

WHO

*Evaluation
Practice
Handbook*



World Health
Organization

WHO

Evaluation Practice Handbook



World Health
Organization

WHO Library Cataloguing-in-Publication Data

WHO evaluation practice handbook.

1.Evaluation studies as a topic – standards. 2.Program evaluation. 3.Organizational policy.
4.World Health Organization. 5.Handbooks. I.World Health Organization.

ISBN 978 92 4 154868 7

(NLM classification: WA 39)

Acknowledgements

We would like to thank the following WHO staff who contributed their knowledge and expertise to the drafting and review of this handbook, in particular Maria Santamaria, Marie Bombin, Guitelle Baghdadi-Sabeti, Batyr Berdyklychev, Erik Blas, Paloma Cuchi, Jane Dyrhaug, Christophe Grangier, Andre Griekspoor, Stephan Jost, Craig Lissner, Ambrogio Manenti, Sivakumaran Murugasampillay, Isabel Noguer, Armando Peruga, Aiga Rurane, Rosina Salerno, Alan Schnur, Deepak Thapa, João Toledo, David Webb.

In addition, discussions at the inception meeting of the Global Network on Evaluation (GNE) in April 2013 guided the development of the handbook. Members of the GNE include Shambu Acharya, Najeeb Al-Shorbaji, Navneet Bahl, Joseph Cabore, Christopher Dye, Christy Feig, Michael Hill, Eileen Jameson, Michael McCullough, Shanti Mendis, Omer Mensah, Hooman Momen, Patanjali Nayar, Dosithee NgoBebe, Mikael Ostergren, Park Kidong, Frank Paulin, Veronica Riemer, Janna Riisager, Aiga Rurane, Yves Souteyrand, Hans Troedsson, Pieter Van Maaren, Wu Guogao, Roelof Wuite.

© World Health Organization 2013

All rights reserved. Publications of the World Health Organization are available on the WHO web site (www.who.int) or can be purchased from WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel.: +41 22 791 3264; fax: +41 22 791 4857; e-mail: bookorders@who.int).

Requests for permission to reproduce or translate WHO publications –whether for sale or for non-commercial distribution– should be addressed to WHO Press through the WHO web site (www.who.int/about/licensing/copyright_form/en/index.html).

The designations employed and the presentation of the material in this publication 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 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.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Printed in Switzerland

Contents

Message from the Director-General	v
About this handbook	vii
PART ONE. PRINCIPLES AND ORGANIZATION	1
Chapter 1. Evaluation in WHO	1
1.1 Definition and principles of evaluation	1
1.2 Evaluation culture and organizational learning	4
1.3 Participatory approach	5
1.4 Integration of cross-cutting corporate strategies: gender, equity and human rights	5
Chapter 2. Organization of evaluation across WHO	8
2.1 Evaluations at WHO	8
2.2 Roles and responsibilities in implementing WHO's evaluation policy	10
2.3 Financing, planning and reporting on evaluation	13
PART TWO. PREPARING FOR AND CONDUCTING AN EVALUATION	17
Chapter 3. Evaluation planning	18
3.1 Defining evaluation questions and criteria	18
3.2 Preparing terms of reference	25
3.3 Choosing a methodological approach	29
3.4 Estimating resources	37
3.5 Determining the evaluation management structure	39
3.6 Managing conflicts of interest	47
3.7 Establishing an evaluation workplan	47
3.8 Preparing the inception report	48
Chapter 4. Conducting the evaluation	50
4.1 Identifying information needs and data collection methods	50
4.2 Briefing and supporting the evaluation team	56
4.3 Ensuring quality	58
Chapter 5. Reporting	61
5.1 Preparing the draft evaluation report	61
5.2 The final evaluation report	63
Chapter 6. Communication, utilization and follow-up of evaluation results	65
6.1 Communication	65
6.2 Utilization and follow-up of evaluation results	66
References	70
Bibliography	72
Annex 1	
WHO Evaluation policy	75

Annex 2	
Typology of assessments other than evaluation conducted at WHO	85
Annex 3	
Basic components of the different types of assessment in WHO	87
Annex 4	
Checklist for compliance with the WHO evaluation policy	89
Annex 5	
Roles and responsibilities – management responses to evaluations	91
Annex 6	
Terms of reference of the Global Network on Evaluation	93
Annex 7	
Advantages, limitations and methodologies of participatory evaluation	99
Annex 8	
Key elements of the joint evaluation process	107
Annex 9	
Evaluation workplan: criteria for selection of evaluation topics	109
Annex 10	
Checklist for evaluation terms of reference	113
Annex 11	
Methodological approaches to impact evaluation	121
Annex 12	
Core competencies for evaluators	123
Annex 13	
Evaluation workplan template	125
Annex 14	
Typology of in-depth interviews	127
Annex 15	
Checklist for evaluation reports	131
Annex 16	
Glossary of key terms in evaluation	139

Message from the Director-General

I welcome this handbook, which is very timely given the World Health Assembly's endorsement of the new WHO Evaluation Policy in May 2012, and the drive to develop a culture of evaluation at all levels of the Organization as we implement reform and move into the new General Programme of Work.



The *Evaluation Practice Handbook* offers comprehensive information and practical guidance on how to prepare for and conduct evaluations in WHO, and gives guidance on the utilization and follow-up of evaluation results and recommendations. Most importantly, it shows how an evaluation culture can be mainstreamed throughout WHO, outlining stakeholders' responsibilities and supporting our staff to commission or carry out high-quality evaluations in accordance with WHO's policy, that conform to current best practices and the norms and standards of the United Nations Evaluation Group.

Evaluation matters. Too often it has been an afterthought in WHO planning, seen as an optional luxury for well-funded programmes, or done only if a donor requires it. This must now change, so that the role of evaluation is understood as an opportunity for organizational and individual learning, to improve performance and accountability for results, and build our capacity for understanding why some programmes and initiatives work, and why others do not. We should not be complacent. Consistent and high-quality evaluation of our work and Organization is essential, and is a tool that will guide programme planning and implementation. We need to build on the example of those successful WHO programmes that regularly evaluate their performance in order to learn from both success and failure and improve results.

Clearly, the ultimate value of evaluations depends on their findings and recommendations being acted upon. An evaluation must be relevant, credible and impartial. It must have stakeholder involvement in order that the recommendations may be accepted and are implementable. There needs to be an appropriate management response, and evaluation findings need to be disseminated to enhance trust and build organizational knowledge. Monitoring the implementation of recommendations and actions will be done in a systematic way, and progress reported annually to the Executive Board. The WHO evaluation website will include copies of all evaluation reports as part of the overall dissemination strategy.

The Global Network on Evaluation (GNE), an Organization-wide network of staff working together to foster the practice of evaluation at WHO, will play an important

role by capturing the institutional experience in evaluation and knowledge, providing strategic direction, ensuring quality control and analysing evaluation findings and lessons learnt.

Through this comprehensive approach, we hope to inspire confidence in our partners and their constituencies, by demonstrating that WHO has the capacity and readiness to learn from failures as well as successes – thereby improving results and ultimately people’s lives.

This handbook will be adapted for e-learning, and will be continuously updated to reflect the latest best practice. I encourage staff and partners to provide comments and suggestions for its improvement, in the light of their experience.



