaluation practitioners and includes high level multi-topic search and sort, the ability through the system to provide analysis of positive and negative sentiment, and a (contained) ChatGPT summarization component.

Finally, the Independent Evaluation Office has a **clear and systematic approach to engaging with its stakeholders**. This is important for ensuring an effective dialogue with the governing body and senior management on evaluation. It includes the use of retreats with the Board and regular meetings with the UNDP Executive Group on thematic and strategic evaluations. There are evaluation focal points that mirror the areas in UNDP's Strategic Plan, as a vehicle closely follow the evolution of UNDP work in different areas and regularly engage with colleagues in UNDP's Global Policy Network. This is one way that evaluation utility is supported.

Areas of best practice in UNDP to inform strengthening WHO evaluation function

- High level of capacity of IEO in human resources and professional skills
- 1% target / 0.77% achieved level of resourcing of evaluation function as a whole.
- Clarity of governance structure and roles, including that the evaluation function's role and mandate supported from the top by the EB and senior management.
- Extensive programme of evaluation training for staff and country M&E focal points.
- Evaluation excellence awards to incentivize quality of decentralized evaluations.
- Comprehensive guidelines and handbook to inform evaluation implementation.
- Strategic approach to strengthening decentralized evaluations – owned and developed by the regions and management – see Road Map 2022.
- Evaluation resource centre – repository of evaluation evidence and methods centre a detailed guide to methodological advances in evaluation and data analysis tools.
- Use of evidence and meta-evaluation – ability to use different levels of evaluation as building blocks for synthesis. Use of artificial intelligence to help analyze and summarize content across multiple evaluations in the Evaluation Resource Centre.
- Well-established approach to country level evaluations.
- Wider contributions to evaluation globally and nationally, through UNEG and ECD products

UNHCR evaluation function

The following draft summary is based on a review of key documents and an interview with Lori Bell, Head of Evaluation.

Documents reviewed include:

1. UNHCR evaluation policy, 2022-2027
2. 2021 Peer Review of the UNHCR Evaluation Function
3. UNHCR Evaluation Office Annual Report 2022
4. 2017-18 MOPAN Assessment

Overview

The Evaluation Office is in practice a very young evaluation entity. Before 2016, UNHCR had a joint Policy Development and Evaluation Service and its capacity and policies to conduct centralized evaluations were relatively undeveloped. In October 2016, the High Commissioner issued UNHCR's revised evaluation policy, establishing a dedicated Evaluation Service reporting directly to him. The policy addressed many of the gaps in policy or areas of weakness identified by previous assessments. Subsequently, the first Head of the Evaluation Service was recruited externally and took her position in June 2017. The name of the entity (to Evaluation Office) was changed as of the new evaluation policy in 2022.

The Evaluation Office (EvO) is establishing the evaluation function during a time of fundamental change and reform in the organization. The EvO has had to ensure that what it is building is attuned to changing organizational structures and to new core businesses processes, as well as an entirely new results-based management system. Further, UNHCR's way of working is also in flux, and this has implications for appetite and capacity to absorb further change at all levels of the organization.

Policy

To operationalize the Policy the Evaluation Office developed a five-year Evaluation Strategy (2023 to 2027). It sets out the vision in which "the use of evaluative evidence is normal and habitual in designing, targeting and delivering strategies, policies, and programmes...." It has a theory of change "focused on increasing both demand and supply of evaluations" and centered around the "3Cs" – Coverage (and quality), Capacity and Culture. UNHCR's evaluation planning and coverage have significantly improved since 2016. The selection of topics for centralized evaluations is more strategic and coverage has increased markedly (source: UNEG Peer Review 2021 and MOPAN 2023).

Evaluation Use and impact

To support the use and impact of its evaluation, UNHCR is increasingly focusing on communications and trying to develop evaluation briefs, infographics, videos, etc. it has also started to harness AI technology to undertake more synthesis exercises. It also ensures that when country case studies are undertaken as part of corporate evaluations,

there is a country presentation/aide memoire produced to support the country's learning rather than it being an extractive process.

Given the relative infancy of the evaluation function within UNHCR, the evaluation culture is also nascent and there is a lack of familiarity with evaluation, with frequent confusion between evaluation and audit. The EvO is working to promote the use of evaluation as a management tool by trying to build relationships with UNHCR senior management and be present at discussions to try to showcase evaluation learning or play an advisory role where possible. It has also worked to embed evaluation within the RBM system.

All corporate evaluations are presented to the High Commissioner and the Senior Executive Team (SET) – and all but Country Strategy Evaluations presented to the Senior Management Committee. Management responses for corporate evaluations need to be signed off a SET member. UNHCR is just starting to implement a process whereby the implementation of recommendations is monitored and reported on. Starting from 2023, management can see all “open recommendations” coming from audits and evaluations in a user-friendly dashboard (hosted by the oversight coordination unit under the IGO) intended to prompt timely follow-up action. One of the features of the new 2022 UNHCR evaluation policy includes a requirement for management reporting on evaluation recommendations for a two-year period after the completion of evaluations.

UNHCR also showcases examples of evaluation use in its annual report.

- The newly revised Policy on Emergency Preparedness and Response was informed by the lessons distilled from the recent evaluations of UNHCR's L3 emergency responses. The evaluation findings provided lessons learned on the design and implementation of policies, guidance and systems used in large-scale rapid-onset emergencies.
- The new UNICEF/UNHCR strategic collaboration framework builds on the lessons learned from the UNHCR-UNICEF Blueprint for Joint Action, thanks to the jointly managed iterative evaluation 2020-2022.
- New focus area strategies under the corporate strategic plan “Strategic Directions” have been heavily informed by the evaluations, notably the Engagement with Humanitarian-Development Cooperation, Statelessness, IDP and Gender-based violence strategies.