Dr Margaret Chan
Director-General



About this handbook

Purpose

The purpose of this handbook is to complement WHO's evaluation policy (Annex 1) and to streamline evaluation processes by providing step-by-step practical guidance to evaluation in WHO. The handbook is designed as a working tool that will be adapted over time to better reflect the evolving practice of evaluation in WHO and to encourage reflection on how to use evaluation to improve the performance of projects and programmes and to enhance organizational effectiveness. Its goal is to promote and foster quality evaluation within the Organization by:

- advancing the culture of, commitment to and use of evaluation across WHO;
- assisting WHO staff to conform with best practices and with the norms and standards for evaluation of the United Nations Evaluation Group (UNEG);
- ensuring the quality control of all evaluations commissioned by WHO at all levels;
- strengthening the quality assurance approach to evaluation in WHO.

The handbook focuses on assessments that qualify as evaluation. It does not address in depth other forms of assessment that take place in WHO (see Annex 2 for a typology of assessments conducted at WHO other than evaluation, and Annex 3 which illustrates the basic components of different types of assessment, including evaluation).

Target audience

This handbook is addressed to WHO staff from three different perspectives.

- Broadly, the handbook targets all staff and partner organizations who may use it as a tool to foster an evaluation culture throughout WHO.
- More specifically, the handbook targets all staff who plan, commission and/or conduct evaluations at the different levels of the Organization, who should use the handbook as a tool to ensure high-quality evaluations in WHO.
- In particular, the handbook targets crucial networks for evaluation such as WHO's senior management and the Global Network on Evaluation (GNE), who should disseminate and promote the handbook and encourage compliance with it across the Organization.

Scope and Structure

This handbook clarifies roles and responsibilities in evaluation and documents processes, methods and associated tools. It describes the main phases of an evaluation – i.e. planning, conducting the evaluation, reporting, and managing and communicating outcomes – and provides operational guidance and templates to assist those responsible for evaluations to comply with the Organization’s evaluation policy.

The handbook is divided into two parts:

- **Part One (Chapters 1 and 2)** covers the definition, objectives, principles and management of evaluation in WHO.
- **Part Two (Chapters 3–6)** provides practical guidance on preparing for and conducting an evaluation, detailing the main steps for carrying out a high-quality evaluation in compliance with WHO’s evaluation policy (Fig. 1).

Annexes provide templates, standard documents and a glossary that can be used for the different phases of the evaluation process.



PART ONE. PRINCIPLES AND ORGANIZATION

Chapter 1. Evaluation in WHO

This handbook is based on WHO's evaluation policy which defines the overall framework for evaluation at WHO. It aims to foster the culture and use of evaluation across the Organization and to facilitate conformity of evaluation in WHO with best practices and with UNEG norms and standards for evaluation.

This handbook draws on WHO experience in evaluation and global best practice, consolidated from the principles of UNEG and the Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC), national evaluation associations, United Nations and other multilateral agencies, regional intergovernmental groups and national governments.

1.1 Definition and principles of evaluation

1.1.1 Definition

WHO's evaluation policy is based on the UNEG definition of evaluation (UNEG, 2012b), which is:

“An evaluation is an assessment, as systematic and impartial as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area, institutional performance (...).”

- It focuses on expected and achieved accomplishments, examining the results chain, processes, contextual factors and causality, in order to understand achievements or the lack thereof.
- It aims at determining the relevance, impact, effectiveness, efficiency and sustainability of the interventions¹ and contributions of the Organization.
- It provides evidence-based information that is credible, reliable and useful, enabling the timely incorporation of findings, recommendations and lessons learned into the decision-making and management processes of the Organization.
- It is an integral part of each stage of the strategic planning and programming cycle and not only an end-of-programme activity.

¹ “Intervention” in this handbook refers to projects, programmes, initiatives and other activities that are being evaluated. Evaluation of interventions per se is a research function and not a management function.

In addition to evaluations, WHO undertakes various assessments at the different levels of the Organization for a variety of purposes. Annex 2 presents a typology of such assessment, and Annex 3 illustrates the basic components of different types of assessments, including evaluation.

1.1.2 Principles

WHO's evaluation policy is based on five interrelated key principles that underpin the Organization's approach to evaluation and provide the conceptual framework within which evaluations are carried out.

1.1.3 Impartiality

Impartiality is the absence of bias in due process. It requires methodological rigour as well as objective consideration and presentation of achievements and challenges. Impartiality contributes to the credibility of evaluation and reduces the bias in data gathering, analysis, and formulation of findings, conclusions and recommendations.

All evaluations should be conducted in an impartial manner at all stages of the process. Establishing an ad hoc evaluation management group ensures oversight of the evaluation process (section 3.5).

1.1.4 Independence

Independence is freedom from the control or undue influence of others. Independence provides legitimacy to an evaluation and reduces the potential for conflict of interest that could arise if policy-makers and managers were solely responsible for evaluating their own activities.

Independence must be ensured at three different levels.

- At the organizational level, the evaluation function must be separated from those individuals responsible for the design and implementation of programmes and operations being evaluated.
- At the functional level, there must be mechanisms that ensure independence in the planning, funding and reporting of evaluations.
- At the behavioural level, there must be a code of conduct that is ethics-based (UNEG, 2008a; WHO, 2009a). This code of conduct will seek to prevent and appropriately manage conflicts of interest (section 3.6).

Evaluators should not be directly responsible for the policy, design or overall management of the subject under review. At the same time, there is a need to reconcile the independence of evaluation with a participatory approach. Often when national programmes are being evaluated, members of the evaluation

team include staff of the programmes that are being evaluated, since they are responsible for supporting the evaluation process and methods and, most importantly, for implementing recommendations for programme change and reform. WHO staff performing evaluations should abide by the ethical principles and rules of conduct outlined in the compilation of WHO policies (WHO, 2009a). External contractors should abide by WHO's requirements for external contractual agreements. Evaluators must maintain the highest standards of professional and personal integrity during the entire evaluation process. They are expected to ensure that evaluations address issues of gender, equity and human rights and that they are sensitive to contextual factors such as the social and cultural beliefs, manners and customs of the local environment.

1.1.5 Utility

Utility relates to the impact of the evaluation at organizational level, on programme and project management, and on decision-making. It requires that evaluation findings are relevant and useful, presented in a clear and concise way, and monitored for implementation. Utility depends on evaluation timeliness; relevance to the needs of the project, programme, systems and stakeholders; credibility of the process, methods and products; and accessibility of reports. Utilization-focused evaluations form the basis on which the results of evaluation inform policy and management.

Utility will be ensured through a systematic prioritizing of the evaluation agenda on the basis of established criteria and consultation with relevant stakeholders, systematic follow-up of recommendations, public access to the evaluation reports and/or other products, and alignment with WHO's management framework founded on results-based performance.

1.1.6 Quality

Quality relates to the appropriate and accurate use of evaluation criteria, impartial presentation and analysis of evidence, and coherence between findings, conclusions and recommendations.

Quality will be ensured through:

- continuous adherence to the WHO evaluation methodology, applicable guidelines and the UNEG norms and standards for evaluation (UNEG, 2012b);
- oversight by the ad hoc evaluation management group (section 3.5);
- the peer review process;
- application of a quality assurance system for evaluation (section 4.3, Annexes 4, 10 and 15).

1.1.7 Transparency

Transparency requires that stakeholders are aware of the purpose and objectives of the evaluation, the criteria, process and methods by which evaluation occurs, and the purposes to which the findings will be applied. It also requires access to evaluation materials and products.

In practical terms the requirements of transparency are as follows.

- The commissioner of the evaluation should ensure continuous consultation and involvement with relevant stakeholders at all stages of the evaluation process.
- The evaluation report should contain details of the purpose and objectives, evaluation methodologies, approaches, sources of information, recommendations and costs incurred.
- In accordance with the WHO disclosure policy, evaluation plans, reports, management responses and follow-up reports should be published on the WHO evaluation web site and on the web sites of WHO country and regional offices, as applicable.

1.2 Evaluation culture and organizational learning

There is no single definition of an evaluation culture. It is a multifactorial concept that is applied differently across various institutional settings (OECD, 1998). WHO considers that an evaluation culture is an environment characterized by:

- organizational commitment expressed through institutionalization of the evaluation function in terms of a structure and process;
- widespread support for evaluation demonstrated through the willingness of managers and decision makers to make effective use of policy advice generated in evaluations;
- strong demand for evaluation generated, specified and articulated by internal and external stakeholders;
- appreciation of innovation and recognition of the need for the Organization to continue learning from feedback on results in order to remain relevant;
- continuous development of evaluation competencies, thus ensuring competent evaluators and well-informed commissioners and users;
- readiness to learn from real situations, sharing information not only about success but also about weaknesses and mistakes made.

In order to mainstream this evaluation culture and organizational learning within WHO's decentralized structure, the Organization needs a mechanism to operationalize related activities. The GNE plays a key role as a

platform to exchange information on evaluation issues of common interest across the Organization, and to promote the use of evaluation and of its products through capacity building and through the development of training materials and information sessions. The GNE is thus a critical element for promoting WHO's culture of evaluation (Annex 6).

1.3 Participatory approach

WHO views the participatory approach to evaluation as a continuation of efforts to foster a culture of evaluation that involves stakeholders at all levels of the Organization and partner entities, including the beneficiaries. The participatory approach is one of the crucial components of equity-focused evaluation (UNICEF, 2011). Participatory approaches engage stakeholders actively in developing the evaluation and all phases of its implementation. Those who have the most at stake in the programme – i.e. decision-makers and implementers of the programmes, partners, programme beneficiaries and funders – play active roles, particularly in evaluations that have a strong learning focus.

A participatory approach ensures that evaluations address equity, share knowledge and strengthen the evaluation capacities of programme beneficiaries, implementers, funders and other stakeholders. The approach seeks to honour the perspectives, voices, preferences and decisions of the least powerful and most affected stakeholders and programme beneficiaries. Ideally, through this approach, participants determine the evaluation's focus, design and outcomes within their own socioeconomic, cultural and political environments.

Full-blown participatory approaches to evaluation require considerable resources, and it is therefore necessary to balance the advantages of these approaches against their limitations to determine whether or how best to use such a methodology for conducting an evaluation (Annex 7).

1.4 Integration of cross-cutting corporate strategies: gender, equity and human rights

At its 60th session in May 2007, the World Health Assembly called for more effective ways of mainstreaming cross-cutting priorities of WHO (WHO, 2007). Gender, equity and human rights are crucial to almost all health and development goals.

Consistent with the Director-General's decision to prioritize the mainstreaming of these issues across all levels of WHO, and in accordance with (i) WHO's Constitution (WHO, 2005), (ii) WHO's strategy on gender mainstreaming (WHO, 2009b), and (iii) UNEG guidance on integrating gender, equity and human rights into evaluation work (UNEG, 2011), all future WHO evaluations should be guided by these principles:

- The *human rights-based approach* entails ensuring that WHO strategies facilitate the claims of rights-holders and the corresponding obligations of duty-bearers. This approach also emphasizes the need to address the immediate, underlying and structural causes of not realizing such rights. Civic engagement, as a mechanism to claim rights, is an important element in the overall framework. When appropriate, evaluations should assess the extent to which a given action has facilitated the capacity of rights-holders to claim their rights and the capacity of duty-bearers to fulfil their obligations (UNDP, 2009). Evaluations should also address the extent to which WHO has advocated for the principle of equality and inclusive action, and has contributed to empowering and addressing the needs of disadvantaged and vulnerable populations in a given society.
- *Mainstreaming a gender perspective* is the process of assessing the implications for women and men of any planned action, including legislation, norms and standards, policies or programmes, in all areas and at all levels. It is a strategy for making gender-related concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of policies and programmes in order to ensure that women and men benefit equally and inequality is not perpetuated. Evaluations should assess the extent to which WHO actions have considered mainstreaming a gender perspective in the design, implementation and outcome of the initiative and whether both women and men can equally access the initiative's benefits to the degree intended (WHO, 2011a).
- *Equity in health*. Equity is the absence of avoidable or remediable differences among populations or groups defined socially, economically, demographically or geographically. Health inequities involve more than inequality – whether in health determinants or outcomes, or in access to the resources needed to improve and maintain health – they also include failure to avoid or overcome such inequality in a way that infringes human rights norms or is otherwise unfair.

Mainstreaming gender, equity and human rights principles in evaluation work entails systematically including in the design of evaluation approaches and terms of reference consideration of the way that the subject under evaluation influences gender, equity and human rights. The aim is to ensure the following.

- Evaluation plans assess the evaluability of the equity, human rights and gender dimensions of an intervention and how to deal with different evaluability scenarios.

- Evaluation of gender, equity and human rights mainstreaming includes assessing elements such as accountability, results, oversight, human and financial resources, capacity.
- Evaluation terms of reference include gender-, equity- and human rights-sensitive questions.
- Methodologies include quantitative and qualitative methods and a stakeholder analysis that is sensitive to human rights, equity and gender and is inclusive of diverse stakeholder groups in the evaluation process.
- Evaluation criteria, questions and indicators take human rights, equity and gender into consideration.
- The criteria for selecting members of the evaluation team are that they should be sensitive to human rights, equity and gender issues, in addition to being knowledgeable and experienced.
- The methodological approach of the evaluation allows the team to select and use tools to identify and analyse the human rights, equity and gender aspects of the intervention.

Chapter 2. Organization of evaluation across WHO

2.1 Evaluations at WHO

2.1.1 Commissioning and conducting evaluations

WHO's evaluation policy outlines a corporate evaluation function that coexists with a decentralized approach to evaluation. Corporate evaluations are undertaken by the Office of Internal Oversight Services (IOS). Decentralized evaluations may be commissioned and conducted by different levels of the Organization, such as:

- headquarters-based departments, technical programmes and units;
- regional technical programmes and units;
- WHO country offices (WCOs);
- IOS as custodian of the evaluation function;

In addition, the WHO Executive Board may, at its discretion, commission an evaluation of any aspect of WHO. Other stakeholders such as Member States, donors or partners (partnerships and joint programmes) may also commission external evaluations of the work of WHO for purposes of assessing performance or accountability, or for placing reliance on the work of the Organization.

Evaluations may be conducted by WHO staff, external evaluators or a combination of both.

2.1.2 Types of evaluation in WHO

Depending on their scope, evaluations are categorized as follows.

- **Thematic evaluations** focus on selected topics such as a new way of working, a strategy, a cross-cutting theme or core function, or address an emerging issue of corporate institutional interest. Thematic evaluations provide insight into relevance, effectiveness, sustainability and broader applicability. They require an in depth analysis of a topic and cut across organizational structures. The scope of these evaluations may range from the entire Organization to a single WHO office.
- **Programmatic evaluations** focus on a specific programme. This type of evaluation provides in-depth understanding of how and why results and outcomes have been achieved over several years and examines their relevance, effectiveness, sustainability and efficiency. Programmatic evaluations address achievements in relation to WHO's results chain and require a systematic analysis of the

programme under review. The scope of programmatic evaluations may range from a country to interregional or global levels. Depending on who commissions them, programmatic evaluations may be corporate or decentralized.

- **Office-specific evaluations** focus on the work of the Organization in a country, in a region or at headquarters in respect of WHO's core roles, function, objectives and commitments. Depending on their scope and who commissions them, these evaluations may be either corporate or decentralized.

Depending on who commissions and who conducts them, evaluations may be further categorized as follows.