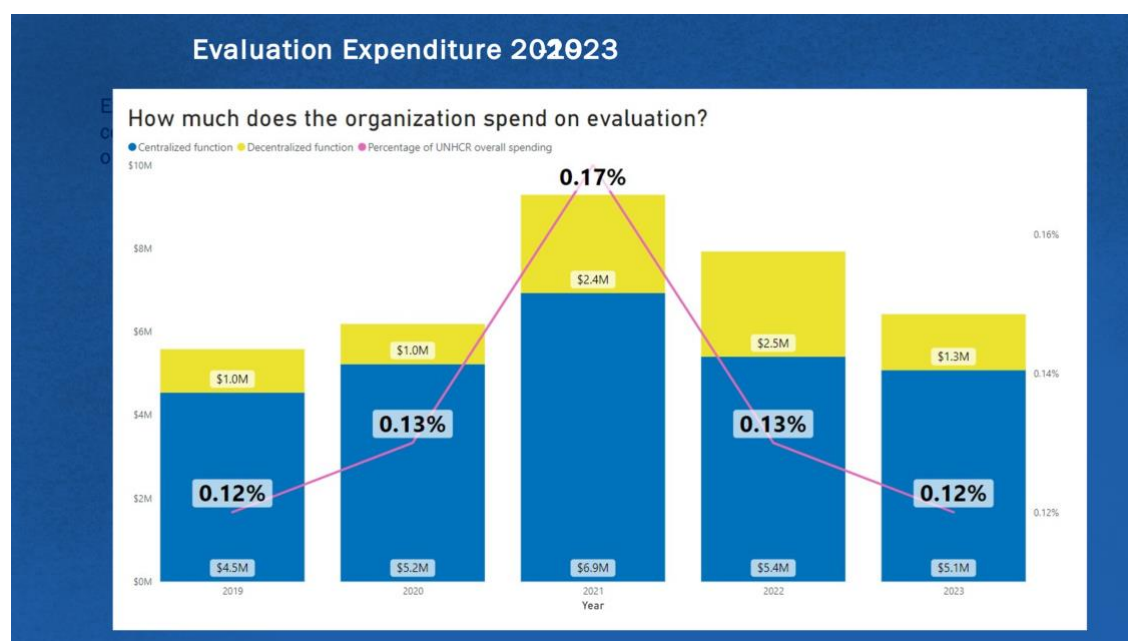
Independence

In terms of independence, the Head of the evaluation office reports to the High Commissioner; and the High Commissioner who, in principle, approves the budget and workplan. The 2021 UNEG/OECD-DAC Independent Peer Review of UNHCR found that the evaluation function falls short of full functional independence as envisaged by UNEG norms and standards. However, the Panel and the Head of the Evaluation Office consider that in practice it is adequate: the Head of Office has sufficient autonomy to manage a pre-agreed budget, select evaluands, and issue evaluation reports. The Review identified a need to further tighten provisions related to the appointment, tenure and dismissal of the position of Head of Evaluation Office (something that has since been addressed in the 2022 evaluation policy).

In some agencies, an evaluation function can rely on an external governing body to nudge the organization into applying learning from evaluations. Governance structures in UNHCR meant this is not quite so relevant here: UNHCR's Executive Committee is an advisory body, and the High Commissioner holds ultimate accountability on behalf of the General Assembly. Thus, responsibility for ensuring the organization learns from evaluation lies more with the office of the High Commissioner and the Senior Executive Team (SET) and less with Member States. Member States still have a role to play, as we set out below.

Resourcing- Human and financial

In terms of resources, the evaluation budget line is submitted to UNHCR's Executive Committee for approval as a separate line, whilst the High Commissioner approves the Evaluation Office's detailed budget internally. The Evaluation Office budget/expenditure had been gradually between 2017 and 2021 but has now levelled off at approximately 6.5 million.



Once spending authority is given it can largely be spent at the discretion of the Head of the Evaluation Office. Evaluation is funded from core resources. Given the organization's 6-8 billion annual budget a year, this only equates to a 0.1% spend on evaluation, below the recommended JIU spend of 0.3-1.5%. Following the 2021 Peer Review of the evaluation function, UNHCR has revised its evaluation policy to outline which types of evaluations should be funded by which type of funding streams.

The Evaluation Office's staffing has increased over the last two years – this growth entirely the result of the creation of Senior Regional Evaluation Officers (SREOs) with direct line reporting to the EvO. Current staffing includes 1 P3, 3 P4 (1 rotational and 2 non-rotation/expert positions), and a P5 managing corporate evaluations and a D1 Head based in Geneva. This is complimented by 5 SREOs (P4s) in 5 of the 7 regional offices. GS staff include 2 resource management staff at the GS5 and GS7 level. This small team is complemented by TA (temporary assignments, consultancies and interns/JPOs).

The Evaluation Office is seeking to align with and support UNHCR's ongoing regionalization and decentralization reforms. In line with its evaluation strategy, the Evaluation Office has expanded its regional presence by posting 5 Senior Evaluation Officers in Regional Bureaux (P4s) to better support decentralized evaluations and it recently published guidance for those conducting decentralized evaluation. These posts are managing country strategy evaluations and advising and quality assuring decentralized evaluations commissioned by management in their region. The intent is to embed these posts within the regional structure but there is a recognition that evaluation is too nascent and that if the posts were funded under the regional budget, they would be more at risk of being cut.

Evaluations can be managed by anyone in the organization and there is not a formal M&E officer position within the organization – which can make it hard to implement effective training for evaluation management. The evaluation office is supporting a global M&E community of practice, co-championed by the Division of Strategic Planning and the Evaluation Office, now offers mentoring support for 217 colleagues around the world.

Quality

The Evaluation Policy introduces a requirement to quality-assure all evaluation processes and products (for both centralized and decentralized evaluations). The Evaluation Office issued “pilot” quality assurance guidelines in 2016 which have since been revised – and established an external quality review service, piloted in 2021 that gives feedback on draft TORs, Irs and Ers.

Annual external assessment exercises have shown that the quality of evaluations has improved from an aggregate rating of 58% in 2018-19 to 67% in 2022. UNHCR has publicly available QA templates for TORs, inception and evaluation reports – adapted from UNEG checklists for the same.

In order to ensure that UNHCR’s evaluations consistently meet UN quality standards, the Evaluation Office has revised its external quality assurance guidance and tools to benefit all evaluations conducted by UNHCR, including those commissioned at the regional and country level. UNHCR has outsourced the post-hoc QA of evaluations to a consulting firm who review evaluations reports and undertake a synthesis of evaluation quality.

UNHCR also have a ‘top-up’ fund that they deploy as needed to improve evaluation quality (i.e. where extra resources are needed/where budget is insufficient).

Relevance and Responsiveness

The number of evaluations has grown steadily from 4 evaluations completed in 2016 to 19 evaluations (13 centralized and 6 decentralized) in 2022 and 26 (12 and 14) in 2023. The increase in decentralized evaluations is linked to changes in UNHCR’s funding; traditionally, the organization had a large unearmarked funding base, but funding is being more earmarked, and the organization is seeing an increasing volume of project evaluations as donor requirements. UNHCR are working to respond to that increased demand by trying to develop a roster of prequalified consultants and have established framework agreements with consulting companies.

In terms of coverage, and in line with the new Evaluation Policy, all major policies, themes, strategic results areas, and geographies of the organization’s operational work will now be evaluated at least once over a 5 to 10-year period. A second coverage target is that all country operations need to undertake some form of evaluative exercise (meaning either a standalone evaluation or take part in a corporate evaluation as a country case study) at least once in a 5-year period. In addition, a recent commitment has been made by Evaluation Directors from across the UN system – to consider systematically in evaluations, how climate change impacts the people it serves, and how development and emergency programmes impact on the environment.

The evaluation workplan is developed collaboratively and there is a corporate workplan and a decentralized plan (for HQ and regionally commissioned evaluations). However, there is a recognition that there needs to be a stronger system in place for early planning of decentralized evaluations.

In 2019, UNHCR started doing Country Strategy Evaluations, which look at the whole of UNHCR’s engagement in a country and there is an increasing demand for these – particularly in the context of multi-year strategic planning (newly introduced in 2022).

Areas of best practice relevant to WHO's evaluation function:

- Establishment of clear coverage norms (all major policies, themes, strategic results areas, and geographies of the organization's operational work will now be evaluated at least once over a 5 to 10-year period, as well as the requirement for countries to take part in an evaluative exercise once every five years).
- Robust external quality assurance and assessment mechanism.
- Evaluation strategy to accompany evaluation policy with clear results framework and ToC.
- Visually enticing annual report on evaluation use, use of infographics.
- Use of a 'top up fund' to add to decentralized evaluation budgets for projects to improve quality.
- Consider role of evaluation staff as political actors to 'infiltrate' management with evaluation messaging.
- Strategic use of joint evaluations for humanitarian response.

UNFPA evaluation function

The following summary draws on available published reports listed below, and draws on a key informant interview with UNFPA Evaluation Office (Deborah McWhinney):

- UNFPA's Annual Report on Evaluation (2022)
- Evaluation Policy (2019), (revised 2024)
- Evaluation Strategy
- Peer Review (2023) and Management Response
- MOPAN Assessment (2017-18)
- Costed Quadrennial Budgeted Evaluation Plan, 2022-2025

Overview

Having been set up as an independent office for the first time in 2013¹⁸, the UNFPA Independent Evaluation Office is now 10 years old. During that period the evaluation function has been steadily evolving and maturing in its capability and influence. According to the latest Annual Report, expenditure on evaluation has increased by 163% from \$3.7m in 2014 to \$9.7m in 2022, and in 2023 accounted for 0.8% of programme expenditure.

The recent strategy for evaluation for 2022-25 seeks to strengthen evaluation further, informed by the strategic context in UNFPA and beyond (see UNFPA strategic plan 2022- 25). It identifies 7 strategic priorities:

- Demand driven evaluation processes.
- Diversification and innovation of evaluation processes and products.
- Quality and credibility of evaluations.
- Enhanced use and utility of evaluations.
- Human-rights based approach, especially gender, youth and disability inclusion.
- Evaluation capacity development.
- Global evaluation advocacy and partnerships to influence evaluation at national level.

The strategy states that

"The evaluation function in UNFPA is moving from a primary focus of reporting on performance to one of generating and communicating value,".

¹⁸ The function had been first set up in 2009 as part of the Oversight Division.

The recent UNEG peer review (2023) found that the function has continued to strengthen since the 2019 policy was introduced and though not a particularly large function it is highly valued for its contribution to corporate level strategies, enhancing programmes and informing country programme documents. It is respected and adds value to decision making. Areas for strengthening include evaluation of humanitarian action and relevance, quality and learning from decentralized evaluations.

The last MOPAN assessment (2017-18) had already found that evaluation independence was well established, accompanied by good quality assurance mechanisms and planning of evaluations to cover all strategic areas. It refers to the consolidation of the IEO and improved evaluation processes, and the effective use of lessons learned syntheses and meta-evaluation to inform decision making.

Policy and enabling environment

The UNFPA evaluation policy was updated in 2019 and again most recently in 2024 and was informed by extensive consultation and external independent strategic review. The review had highlighted the need to align the policy with UNFPA strategic plan and global norms - and the policy does indeed do this by linking to UNFPA's Strategic Plan and UN wide processes (the 2016 QPCR). The evaluation policy is comprehensive and in line with UNEG norms and standards.

The 2024 policy covers all types of evaluations, both centralized and decentralized, sets out clear roles and accountabilities, guiding principles and how evaluation topics are selected. One aspect is the role of the Executive Director in ensuring support and an enabling environment for evaluation, and safeguarding independence, including in how the director of the EO is appointed. Like in UNICEF, the function reports administratively to the ED and functionally to the Executive Board. The roles of the ED and the policy and strategy division, among others in ensuring that the organization supports and uses evaluation effectively are clearly set out.

It is interesting that the policy also puts particular emphasis on UNFPA's partnership role and support to UN wide and global initiatives - including how it contributes to system wide evaluation, joint evaluations, Inter-Agency Humanitarian evaluations and evaluation capacity building.

Systems and practice

UNFPA's Evaluation Quality Assurance and Assessment system is clear and comprehensive. It includes guidance and tools, approval of terms of reference and pre-qualification of evaluators, and a role of evaluation advisers in approval of the inception reports. It sets out quality criteria for assessing reports; and the quality assessment is then carried out through a process which involves independent assessors, run by the IEO.

Planning of evaluations follows a clear process at all 3 levels which supports independence and relevance of evaluations. A quadrennial budgeted evaluation plan covers both centralized and certain types of decentralized evaluations (regional and country evaluations), based on consultations with the Executive Board, senior management and UNFPA offices and other stakeholders. In addition, costed evaluation plans are prepared at regional and country level. At country level these are approved concurrently with CPDs.

The independent Evaluation Office has the authority to determine the scope, design, conduct and commissioning of evaluations, and to submit reports directly to the appropriate decision makers, including the Executive Board.

Coverage norms are set out in the policy. In the 2019 policy the requirement was focused on CPE's, requiring that they be conducted at least once in every two programme cycles, or more often if the quality was unsatisfactory for the last CPE and/or significant changes in country contexts have occurred. This target is met 95% or more for the last

few years. In the latest update of the policy, coverage norms are expected to be applied more broadly across different types of evaluations.

One area of practice which is specifically targeted in the strategy and policy is use of diversified and innovative approaches to evaluation, in at least 50% of centralized evaluations.

Quality

The annual report notes that steady and continuous investments in evaluation have brought sustained results in improving quality, coverage and utility. The assessed quality of evaluations is high, having been 100% for 3 years running and then 96% in 2022. However, there is scope to enhance the effectiveness of the decentralized evaluation function, including the skills and competences of evaluation officers and focal points at country level. One area requiring further attention is to ensure that CPEs are timely and useful.

Evaluation Use and Impact

Management responses are required within six weeks of submission of an evaluation report. This target is consistently met (100% for last 9 years). The implementation of management response actions is also monitored in the annual report. In most regions it is achieved either 95% or 100%. There has been an improvement in the extent to which new CPDs are informed by country programme evaluations. This has been achieved partly because of the EO participating in strategic dialogues and Programme Review Committee meetings.

Financial and Human Resources

As set out in the evaluation policy, UNFPA commits to allocating a minimum of 1.4% and up to 3% of programme expenditure towards evaluation. There is a clear separate budget line in the UNFPA integrated budget. There is also a mechanism for ring-fenced funds to support evaluations in country offices facing financial constraints.

The Evaluation Office has 12 posts, including 10 professional and 1 director. In addition, there are a further 6 regional M&E advisers at P5 level and between 15 and 25 M&E officers or focal points in each region to provide support at country office level. One challenge has been relatively high turnover of regional advisers.