- **Internal evaluations** are commissioned and conducted by WHO, at times with some inputs from external evaluators.
- **Joint evaluations** are commissioned and conducted by WHO and at least one other organization. Annex 8 provides guidance on the conditions under which joint evaluations are usually undertaken.
- **Peer evaluations** are commissioned by WHO and conducted by teams composed of external evaluators and programme staff. These evaluations combine internal understanding with external expertise and often focus on strengthening national capacities for selected programmes.
- **External evaluations** are typically commissioned by WHO, or by Member States, donors or partners, and are conducted by external evaluators. The evaluations usually assess the performance and accountability of WHO prior to placing reliance on its work. WHO cooperates fully in such evaluations, and the GNE and IOS can facilitate such processes by providing appropriate information and by connecting external evaluation teams with internal WHO units, departments and other stakeholders.

2.1.3 Use of and approach to evaluation

Evaluation needs to address both organizational learning and accountability, and the balance between these two purposes will guide the terms of reference and the methodology of the evaluation. Finding the right balance is an important role of the commissioner of the evaluation. The timing of the evaluation in relation to the programme's life-cycle is also important because it will influence the methodological approaches and the specific focus of the evaluation. Three types of evaluation are possible from this perspective (section 3.3).

2.2 Roles and responsibilities in implementing WHO's evaluation policy

WHO's approach to evaluation is characterized by the principles of decentralization and transparency, and by the availability of a central corporate evaluation function and a global network on evaluation. The roles and responsibilities of the stakeholders and related parties in the implementation of the WHO evaluation policy are outlined below.

IOS is the custodian of the evaluation function. Through its annual report, IOS reports directly to the Director-General and to the Executive Board on matters relating to evaluation in WHO. IOS is responsible for commissioning corporate-level evaluations and for the following functions:

- leading the development of a biennial Organization-wide evaluation workplan;
- informing senior management of evaluation-related issues of Organization-wide importance;
- facilitating the input of evaluation findings and lessons learned for programme planning;
- coordinating the implementation of the evaluation policy across the three levels of the Organization;
- maintaining a system to monitor management responses to evaluations;
- maintaining an online registry of evaluations performed across WHO;
- maintaining a roster of experts with evaluation experience;
- providing guidance material and advice for the preparation, conduct and follow-up of evaluations;
- reviewing evaluation reports for compliance with the requirements of the evaluation policy;
- strengthening capacities in evaluation among WHO staff (e.g. making available standardized methodologies or training on evaluation);
- submitting an annual report on the implementation of the biennial Organization-wide evaluation workplan to the Executive Board through the Director-General;
- supporting the periodic review and updating of the evaluation policy as needed;
- acting as the secretariat of the GNE.

The GNE is a network of staff from all levels of the Organization who act as focal points to support the implementation of the evaluation policy and

promote the culture of evaluation, as well as facilitating information-sharing and knowledge management (Annex 6). In particular, GNE members:

- participate in the preparation of the biennial Organization-wide evaluation workplan and its annual update;
- submit reports of evaluation done in their areas of responsibility to the evaluation registry;
- follow up on the status of management responses to evaluation recommendations;
- act as focal points for evaluation in their respective constituencies;
- champion evaluation throughout the Organization;
- advise programmes across WHO on evaluation issues, as needed.

GNE members are appointed by assistant directors-general at headquarters and by regional directors at regional offices to represent:

- country office level: heads of WHO country offices who have a strong background in evaluation and have the capacity to champion evaluation issues at country level within their region;
- regional level: staff working at the regional level whose current functions include monitoring and evaluation (ideally, these staff could be working in the office of the director of programme management, the assistant regional director or deputy regional director, depending on the regional office);
- WHO headquarters level: staff working at headquarters with responsibilities for monitoring and evaluation within their clusters;
- global level: staff working on monitoring and evaluation within the WHO departments that address cross-cutting issues of special relevance to evaluation, such as Country Collaboration (CCO), Communications (DCO), Gender, Equity and Human Rights (GER), IOS, Knowledge Management and Sharing (KMS), Information Technology (ITT), and Planning, Resource Coordination and Performance Monitoring (PRP).

The Executive Board of WHO:

- determines the evaluation policy and subsequent amendments, as needed;
- provides oversight of the evaluation function within the Organization;
- encourages the performance of evaluations as an input to planning and decision-making;

- provides input to the biennial Organization-wide evaluation workplan on items of specific interest for Member States;
- approves the biennial Organization-wide evaluation workplan;
- considers and takes note of the annual report on the implementation of the biennial Organization-wide evaluation workplan.

The Global Policy Group (GPG):

- is consulted on the proposed contents and subjects prior to finalization of the biennial Organization-wide evaluation workplan;
- ensures that there are adequate resources to implement the biennial Organization-wide evaluation workplan;
- considers the report on the implementation of the biennial Organization-wide evaluation workplan.

The Director-General shall:

- be consulted on the proposed contents and subjects prior to finalization of the biennial Organization-wide evaluation workplan;
- ensure that there are adequate resources to implement the biennial Organization-wide evaluation workplan;
- consider the report of the implementation of the biennial Organization-wide evaluation workplan.

Regional directors and assistant directors-general:

- assist with the identification of topics for the biennial Organization-wide evaluation workplan;
- ensure that financial resources for evaluation are included in projects and workplans;
- ensure that evaluation recommendations relating to their areas of work/programmes are monitored and implemented in a timely manner;
- assign a focal point for evaluation in the region and/or cluster for the GNE.

Programme directors and heads of country offices should:

- ensure that all major programmes are evaluated at least once in their strategic planning life-cycle in accordance with established criteria;
- ensure that all programmes have a well-defined performance framework with a set of indicators, baselines and targets that

contributes to their evaluability for process, outputs, outcomes and impact, as appropriate;

- ensure that evaluations are carried out in accordance with WHO evaluation policy;
- ensure that responsible officers in the programmes prepare management responses to all evaluations and track implementation of the recommendations;
- ensure timely implementation of all evaluation recommendations;
- utilize evaluation findings and recommendations to inform policy development and improve programme implementation;
- through their representative at the GNE, report on evaluation plans, progress of implementation and follow-up of recommendations on at least a six-monthly basis.

The director of PRP at headquarters is responsible for the coordination of tools and systems to provide the information to determine the evaluability of projects, programmes and initiatives, as appropriate.

The Independent Expert Oversight Advisory Committee (IEOAC) provides oversight and guidance on the evaluation function.

2.3 Financing, planning and reporting on evaluation

2.3.1 Financing evaluation

In determining the amount required to finance the evaluation function, other organizations have estimated that 3–5% of the programme budget should be used for evaluation. WHO has adopted these figures, which will be revised in due course. It is the responsibility of the Director-General's Office, regional directors, assistant directors-general, directors of departments and heads of WHO country offices to ensure that resources are adequate to implement their respective components of the Organization-wide evaluation plan. An appropriate evaluation budget needs to be an integral part of the operational workplan of a department, programme and project, and should be traceable in the workplan, along with resource use/expenditures, to facilitate reporting. The appropriate evaluation budget should be discussed as necessary with stakeholders during the planning phase of each project/programme/initiative.

2.3.2 Cost of an evaluation

In its 2008 internal review of evaluative work at headquarters, IOS estimated the direct cost of an evaluation ranged between US\$ 267 000 and US\$ 1.3 million for external evaluations (some impact evaluations have cost over US\$ 3 million), and between US\$ 53 000 and US\$ 86 000 for programme/country evaluations.