Areas of best practice that can inform WHO evaluation function:

- Strong approach to strategic planning of evaluation, informed by UNFPA strategic plan 2022-25.
- This is a balanced approach starts from demand and intent to create value through evaluation use. It emphasizes both centralized and decentralized evaluation and includes broader partnerships for evaluation.
- Despite its small size the evaluation office has been able to achieve high standards:
 - Comprehensive system of quality assurance (EQAA), is in place despite the relatively small capacity of the independent Evaluation Office.
 - High level of value added and respect from across the organization, well supported by interactions with governing body.
 - KPIs are met typically at 95% to 100% level.
- Financial spend on evaluation has nearly tripled in \$ terms and is monitored relative to programme expenditure (target is ambitious i.e. 1.4%; actual is impressive, ranges from 0.8% – 1.0%).

-
- A strategy to enhance evaluation use through communications and knowledge management was put in place in 2018-21. It aligns with UNFPA's wider strategies on communication and knowledge management and includes a range of approaches such as infographics, videos and podcasts.
 - Coverage norms are being applied for country programme evaluations and this has encouraged an increase in coverage to 100% of target recently.
 - Strong approach to external partnerships on evaluation, including advocacy in areas such as Eval4Action and the Global Parliamentarians Forum. More than half of centralized evaluations are joint or system wide.

UNESCO evaluation function

The following draft summary is based on a review of key documents and an interview with Bernardin Assiene, Director, Division of Internal Oversight Services.

Documents reviewed include:

5. UNESCO evaluation policy, 2022-2029
6. 2019 Peer Review of the UNESCO Evaluation Function
7. 2017-18 MOPAN Assessment
8. 2022 Annual Report on Evaluation

Overview

UNESCO's Evaluation Function is perceived as mature and corporate evaluations are considered to be of high quality, although challenges remain around the quality of decentralized evaluations.

The evaluation function was originally in the Bureau of Strategic Planning and the Evaluation Section (now the Evaluation Office) was established in 2000 as a unit within the Internal Oversight Service (alongside audit and investigation). The head of the Evaluation Office reports to the Director of IOS who is accountable for the quality of UNESCO's evaluation function. UNESCO's evaluation system consists primarily of corporate evaluations, conducted by the IOS Evaluation Office, and decentralized evaluations, managed by other UNESCO entities such as Field Offices, Programme Sectors and Category 1 Institutes.

Policy

A clear framework guiding evaluation is established via UNESCO's Evaluation Policy 2022-2029. The policy outlines the evaluation function's objective to provide credible evidence to support UNESCO to achieve its mandate. It aspires to position evaluation as a management tool for enhancing accountability, promoting organizational learning and knowledge management, and informing decision-making processes.

To support the implementation of the evaluation policy, the Evaluation Office developed an accompanying Evaluation Strategy which focuses on the strategic actions to be taken by the IOS Evaluation Office in support of its expected results. The strategy provides a situation analysis outlining the key challenges, the main result areas and the strategic actions that will be implemented to achieve the results. It includes Key Performance Indicators to help monitor and

manage evaluation results. The strategy also assesses major risks to implementation, along with mitigating measures, resource requirements and arrangements for monitoring, reporting and reviewing the strategy.

Relevance and responsiveness

The number of corporate evaluations has remained steady in the past 4-5 years while the volume of decentralized evaluations has increased steadily. This is credited to the requirements stated in the Evaluation Policy (all UNESCO initiatives funded by voluntary contributions with a budget larger than USD 1.5 million must be subject to an external independent evaluation.) It can also be traced to the outreach work by the Evaluation Office including enhanced evaluation capacity development activities and the Evaluation Focal Point network. In 2023, there were five corporate and 31 decentralized evaluations, compared to 6 and 24 in 2018.

The centralized evaluation workplan is developed every two years in a consultative way, according to strategic priorities, management and board requests. There are no specific coverage norms. For decentralized evaluations, the IOS Evaluation Office extracts the individual workplans from the corporate Core data planner.

Enabling environment and culture

UNESCO is perceived as having a relatively strong culture with regards to evaluation. The IOS-EV Office is encouraging an open conversation on evaluation issues through the delivery of monthly webinars on different topics related to evaluation and open for all interested staff. An updated UNESCO Manual geared primarily to managers of decentralized evaluations offers concrete guidance on the evaluation process. The Evaluation Knowledge Hub a one-stop shop for all things evaluation in UNESCO consolidates the UNESCO, UNEG and other guidance as well as a repository of decentralized evaluations and other support tools. IOS Office delivers targeted and in-depth training events both in-person and online. In 2023, it organized three in-depth trainings in UNESCO Regional and National Offices.

Evaluation Use and impact

The most notable results of UNESCO's recent capacity building efforts are more systematic planning, and resourcing of decentralized evaluations as well as an enhanced use of evaluation findings and recommendations during project design and enhanced implementation rates of recommendations.

To facilitate evaluation use across the organization, UNESCO produces an annual evaluation synthesis report which integrates most decentralized and all corporate evaluations produced in a year. The report contributes to accountability and transparency to Member States and partners, identifies cross-cutting issues and lessons learned, and helps build a stronger evidence base about how programme initiatives, implementation modalities and working methods contribute to the Outcomes and Outputs set for the Organization. It enables UNESCO to build a 'culture of results' so that decisions are based on robust evidence of what works. It also provides reflections on the quality of the evaluation reports themselves.

Furthermore, all evaluations must develop a communication plan which outlines how findings will be disseminated and targeted with user-friendly modalities. All UNESCO evaluations, whether corporate or decentralized, require a Management Response and Action Plan. The Management Response is management's overall acknowledgement of the report findings and recommendations. The Action Plan specifies in detail what actions are needed, and the timeframe and accountability for implementing the recommendation.

Independence

The 2017-18 MOPAN report highlighted the independence of UNESCO's evaluation function as 'Highly Satisfactory'. Structurally, the Evaluation Office is positioned independently from UNESCO management and programmatic functions within IOS, alongside audit and investigation functions. The Evaluation Office has operational independence and a discreet budget allocation, and UNESCO's procedures safeguard the ability of evaluators to access any information or persons. It consults, but autonomously sets the UNESCO corporate evaluation work plan and decides if, and when, to present evaluation reports to the Executive Board, other governing bodies or other appropriate levels of decision making. The Head of Evaluation can submit reports directly to the Executive Board as well as the Director-General. The Director of IOS has full discretion and control over the corporate evaluation budget. To further enhance independence, the Director of IOS is only allowed to serve a term of 6 years and must then leave UNESCO.

Quality

Since 2022, the IOS Evaluation Office developed a tailored quality assurance checklist based on UNESCO specificities and UNEG Norms and Standards. All evaluations (decentralized and corporate) are externally reviewed using this tool.

The 2017-18 MOPAN report and 2020 Peer review of the UNESCO evaluation function noted that corporate evaluation reports are generally of a high standard but that there were quality issues with some decentralized evaluations. As a result of this, the evaluation office has worked to build capacity (training on new evaluation policy, UNEG norms and standards, developing ToRs, hiring consultants, tools and checklist to support evaluation quality). This appears to be bearing fruit as the 2023 Evaluation Synthesis found that there had been a significant improvement in the quality of evaluation reports, particularly with respect to decentralized evaluations. This is a significant development and reflects UNESCO's growing commitment to accountability and learning.

Gender Equality is a corporate UNESCO priority area. For two years in a row UNESCO has 'exceeded requirements' for Performance Indicator 4: Evaluation, which integrates the corporate portfolio. For decentralized evaluations, integration of gender and human rights considerations requires improvement.

Resourcing- Human and financial

The UNESCO evaluation office has 6 Professional posts: (one chief of section, four senior / principal evaluation specialists, one evaluation specialist) on fixed-term contracts. The Evaluation Activity Budget in 2024-2025 is 300k. The corporate evaluations as well as the support to the decentralized function and all other related activities are funded by the '3%' budget. This budget was 1.3 million in 2020-2021; 1.1 million in 2022-2023 and 1.9 million in 2024-2025. This enhances continuity of the function and means its staffing is less precarious against peaks and troughs of programme funding.

In line with best practice across UN agencies, UNESCO sets an overall target of 3% of programme expenditure from both regular programme resources and voluntary contributions as the recommended minimum level of investment in evaluation. The 3% allocation from regular budget in the current biennium has allowed the Evaluation Office to undertake corporate evaluations without significant budget constraint. Corporate evaluations led by the Evaluation Office are financed by the "3%" budget. In addition, all voluntary contribution projects must set aside 3% for evaluation purposes. Any project above \$1.5 million must undergo an independent external evaluation.

The Evaluation Office has also engaged in targeted resource mobilization on a bilateral or multilateral basis for specific evaluations or capacity development activities. Member States and donors are encouraged to voluntarily

contribute resources to a Special Account 'to support the UNESCO evaluation function or specific corporate evaluations under the management of the IOS Evaluation Office'. As part of this strategy, the IOS Evaluation Office also encourages in-kind non-monetary contributions from Member States and other sources (e.g. the private sector) in the form of secondments and/or provision of short-term expertise.

Each field office has an evaluation focal point who is trained by the central evaluation function regarding the expected norms and standards expected but any evaluation work is performed on top of their existing role. This is likely to become an increasing challenge as the volume of decentralized evaluation increases. There are currently no evaluation staff at a regional level, but the evaluation office is advocating for this.

UNESCO has an LTA of evaluation firms for evaluation in the education sector and has a roster of consultants and can undertake open calls to tender as required.

Areas of best practice relevant to WHO's evaluation function:

- Annual synthesis of evaluation finding (mapped to the SDGs) and of evaluation quality.
- Roll out of structured training on evaluation policy/standards to build capacity and understanding of evaluation and its use across the organization.
- Ensure all evaluations include a communication plan which outlines how findings will be disseminated and targeted with user-friendly modalities.
- Ensure implementation guidance has detailed roles and responsibilities with regards to evaluation use.
- Funding of evaluation staff from central human resources budget to ensure continuity/sustainability.
- Consider applications of mandatory 'action plan' to facilitate evaluation use and uptake of recommendations.
- 3% PB funding allocation to evaluation.

UNICEF evaluation function

Overview

UNICEF has a mature and well-established evaluation function with strong coverage both at central and decentralized level, whose credibility has been enhanced by investment in human and financial resources. The function was assessed by the recent peer review as providing high quality, useful and credible evaluations - while noting some areas for enhancement including on resource incentives and accountability at DE level.

The MOPAN assessment 2020 had also found (similarly) that UNICEF has a strong evaluation function with a high degree of independence. It noted that its evaluation plan is well-funded, and coverage has increased over time. MOPAN noted that UNICEF has a robust system in place to ensure the quality of its evaluations and has made efforts to increase staff capacity. UNICEF also has clear accountability mechanisms to ensure that evaluation recommendations are acted upon in a timely manner. The assessment noted that UNICEF has made some progress in using evidence to inform the design of programming but did not yet do so consistently.

Recent work led by the Evaluation Office has sought to take UNICEF's evaluation function to the next level, informed by a 'next generation' vision paper which has informed the policy and taking on ambitious new evaluations such as of the UNICEF Strategic Plan. The 2022 Annual Report notes that contribution of evaluation to the UNICEF oversight and accountability structure was clarified, and significant progress made in the use of innovative evaluation approaches, national evaluation capacity development, and impact evaluation.

However, progress on key performance indicators has been mixed in some respects, with the number of evaluations higher than ever and quality high, although the measured quality of evaluations declined (partly due to changes in how quality is assessed – a stricter assessment framework was introduced, piloted by a new external firm). Discussion with EO staff highlighted that a perennial challenge is ensuring funding, that is actually made available within a given planning period, matches that planned for in the costed workplan. The overall target of 1% is not yet achieved. This affects planning and timely delivery of both centralized and decentralized evaluations.

Policy and enabling environment

The UNICEF evaluation policy is up to date and comprehensive (2023), covers both the centralized and decentralized parts of the function, sets out clear roles and accountabilities and is well linked to UNICEF's overall strategic plan and goals. More work is needed on accompanying guidance/handbook to inform practice and implementation of policy.

- The evaluation function has been subject to regular peer review, the latest being in 2023. It found that:
- UNICEF's progress towards delivering the 2018 Evaluation Policy aims is broadly good, with some areas for enhancement.
- Policies, guidelines, data on evaluation quality, management responses and the existence of a specific funding target for evaluations, are areas of strength in UNICEF's evaluation function.
- The number of evaluations has doubled since 2018.

-
- The function can produce high-quality, useful, and credible evaluations.

The policy helps to ensure independence, credibility and to build an evaluation culture:

- The executive director has clear responsibility for safeguarding independence and promoting evaluation culture.
- The evaluation office led by the D2 Director who is a professional evaluator by training reports administratively to ED and annual to the Executive Board through the annual report and by sharing evaluation findings. There is a strong level of engagement with governing body on evaluation and good access to senior managers.
- Many aspects of an evaluation culture exist and well institutionalized, but further enhancement is required at DE level to support accountabilities and incentives for funding.

Systems and practice

UNICEF has a diverse and growing range of evaluation products, with a very high proportion (around 90%) managed at decentralized level. 199 evaluation products were completed in 2022, including the addition of 23 country reports from multi- country evaluations, a new feature. Evaluability assessments have increased significantly. The number of impact evaluations is another high priority but is rather small and has declined in the last 2 years. UNICEF has a robust system of quality assurance of centralized and decentralized evaluations (GEROS). One key informant suggested (and we agree) that there may be merit in considering a joint approach with other agencies that have similar systems, also allowing smaller functions to piggy-back on a shared system.

There is a clear system of evaluation planning in UNICEF, informed by extensive consultation on selection of evaluation topics and by coverage norms. The Global Evaluation Plan covers 4 years and provides detail on the type of evaluation to be conducted over the strategic cycle. At decentralized level, each country office is responsible for developing a multi-year Costed Evaluation Plan as part of their CPD. The linkages between the Evaluation Office and the regional level are strong.