2.3.3 The biennial Organization-wide evaluation workplan

The evaluation policy defines a biennial Organization-wide planning and reporting process as part of the Organization's planning and budgeting cycle. The workplan is established in consultation with senior management at headquarters and regions and with country offices, based on established criteria (Annex 9). The biennial Organization-wide evaluation workplan will be updated annually on the basis of the annual report. The workplan is submitted to the Executive Board for approval through the Programme Budget and Administration Committee.

The following categories will be considered in the development of criteria for the selection of topics for evaluation:

- *organizational requirements* relevant to global, international or regional commitments; specific agreements with stakeholders, partners or donors; requests from governing bodies (e.g. global partnership, Millennium Development Goals or a donor request);
- *organizational significance* relating to the priorities and core functions of the General Programme of Work; level of investment; timing since the last evaluation; complexity and associated inherent risks; impact on reputational risk; evaluability (technical, operational); performance issues or concerns in relation to achievements of expected results, such as a significant problem identified in the course of monitoring;
- *organizational utility* relating to a cross-cutting issue, theme, programme or policy question; potential for replication of innovative/catalytic initiatives; utilization of evaluative findings; potential for staff or institutional learning (innovation); degree of comparative advantage of WHO; or changes in the international health landscape and/or in scientific evidence;
- *mandatory evaluations* for programmes and initiatives once in their life-cycle when at least one of the following conditions applies:
 - WHO has agreed to a specific commitment with the related stakeholders over that life-cycle;
 - the programme or initiative exceeds the period covered by one General Programme of Work;
 - the cumulative investment size of the programme or initiative exceeds 2% of the programme budget.

The duration of the programme/initiative as well as the stage in the programme life-cycle needs to be considered when selecting the evaluation.

The identification of evaluations for the biennial Organization-wide evaluation workplan will be coordinated by the GNE through an effective consultation process involving:

- for corporate evaluations: the Director-General, regional directors, advisers to the Director-General;
- for decentralized evaluations: regional directors, advisers to the Director-General, directors and heads of country offices.

2.3.4 Reporting on the biennial Organization-wide evaluation workplan to the governing bodies

IOS coordinates the preparation of an annual evaluation report and presents it to the Executive Board through the Programme, Budget and Administration Committee. The report is reviewed by the GNE for comments and additions, as applicable, before it is finalized by IOS. The report consists of two parts.

- **Part 1** reports on the implementation of the evaluation policy. The report is designed to inform the Organization's governing bodies of progress in the implementation of the biennial evaluation workplan. It conveys information on the status of planned evaluations at both corporate and decentralized levels, and gives a summary account of their main findings and recommendations, as well as lessons learned. The report also gives an account of the functioning of the GNE throughout the year. The report suggests modifications that need to be made to the biennial evaluation workplan as a result of the analysis of progress made in its implementation and resulting findings or comments.
- **Part 2** covers utilization and follow-up of recommendations. The report relates the implementation status of the recommendations of all evaluations included in the evaluation registry and provides details on the level of compliance of WHO's commissioning entities with the follow-up of their respective evaluations. Those who commission an evaluation are ultimately responsible for the use made of the evaluation's findings. They are also responsible for issuing a timely management response through the appropriate assistant director-general at headquarters or through the regional directors and heads of WHO country offices. Management responses should contain detailed information on the actions taken to implement the evaluation's recommendations.

To support analysis and reporting, IOS has established a central tracking process to monitor management responses throughout the Organization.

- The GNE monitors the follow-up of the implementation of evaluations in a systematic manner, coordinating efforts with those who commissioned the evaluations.

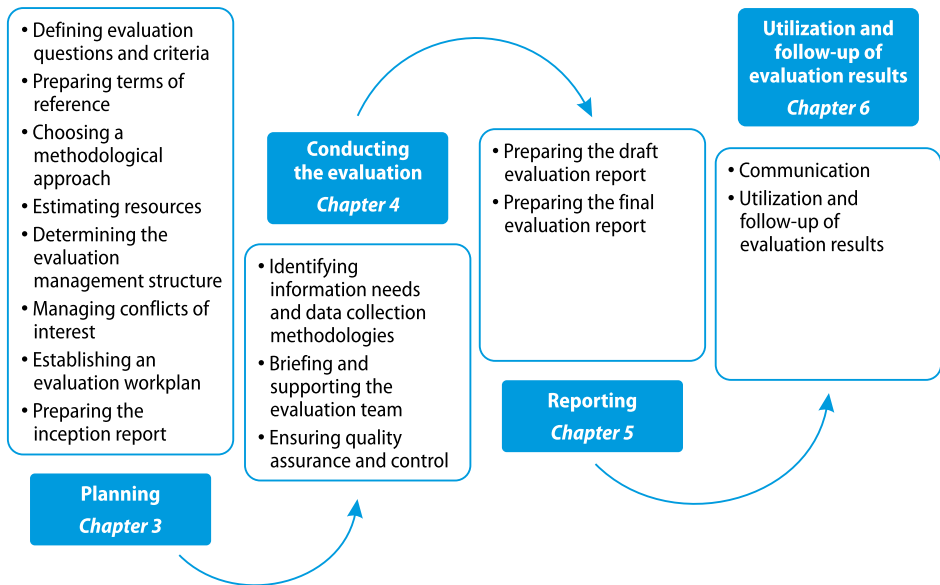
- IOS, based on inputs from the GNE, issues periodic status reports to senior management on progress made in the implementation of recommendations.
- IOS includes a section on implementation of recommendations in its annual evaluation report to the Executive Board.

PART TWO. PREPARING FOR AND CONDUCTING AN EVALUATION

In this second part of the *Evaluation practice handbook*, Chapter 3 outlines a step-by-step approach to the evaluation planning process, Chapter 4 reviews the activities necessary to conduct an evaluation, Chapter 5 provides details of the requirements of reporting, and Chapter 6 describes the utilization and follow-up of evaluation results (Fig. 1).

Fig. 1

Structure of Part Two and the different steps of the evaluation process



Chapter 3. Evaluation planning

This chapter provides a description of the evaluation planning process and outlines the considerations that form the basis of commissioning an evaluation.

The chapter starts by examining the requirements for defining adequate evaluation questions and linking them to evaluation criteria. It also spells out the necessary components of an evaluation plan and provides guidance on drafting clear terms of reference that will hold the evaluation team accountable. The chapter describes the main points to be considered when selecting a methodological approach and ensuring the availability of resources. It also includes guidance on determining a workable evaluation management structure, selecting an evaluation team and preparing an inception report.

3.1 Defining evaluation questions and criteria

The most crucial part of an evaluation is to identify the key questions that it should address. These questions should be formulated by the evaluation commissioner, and should take into account the organizational context in which the evaluation is to be carried out, and the life-cycle of the programme or project. The key questions will serve as the basis for more detailed questions.

Evaluation questions may be:

- **descriptive**, where the aim is to observe, describe and measure changes (what happened?);
- **causal**, where the aim is to understand and assess relations of cause and effect (how and to what extent is that which occurred contributing to and/or attributable to the programme?);
- **performance-related**, where evaluation criteria are applied (are the results and impacts satisfactory in relation to targets and goals?);
- **predictive**, where an attempt is made to anticipate what will happen as a result of planned interventions (will the measures to counter a particular issue in a given area create negative effects in other areas or be taken at the expense of other pressing public health problems?);
- **probing**, where the intention is to support change, often from a value-committed stance (what are the effective strategies for enhancing women's access to care?).

Ideally, evaluation questions should have the following qualities:

- The question must correspond to a real need for information or identification of a solution. If a question is of interest only in terms

of new knowledge, without an immediate input into decision-making or public debate, it is more a matter of scientific research and should not be included in an evaluation.