Coverage of evaluation is strong. The MOPAN assessment noted that there had been a notable increase in evaluation coverage at country level, with 95% of country offices having undertaken an evaluation during 2017-19, compared to 89% in 2016-18. The 2022 annual report notes a further increase in evaluations in the last 2 years. UNICEF has a clear set of coverage norms which are set out in policy, including for example on how frequently country plans should be evaluated and requirements on evaluating all level 3 emergencies. Engagement in joint evaluations and contributions to UNEG are both strong.

Credibility

The MOPAN assessment noted that UNICEF has a strong corporate evaluation function which is independent (operationally and financially) from management. The Office of Evaluation reports directly to the Executive Director and prepares and submits an independent plan for global evaluations. Budget lines for evaluation are required in the UNICEF Integrated Budget and the multi-year work plans of ROs and COs (1%+ of programme resources) As per the policy, the Director of Evaluation is appointed by the Executive Director in consultation with the Board - and has full accountability for the oversight of the evaluation function. He/she is responsible for establishing and updating the evaluation policy, setting performance standards for evaluations and for submitting independent annual reports to the Executive Board on the implementation of the evaluation policy.

The Director of Evaluation has full discretion and control over resources allocated by the Board. The budget for the Evaluation Office is approved by the Executive Board as part of the UNICEF Integrated Budget, thereby safeguarding independence at the highest level, although obtaining the budgeted resourcing in practice is a challenge. Through the Global Evaluation Report Oversight System (GEROS), UNICEF tracks the quality and credibility of evaluation reports across the organization, including evaluation terms of reference (TORs) and real-time feedback on the quality of evaluation design.

At the time of the MOPAN assessment almost all evaluations (99%) were judged as satisfactory or above in 2019, and this high level was maintained through 2021. Quality was reported to have reduced in 2022, although it remains at a high level overall. According to the Annual Report, of the 174 evaluations that were independently assessed by an external firm using GEROS, 3 were rated “exceptional”, compared with 9 in 2021; 41 were “highly satisfactory”, compared with 81 in 2021; and 100 were “satisfactory”, compared with 62 in 2021. Thirty evaluations were rated “fair”, compared with two in 2021. No evaluation received a grade of “unsatisfactory”. This means that 83% were rated satisfactory or better.

Follow up and Use / Utility

The Evaluation Policy requires UNICEF offices to prepare a formal management response to recommendations within 60 days of an evaluation being completed. Management responses identify the level of management agreement with the recommendations, the actions and timeline required to implement them, and the lead person responsible for implementing each recommendation. An integrated system stores all management responses, and the status of the implementation of recommendations is reported annually in the Annual Report on the Evaluation Function.

In 2019, UNICEF launched an Evidence Information Systems Integration (EISI) System to plan and manage the evaluation and research function. The EISI system stores all management responses and is used by management to follow up on the implementation of recommendations. In 2019, 96% of management responses were issued within the agreed timeline, compared to 100% for 2018.

The EISI platform stores evaluations, research and studies conducted since 2014 and is easily accessible to UNICEF staff with advanced search functions.

The 2019 Quality Review of the CPDs found that lessons learned do not sufficiently inform the development of CPDs.

On the procurement side, the evaluation office has recently finalized a new long-term agreement with 24 organizations, including expending to include firms from the Global South and in different areas including on impact evaluation and evaluability assessments.

Financial and Human Resources

A strength of UNICEF (in policy terms) is that it has a concrete target of 1% of programme resources to be spent on evaluation.

- This is tracked in the annual evaluation report to the Executive Board.
- \$65m expenditure on evaluation in 2022 (compared with \$7.2bn programme resources). This had increased by \$15m in 3 years, at a time of sharp increases in programmatic expenditure.

However, the practice is not yet matching the commitment – both for the Evaluation Office’s work in commissioning centralized evaluations and at regional and country level. The process of ensuring that resources are actually made available is ongoing and challenging.

The evaluation function is comparatively well resourced in terms of human resources. Overall staffing of evaluation function is around 40 staff. The Evaluation Office has 26 staff including a new Principal Adviser position in support of the Director. Regional evaluation adviser staffing includes 7 P5 posts. These are supported by teams in each region – in some cases the regional teams are larger than the whole of EVL’s central function staffing in WHO. A challenge is in ensuring that the REAs have the financial resources to deliver what is planned/needed. The peer review notes that at CO level, budgets are not always made available, or predictable, which affects planning.

Areas of best practice that can inform WHO evaluation function

- High level of institutionalization and engagement with governing body and senior management, strong culture of evaluation.
- Up to date and comprehensive evaluation policy informed by external peer review and visioning process for a next generation evaluation function.
- High level of evaluation coverage including at DE level, with a diverse range of evaluation products. Large number of decentralized evaluations.
- GEROS system of quality.
- Investment in financial resources for evaluation and 1% target for evaluation funding relative to programme resources.
- Strong investment in the level of capacity in human resources at EO and regional level and in financial terms (\$65m spend on evaluation).
- Coverage norms in all the key areas of evaluation.
- EISI system for tracking use of evidence and management responses.

(Note that UNICEF and WHO have significant differences in organizational context. UNICEF is more project-driven and is able to call on a wider funding base at national level through its committees. WHO has a stronger focus on normative work. The governance arrangements in WHO, including role of MS, are quite different from UNICEF. The regional committees role in WHO means that accountabilities are driven more from regional level not just globally).

WFP evaluation function

Overview

WFP has an evaluation function that is mature, comprehensive, well-resourced, and highly structured and integrated across central, regional and country level. Evaluation has been a feature in the organization for nearly 60 years and the evaluation policy has evolved through various iterations since 2003. The function has been through 3 independent peer reviews, and these have helped to encourage strategic improvements to every aspect of the function with a high degree of visibility, ownership and support from the Executive Board, Executive Director, and senior managers.

WFP's investment in evaluation has risen sharply over the last few years. By 2023 financial resources were planned to reach over \$34m, or 0.31% of WFP total contributions income.

Policy and Enabling Environment

Key features of the evaluation function include:

- A long established and regularly updated evaluation policy (updated 2022, now in its 4th incarnation) and strategy (2022). Together these provide a clear institutional framework for policy and implementation, with annual reporting to the Executive Board.
- The budgetary framework for evaluation is approved by the Board in the context of the WFP management plan; the Director of Evaluation has full discretion and control over resources allocated to OEV. The appointment of the Director of Evaluation is made by the Executive Director with the approval of the Executive Board.
- Coverage norms are clearly set out for each area of the business e.g. policy, strategy, emergency response, country strategy and impact evaluations.
- Assurance statements issued by the Executive Director based on the statements of directors commit to ensuring the impartial conduct of decentralized evaluations.
- A large number of centralized and decentralized evaluations are conducted. In 2022, 54 evaluations were completed, being 27 centralized and 27 decentralized.
- An unusual feature of WFP's evaluation work is that it has 3 rather than 2 main categories i.e. centralized, decentralized and impact evaluations. WFP is unusual among UN agencies in having a strategy and workplan for impact evaluations.

The UNEG/OECD-DAC Peer Review (2021) concluded that at WFP a “highly strategic independent corporate evaluation function oversees the production of high-quality centralized and decentralized evaluations”. The peer review noted that the Office of Evaluation (OEV) has adapted the evaluation function to keep it aligned with WFP priorities and organizational changes.

The Peer review also noted that WFP has increased the number of centralized and decentralized evaluations, achieving the coverage required by its current coverage norms. The majority of its policies have been evaluated, and OEV is on track with the roll out of country strategic plan evaluations.

The peer review recommended ways to enhance the utility and added value of the evaluation function through adequate and sustained financing; the development of appropriate skills and capacity within the evaluation cadre; the implementation of a more rigorous approach to organizational learning from evaluation; and the strengthening of evaluation partnerships in support of national evaluation capacity and joint and system-wide evaluations in pursuit of nationally owned SDGs.

Systems for quality enhancement and assurance

The Evaluation Quality Assurance System (EQAS) and Decentralized Evaluation Quality Assurance Systems provide comprehensive and structured approaches for quality assurance for centralized and decentralized evaluation including process maps, templates, checklist and technical notes for all evaluation types. All evaluations are subject to independent post hoc quality assessment. Quality is consistently high: in 2022 86% of centralized evaluations (100% in the previous two years), and 96% of decentralized evaluation were satisfactory or highly satisfactory. A comprehensive and detailed Management Response System requires the identification of clear actions, responsibilities, and timeframes for addressing evaluation recommendations. Progress on recommendation implementation is reported annually to the Executive Board.

The annual evaluation report is the primary instrument for reporting on the entire evaluation function and includes an assessment of the quality of all evaluations, progress on key performance indicators for the evaluation function, the performance of regional evaluation units, and OEV's performance against its workplan.

OEV has developed a management information system and a set of key performance indicators that support Board oversight of evaluation across WFP and provide information on progress made towards achieving the outcomes set out in the policy. Areas of reporting include evaluation coverage, the quality of evaluation reports, use of evaluations, evaluation partnerships and joint evaluations, and financial and human resources.

Evaluation use and follow up

The MOPAN report had noted that WFP has a comprehensive repository of evaluations, though since the organization lacks a full and comprehensive Knowledge Management function, uptake and use of lessons learned, and best practices from evaluations depends on the systems for implementing evaluation recommendations. Guidance proposes using lessons learned to inform intervention design in some areas, but this is not comprehensive or systematic.

- Since then, WFP has developed a full Evaluation Communications and Knowledge Management Strategy (2021-26). In relation to evaluation follow up, the 2022 evaluation policy states that:
- All WFP evaluations and management responses will continue to be publicly available. The Executive Board considers all OEV-commissioned evaluations and their management responses. Strengthened mechanisms for following up on management actions in response to evaluation recommendations will include an annual analytical report produced by WFP management on the implementation status of evaluation recommendations for consideration by the Oversight and Policy Committee and the Executive Board.
- The WFP risk and recommendation tracking tool (R2) and its associated improvements in the tracking and reporting of the follow-up to recommendations will enhance the use of evaluations for learning and accountability purposes. To ensure the independence of this process, the Corporate Planning and Performance Division coordinates the development of management responses to centralized evaluations and monitors the implementation of management responses to all evaluation recommendations at the corporate level.

Resourcing

The resources budgeted for evaluation in WFP are large in absolute terms and are explicitly approved and identified up front. They come from various sources but mainly from regular and programme resources. The policy requirement is that

‘WFP will allocate at least 0.4 percent of its total contribution income to its evaluation function. Expenditure on evaluation is expected to rise progressively with the continued growth of the function, particularly at the regional and country levels, reaching up to 0.6 percent of total contribution income in 2026.’

Total financial resources for the evaluation function in 2023 was planned to reach USD 34.3 m, or 0.31% of WFP’s total contribution income in 2023. The intended floor for this is set at 0.4% but has not been achieved due to sharper than expected increases in WFP’s overall income. There are various different sources of evaluation resources including:

- Programme support and administrative resources (PSA), which fund the OEV budget and regional evaluation unit costs;
- Country portfolio budgets, which fund CSP evaluations, decentralized evaluations commissioned by country offices and contributes to data collection costs for impact evaluations;
- A multi-donor trust fund, which receives and channels resources dedicated to impact evaluations; and
- Multilateral resources allocated to the Contingency Evaluation Fund, which supports country offices that face genuine resource constraints to fund evaluations.

The evaluation function budget in 2022 was USD 23.55m, of which USD 15.17m was from PSA. Programme funds from country portfolio budgets (USD 4.5 m) were allocated for CSP evaluations. USD 1.1 m was received through the multi-donor trust fund for impact evaluations, adding to a balance on the fund from previous contributions at the start of the year of USD 2.1 m. The decentralized evaluation function budget was USD 9.3 m in 2022. USD 1.5 m was available for the contingency evaluation fund, to support nine country offices for the conduct of decentralized evaluations and CSP evaluations.

In terms of human resources, in 2022 OEV had 62 staff positions, with a further 24 in regional bureaux, and this has increased over time and in diversity and the proportion in fixed-term posts. The function is headed at D2 level with 2 D1 deputy heads.

According to the most recent annual report on evaluation, 58 centralized evaluations were completed or ongoing in 2022. These include policy, strategic, corporate emergency, impact and country strategy and programme (CSP) evaluations.

CSP evaluations are an important evaluation product in WFP, and in volume terms are a major part of OEV’s work programme. Twenty CSP evaluations were completed in 2022 and a similar number were ongoing or planned. They are the main instrument for institutional accountability and learning related to WFP’s activities at the country level. The evaluation process is timed to ensure that the final report on the evaluation of a CSP is ready when the country office starts designing the new CSP.

Coverage norms are an important feature of evaluation in WFP, something which is likely to be relevant for WHO in learning how it can strengthen its own function. Coverage is monitored for policies/strategies and for corporate emergency responses and the latest annual report for 2022 shows that:

- Just over half of active policies, approved since 2011, having been evaluated as of 2022.
- More than three quarters of the 18 emergency responses had been evaluated.
- CSP evaluation is required in the penultimate year of each CSP. For interim CSPs, an evaluation is required every 5 years for the 10 largest country offices and every 10–12 years for all other country offices. As of 2022, of the 65 first-generation CSPs, 28 had been evaluated, 19 were underway and 10 were planned for evaluation.
- The coverage norm for decentralized evaluations relates to the percentage of country offices with at least one DE conducted in each CSP or interim CSP cycle. In 2022, 69% (18 out of 26 offices) had met this standard.