- The question concerns a need, a result, an impact or a group of impacts. If a question concerns only the internal management of resources and outputs, it can probably be treated more efficiently in the course of monitoring or audit.
- The question concerns only one judgement criterion. This quality of an evaluation question may sometimes be difficult to achieve. However, experience has shown that it is a key factor and that, without judgement criteria clearly stated from the outset, evaluation reports rarely provide appropriate conclusions.

3.1.1 Risks

There are three major risks in drafting evaluation questions (European Commission, 2012):

- gathering large quantities of data and producing sometimes technically sophisticated indicators that make little contribution to practice or policy;
- formulating evaluation questions that are not answerable;
- defining the overarching concern for utility too narrowly and limiting the user focus to the instrumental use of evaluation by managers, rather than including uses that beneficiaries and civil society groups may make of evaluation in support of public health and accountability.

In practice, not all questions asked by evaluation commissioners and programme managers are suitable as evaluation questions; some are complex, long-term and/or require data that are not available. In some cases, questions do not even require evaluation and can be addressed through existing monitoring systems, by consulting managers or by referring to audits or other control systems.

3.1.2 Evaluation criteria

The expected purpose of the evaluation will determine the criteria that need to be included. The criteria may then be used to define the evaluation questions (Table 1). Some of these criteria have been adapted to specific evaluations such as those related to humanitarian programmes (ALNAP, 2006).

Table 1
Evaluation criteria and related questions

Criterion	Measure	Sample questions
Relevance	<p>The extent to which the objectives of an intervention are consistent with the requirements of beneficiaries, country needs, global priorities and the policies of partner organizations and donors. Retrospectively, questions related to relevance may be used to evaluate whether the objectives of an intervention or its design are still appropriate given changed circumstances.</p> <p>The appropriateness of the explicit objectives of the programme in relation to the socioeconomic problems it is intended to address. In ex ante evaluations, questions of relevance are the most important because the focus is on choosing the best strategy or justifying the one proposed. In formative evaluations, the aim is to check whether the public health context has evolved as expected and whether this evolution calls into question a particular objective.</p>	<p>To what extent are the programme objectives justified in relation to needs? Can their <i>raison d'être</i> still be proved? Do they correspond to local, national and global priorities?</p>
Efficiency	<p>How economically resources/ inputs (funds, expertise, time, etc.) are converted to outputs/results.</p> <p>Comparison of the results obtained or, preferably, the outputs produced, and the resources spent. In other words, are the effects obtained commensurate with the inputs? (The terms “economy” and “cost minimization” are sometimes used in the same way as “efficiency”.)</p>	<p>Have the objectives been achieved at the lowest cost? Could better effect be obtained at the same cost?</p>

Table 1 *continued*

Criterion	Measure	Sample questions
Effectiveness	<p>The extent to which the programme/initiative's objectives were achieved or are expected to be achieved, taking into account their relative importance. Effectiveness is also used as an aggregate measure of (or judgement about) the merit of worth of an activity – i.e. the extent to which a programme has achieved, or is expected to achieve, its major relevant objectives and have a positive institutional impact.</p> <p>Whether the objectives formulated in the programme are being achieved, what the successes and difficulties have been, how appropriate the solutions chosen have been and what the influence is of factors external to the programme.</p>	<p>To what extent has the outcome/impact been achieved? Have the intervention and instruments used produced the expected effects? Could more results be obtained by using different instruments?</p>
Sustainability	<p>The continuation of benefits from an intervention after major assistance has been completed, the probability of continued long-term benefits, the resilience to risk of the net benefit flows over time.</p> <p>The extent to which the results and outputs of the intervention are durable. Evaluations often consider the sustainability of institutional changes as well as public health impacts.</p>	<p>Are the results and impacts, including institutional changes, durable over time? Will the impacts continue if there is no more public funding?</p>
Impact	<p>Grouping of the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.</p> <p>The measurement of impact is a complex issue that requires specific methodological tools to assess attribution, contribution and the counterfactual (section 3.3).</p>	<p>Are the results still evident after the intervention is completed?</p>

Source: adapted from definitions in OECD (2010b).

The terms “effectiveness” and “efficiency” are commonly used by managers who seek to make judgements about the outputs and the general performance of an intervention. There is likely to be a fairly large set of questions that will be grouped under these criteria.

3.1.3 Additional considerations

Additional considerations may be taken into account in evaluation, as outlined in Table 2.

Table 2
Additional considerations

Criterion	Measure	Sample questions
Utility	Judges the outcomes produced by the programme in relation to broader public health needs. Utility is an evaluation criterion that reflects the official objectives of the programme. A question on utility should be formulated when programme objectives are unclear or when there are many unexpected impacts. In this case, stakeholders, and in particular intended beneficiaries, should be involved in the selection of utility questions.	Are the expected or unexpected effects globally satisfactory from the point of view of direct or indirect beneficiaries?
Equity	Mainly used to refer to equal access for all population groups to a service without any discrimination. This concept relates to the principle of equal rights and equal treatment of women and men. It means, firstly, that everybody is free to develop personal aptitudes and to make choices without being limited by stereotyped gender roles, and secondly, that particular differences in behaviour, aspirations and needs, between women and men, are not to be valued too highly or considered too critically. The principle of equity may require unequal treatment to compensate for discrimination.	Have the principles of gender equality, human rights and equity been applied throughout the intervention?

Table 2 *continued*

Criterion	Measure	Sample questions
	<p>The evaluation of equity includes the mainstreaming of gender at all stages. Equity can be applied to characteristics other than gender, such as social and economic status, race, ethnicity or sexual preferences.</p>	
Coherence	<p>The need to assess security, developmental, trade and military policies as well as humanitarian policies to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations.</p> <p>Coherence may be difficult to evaluate, in part because it is often confused with coordination. The evaluation of coherence focuses mainly on the policy level while that of coordination focuses more on operational issues.</p> <p>Addressing coherence in evaluations is important where there are many actors and increased risk of conflicting mandates and interests.</p>	<p>To what extent were the different interventions or components of an intervention complementary or contradictory?</p>
Synergy	<p>Several interventions (or several components of an intervention) together produce an impact that is greater than the sum of the impacts they would produce alone.</p> <p>Synergy generally refers to positive impacts. However, phenomena that reinforce negative effects, negative synergy or anti-synergy may also be referred to (e.g. an intervention subsidizes the diversification of enterprises while a regional policy helps to strengthen the dominant activity).</p>	<p>Is any additional impact observed that is the positive or negative result of several components acting together?</p>

Table 2 *continued*

Criterion	Measure	Sample questions
Additionality	<p>The extent to which something happens as a result of an intervention that would not have occurred in the absence of the intervention.</p> <p>Additionality means that an intervention does not displace existing efforts by other players in the same area. In other words, other sources of support remain at least equal to that which existed before the intervention.</p>	To what extent did the intervention add to the existing inputs, instead of replacing any of them, and result in a greater aggregate?
Deadweight	<p>Change observed among direct beneficiaries following an intervention that would have occurred even without the intervention.</p> <p>The difference between deadweight and counterfactual is that the former underscores the fact that resources have funded activities that would have taken place even without public support.</p>	Did the programme or intervention generate outputs, results and impacts that would in any case have occurred?
Displacement	<p>The effect obtained in an area at the expense of another area, or by a group of beneficiaries at the expense of another group within the same territory.</p> <p>Evaluation can best contribute to answering questions about deadweight and displacement when the scale of an intervention or programme is large.</p>	Did the intervention cause reductions in public health development elsewhere?

Sources: Danida, 2012; European Commission, 2012; OECD, 2010a; UNEG, 2012b.

In addition, evaluation questions that derive from these considerations may relate to the unintended negative and positive consequences of interventions. Although programmes have their own logic and goals, they are embedded in policies that define a broader purpose. Evaluators should also consider results of a programme that goes beyond formal goals, such as:

- the experiences and priorities of intended beneficiaries who have their own criteria for programme effectiveness that may not accord with those of programme architects and policy-planners;

- perverse effects that may run counter to programme intentions, reducing opportunities rather than increasing them;
- results suggested by other research and evaluation, possibly drawing on theories of public health or comparative experience in other countries.

3.1.4 Evaluability of evaluation questions

Once the key evaluation questions have been identified, their evaluability has to be considered. A preliminary assessment has to be made of whether the evaluation team, in the time available and using appropriate evaluation tools, will be able to provide credible answers to the questions asked.

For each evaluation question, there is a need to check:

- whether the concept is clear;
- whether explanatory hypotheses can be formulated;
- whether available data can be used to answer the question, without further investigation;
- whether access to information sources will pose major problems.

A number of factors can make a question difficult to answer, such as if a programme is new, if it has not yet produced significant results, or if there are no available data or the data that are available are inappropriate. These reasons may lead to the decision to reconsider the appropriateness of the evaluation questions, to postpone the evaluation or not to undertake it.

Other questions that are relevant and should be considered even before the key questions are identified, include the following.

- Will the recommendations be used? By whom? For what purpose (deciding, debating, informing)? When?
- Is it appropriate to perform such an evaluation at a given time or in a particular political context? Is there a conflict that could compromise the success of the exercise?
- Has a recent study already answered most of the questions?

All evaluation questions need to be narrowed down and clarified so that they are as concise as possible.

3.2 Preparing terms of reference

Once there is agreement on the objectives of the evaluation and the questions that it will need to answer, it is essential to formalize planning by establishing

the terms of reference. The terms of reference serve as the guide and point of reference throughout the evaluation.

While the initial draft of the terms of reference is usually the responsibility of the commissioning office, evaluation terms of reference should be completed in consultation with key stakeholders and evaluation partners in order to ensure that their key concerns are addressed and that the essential audience for the evaluation will view its results as valid and useful.

The terms of reference should be explicit and focused and should provide a clear mandate for the evaluation team regarding what is being evaluated and why, who should be involved in the evaluation process, and the expected outputs (Annex 10).

The terms of reference should be unique to the circumstances and purposes of the evaluation. Adequate time should be devoted to preparing evaluation terms of reference – in particular by the evaluation manager – as they play a critical role in establishing the quality standards and use of the evaluation report.

The outcome, project, thematic area or other initiatives selected for evaluation along with the timing, purpose, duration and scope of the evaluation will dictate much of the substance of the terms of reference. However, because an evaluation cannot address all issues, developing the terms of reference also involves strategic choices about the specific focus, parameters and outputs for the evaluation within available resources.

3.2.1 Content of terms of reference

The terms of reference for an evaluation should include detailed information on the following elements (see Annex 10 for a quality checklist):

- **context** of the evaluation and framework analysis of the subject under evaluation;
- **purpose and objectives** of the evaluation;
- **scope and focus** of the evaluation;
- **evaluation criteria** (relevance, efficiency, effectiveness, sustainability and impact);
- **key evaluation questions**;
- **adherence** to WHO cross-cutting strategies on gender, equity and human rights;
- **users** (owners and audience) of the evaluation results;
- **methodology** (involvement of stakeholders, approach for data collection, and analysis methods required to answer the evaluation questions);

- **evaluation team** (team size, knowledge, skills and required qualifications of evaluators), with specific mention of how conflicts of interests are addressed, and how the independence and objectivity of the team are assured;
- **a detailed workplan** (timetable, organization and budget);
- **deliverables** (including timing of inception/draft/final report, report distribution strategy, follow-up);
- as applicable, composition of the **ad hoc evaluation management group** (including relevant technical requirements).

3.2.2 Context of the evaluation

Evaluations are usually scheduled on completion of a critical phase, or at the end, of the project/programme planning and management cycles. Timeliness is critical to the degree of utility of the results of a given evaluation. It is also important to assess the scheduling of an evaluation in the light of local circumstances since these may jeopardize the course of the evaluation or have a significant bearing on its findings or its relevance.

Moreover, an evaluation may be deferred until other assessments provide clear information on the successes or failures of a project or programme.

3.2.3 Purpose of the evaluation

The initial step in planning an evaluation is to define why the evaluation is being undertaken, i.e. to identify and prioritize the evaluation objectives. This entails determining who needs what information and how the results of the evaluation will be used.

All potential evaluation users, beyond those who commission the evaluation, should be identified. Typically, users would include, according to the situation, responsible WHO staff, implementing partners, partnership members, recipients of the intervention, policy-makers, those with a stake in the project or programme, and individuals in organizations related to the activity being evaluated.

3.2.4 Evaluation scope and focus

Determining the scope of an evaluation includes identifying the nature of the activity and the time period that the evaluation should cover, which may already have been specified with the project or programme during planning.

Other options can be considered, including looking at one activity in several programmes to compare the effectiveness of various approaches, or looking at several projects in a particular area to provide insight into their interactions and relative effectiveness.

An evaluation should:

- describe and assess what output, outcome and impact the activity or service has accomplished, and compare this with what it was intended to achieve;
- analyse the reasons for what happened or the changes that occurred;
- recommend actions for decision-makers to take, based on the answers to the evaluation questions.

An evaluation may focus on different levels of service/project/programme inputs, outputs, processes, outcomes and impacts. A key element underlying evaluations is the need to examine changes and their significance in relation to effectiveness, efficiency, relevance, sustainability and impact (UNICEF, 1991). While any single evaluation may not be able to examine each of these elements comprehensively, they should be taken into consideration.

3.2.5 Deliverables

The terms of reference should clearly describe the deliverables expected from the evaluation exercise, i.e. the evaluation report (inception, draft and final reports). They need to clearly state who will make inputs to the final report, who has final control over the report, the structure and expected content of the report, and the target audience. All these elements should be clearly agreed with the evaluation team leader early in the evaluation process, so that data collection is focused on what is required for the report.

The terms of reference need to consider the following aspects of the report in relation to the report's final format and content (see Annex 10):

- timing of the draft and final report
- need for an executive summary;
- clarity of content;
- suitability of format for the intended audience;
- who will make inputs to the report, and who has final control over its structure and content;
- distribution list and distribution strategy of the report.

During the course of the evaluation, it may become necessary to change some aspects of the expected structure or format of the report on the basis of the actual situation and findings. On occasion, the evaluation team may propose amendments to the terms of reference, provided that those who commissioned the evaluation are informed of the progress of the evaluation and the reasons for revising the terms of reference.

While there is a need to demonstrate adequate flexibility to preserve the relevance of the evaluation, it is important to ensure that any amendments to the terms of reference do not affect the suitability and effectiveness of the evaluation adversely.

3.3 Choosing a methodological approach

3.3.1 Evaluation approach

Each evaluation should have clear objectives, and its purpose and emphasis should be tailored to meet the objectives most appropriately. It should be clear whether the emphasis is on policy, process and management issues; or on results, including outcomes and impact of the interventions under study; or on a mix of both process issues and results at various levels (Danida, 2012).