Areas of best practice relevant to strengthening WHO's evaluation function

- Clarity of governance structures, policy, enabling environment and support/commitment to evaluation from governing body and senior management.
- Strategic approach to building the evaluation function.
- Comprehensive coverage norms with regular monitoring of progress.
- Quality assurance and assessment system (EQAS, DEQAS) including post hoc quality assessment and reporting.
- Strong capacity both at centralized level (OEV) and in regional evaluation units.
- Balanced approach to resourcing, using multiple sources of funding.
- Impact evaluation strategy.

Annex 4: Bibliography

Davies, I. C., & Brummer, J. (2013). Lessons- Learned Study of Peer Reviews of UNEG Evaluation Functions.

International Labour Office (2018). ILO results-based Evaluation Strategy 2018-2021. Programme, Financial and Administrative Section.

Prom-Jackson, S., & Bartsiotas, G. A. (2014). Analysis of the evaluation function in the United Nations System. *Geneva: United Nations Joint Inspection Unit*.

United Nations Children's Fund (2023). Annual Report for 2022 on the Evaluation Function in UNICEF. Executive Board. Annual Session 2023.

United Nations Children's Fund (2023). Draft Revised Evaluation Policy of UNICEF. Executive Board. Annual Session 2023.

United Nations Children's Fund (2022). Plan for Global Evaluations ,2022-2025. Executive Board. First Regular Session 2022.

United Nations Children's Fund (2021). UNICEF Integrated Budget, 2022-2025. Executive Board. 2021 Second Regular Session.

United Nations Children's Fund. Global Evaluation Report Oversight System (GEROS). Guidance Handbook. UNICEF Evaluation Office.

United Nations Children's Fund (2018). UNICEF Procedure on the Implementation of the 2018 UNICEF Evaluation Policy. Document number: EVAL/PROCEDURE/2018/001.

United Nations Development Programme (2019). The Revised UNDP Evaluation Policy.

United Nations Development Programme (2023). Evaluation 2022 Annual Report.

United Nations Development Programme (2019). Independent Review of UNDP Evaluation Policy.

United Nations Development Programme (2021). UNDP Evaluation Guidelines.

United Nations Evaluation Group (2023). UNEG Peer Review of the Evaluation Function of UNICEF.

United Nations Evaluation Group (2023). UNEG Peer Review of UNFPA's Evaluation Function.

United Nations Population Fund (2023). Annual Report on the Evaluation Function, 2022.

United Nations Population Fund (2022). Quadrennial Budgeted Evaluation Plan, 2022-2025.

UNFPA. UNFPA Evaluation Strategy 2022-2025.

UNFPA. UNFPA Response to the UNEG-OECD/DAC Peer Review of the UNFPA Evaluation Function.

UNFPA (2019). UNFPA Evaluation Policy.

UNFPA. Decentralized Programme Level Evaluations (2022-2025).

UNFPA. Strategy to Enhance Evaluation use through Communications and Knowledge Management 2018-2021. UNFPA Evaluation Office.

UNHCR (2021). Peer Review of the Evaluation Function at the Office of the United Nations High Commissioner for Refugees (UNHCR). United Nations Evaluation Group.

OECD DAC Network on Development Evaluation.

UNHCR (2021). UNHCR Management Response- UNEG/DAC Peer Review.

UNHCR (2022). UNHCR Evaluation Service Annual Report 2021, Evaluating what works for the people we serve.

UNHCR (2023). UNHCR Evaluation Service Annual Report 2022, Evaluating what works for the people we serve.

UNHCR. UNHCR Evaluation Workplan 2023/2024.

UNHCR (2022). Policy for Evaluation in UNHCR. Using evidence to drive results towards safeguarding the rights and well-being of people we serve.

World Food Programme (2019). In Brief WFP's Evaluation Function.

World Food Programme (2023). Management Response to the Annual Evaluation Report for 2022. Executive Board. Annual Session.

World Food Programme (2023). Annual Evaluation Report for 2022. Executive Board. Annual Session.

World Food Programme (2022). Management Response to the Annual Evaluation Report for 2021. Executive Board. Annual Session.

World Food Programme (2022). Annual Evaluation Report for 2021. Executive Board. Annual Session.

World Food Programme (2021). Decentralized Evaluation Guidance for Process and Content. Decentralized Evaluation Quality Assurance System.

World Food Programme. WFP Impact Evaluation Strategy (2019-2026). WFP Office of Evaluation.

World Food Programme (2022). WFP Evaluation Policy 2022.

World Food Programme (2021). WFP Evaluation Communications and Knowledge Management Strategy (2021-2026).

World Food Programme (2019). Implementation Status of Evaluation Recommendations.

World Food Programme (2021). Summary Report on the Peer Review of the Evaluation Function at the World Food Programme.

World Food Programme (2021). Peer Review of the Evaluation Function at the World Food Programme. United Nations Evaluation Group. OECD DAC Network on Development Evaluation.

World Food Programme. Annual Evaluation Report 2022 in Review.

World Food Programme. Annual Evaluation Report 2021 in Review.

World Health Organization (2018). Evaluation: Evaluation Policy (2018). Executive Board 143rd Session.

World Health Organization (2022). Terms of Reference of the Global Network on Evaluation (GNE). WHO Evaluation Office.

World Health Organization (2017). Independent Review of the Implementation of the WHO Evaluation Policy and the Framework for Strengthening Evaluation and Organizational Learning. Review Report.

World Health Organization (2022). Support Provided by the Evaluation Office for Decentralized Evaluation in WHO. WHO Evaluation Office.

World Health Organization (2022). Guidance Note for the Role of Quality Assurance Advisor for Decentralized Evaluations. WHO Evaluation Office.

World Health Organization (2022). Implementation Framework of the WHO Evaluation Policy. WHO Evaluation Office.

World Health Organization (2023). Guidance Note on Integrating Health Equity, Gender Equality, Disability Inclusion and Human Rights in WHO Evaluations. WHO Evaluation Office.

World Health Organization (2023). Independent Evaluation of WHO's Results-Based Management Framework. Final Report. WHO Evaluation Office.

World Health Organization. A Framework for Strengthening Evaluation and Organizational Learning in WHO.

World Health Organization (2023). Mid-term Evaluation of the Global Strategy to Eliminate Yellow Fever Epidemics (EYE) 2017-2026.

World Health Organization (2023). Evaluation of WHO 13th General Programme of Work.

World Health Organization (2022). Framework for Evaluations of WHO's Contribution at Country Level. WHO Evaluation Office.

World Health Organization (2023). Practical Guide to Evaluation for Programme Managers and Evaluation Staff. WHO Evaluation Office.

World Health Organization. Update and Proposed Workplan for 2024-2025. Executive Board. 154th Session.

World Health Organization (2023). Evaluation: Annual Report. 153rd Session.

World Health Organization (2022). Evaluation: Annual Report. 151st Session.

Any enquiries about this evaluation should be addressed to:
Evaluation Office, World Health Organization
Email: evaluation@who.int
Website: Evaluation (who.int)