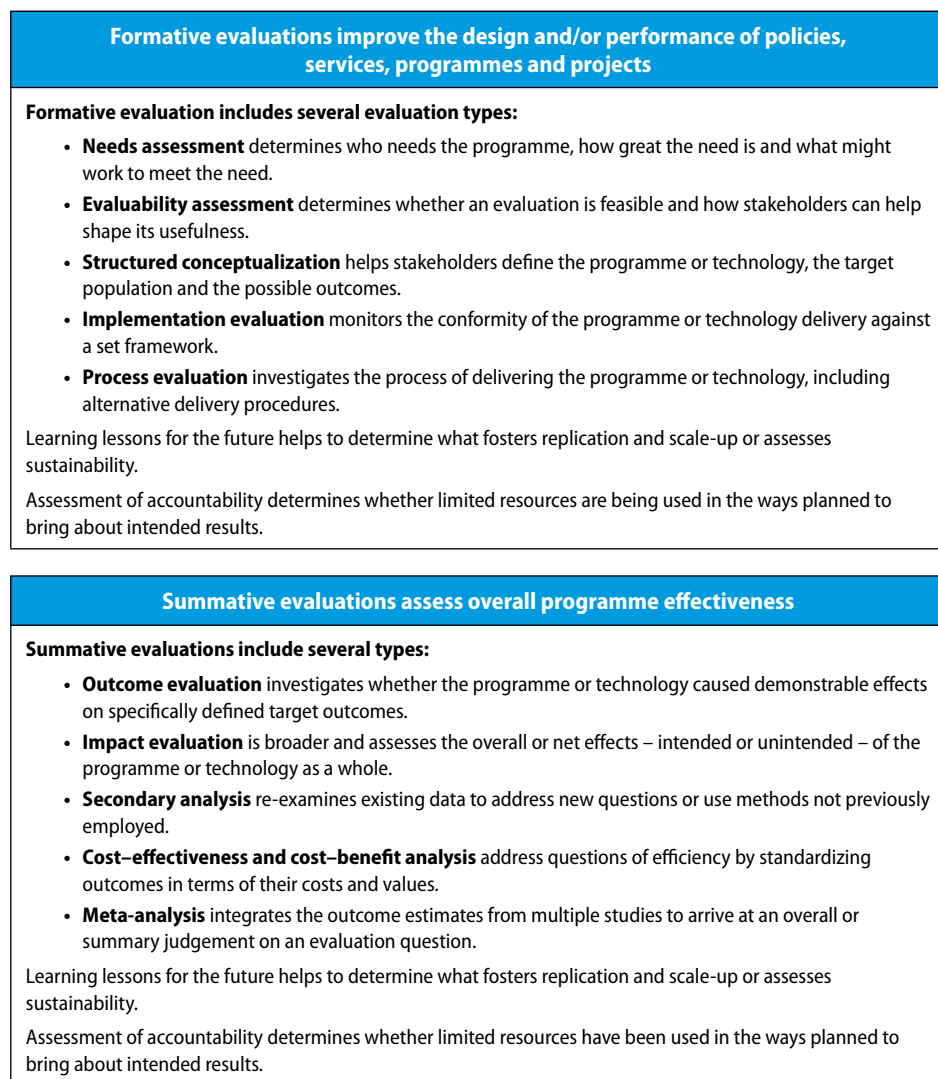
Over the years, evaluation approaches have evolved from classical categorizations such as summative and formative approaches to include combined approaches and impact evaluation.

The purpose, scope and evaluation questions determine the most appropriate way to inform the selection of an evaluation approach.

3.3.2 Formative, summative and real-time evaluations

- **Formative evaluations** (often called process evaluations) are generally conducted during implementation to provide information on what is working and how efficient it is, in order to determine how improvements can be made.
- **Summative evaluations** (often called outcome/impact evaluations) are undertaken: (i) at or close to the end of an intervention, or at a particular stage of it, to assess effectiveness and results; and (ii) after the conclusion of an intervention to assess impact. The timeframe will depend on the type of intervention, and may range from a few months to several years. Fig. 2 outlines methodological approaches commonly used in relation to summative and formative evaluations. Both approaches need to ensure internal consistency as well as consistency with the WHO results chain.
- **Real-time evaluations** are special evaluations that are particularly applied in humanitarian assistance, within three months of the start of a major new international humanitarian response. A real-time evaluation is an evaluation in which the primary objective is to provide feedback in a participatory way in real time (i.e. during the evaluation fieldwork) to those executing and managing the humanitarian response (ALNAP, 2006).

Fig. 2
Methodological approaches to evaluation



Source: adapted from Trochim, 2006.

3.3.3 Evaluation methodology

The evaluation methodology is developed in line with the evaluation approach chosen. The methodology includes specification and justification of the design of the evaluation and the techniques for data collection and analysis (Table 3). The methodology should also address quality.

Table 3
Evaluation methodology – quality aspects and tactics to ensure them

Criterion	Tactic	Phase in which tactic is applied
Construct validity	• Using multiple sources of evidence, triangulation	Data collection
	• Establishing chain of evidence	Data collection
	• Having key informants review draft case-study report	Composition
Internal validity	• Pattern-matching	Data analysis
	• Explanation-building	Data analysis
External validity	• Using analytical generalization:	Data analysis
	– theory in single case-studies	
	– replication logic in multiple case-studies	
	• Using statistical generalization (for relevant embedded subunits)	Data analysis
Reliability	• Using case-study protocol	Data collection
	• Developing case-study database	Data collection

The methodology selected should enable the evaluation questions to be answered using credible evidence. A clear distinction needs to be made between the different result levels with an explicit framework analysis or theory of change. The framework analysis or theory of change should make explicit the intervention logic. In addition to containing an objective–means hierarchy stating input, process (activity), output, outcome and impact, it describes the contribution from relevant actors and the conditions needed for the results chain to happen (OECD, 2010a).

The evaluation methodology addresses:

- the scope of the evaluation (duration of evaluation period and activities to be covered);
- data collection techniques at various levels (countries, sectors, themes, cases);
- data analysis to answer the evaluation questions;
- quality of the evaluation exercise.

The available budget and timeframe influence methodological choices, and the methodology chosen has implications for the budget.

The evaluation methodology selected should ensure that the most appropriate methods of data collection and analysis are applied in relation to the evaluation objectives and questions. Evaluation methodologies are derived

from research standards and methods. Research methods that are both tested and innovative inspire and strengthen the methodological rigour of evaluations (Danida, 2012).

There are many possible methodological combinations, mixing quantitative and qualitative methods, which makes each evaluation unique. WHO encourages triangulation of methods, data collection and data analysis based on a thorough understanding of the evaluation topic. All evaluations must be based on evidence and must explicitly consider limitations related to the analysis conducted (e.g. resulting from security constraints or lack of data).

The level of participation of stakeholders in conducting an evaluation is often crucial to its credibility and usefulness. Participatory approaches are time-consuming, but the benefits are far-reaching. However, the advantages of participatory approaches to evaluation need to be balanced against objectivity criteria and the cost and time requirements of carrying out participatory evaluations (Annex 7).

3.3.4 Determining the information needed to answer the evaluation questions

The evaluation commissioner must make sure that the evaluation team starts by using the information that is available, reviewing existing data and assessing their quality. Some available data can be used to assess progress in meeting the objectives of a project/programme, while other existing data may be helpful for developing standards of comparison. Existing data sources may include:

- WHO governing body documentation (e.g. Executive Board/World Health Assembly resolutions, Programme, Budget and Administration Committee guidance);
- WHO's results-based management framework planning documents (e.g. General Programme of Work, Programme budget, and operational Global Management System workplans), country-level and/or regional-level documents (e.g. country cooperation strategy documents, national health plan and regional programme budget) and, as applicable, the United Nations Development Assistance Framework and/or partnership documents;
- WHO's results-based management monitoring and assessment documents in the context of the new approach to assessing the Twelfth General Programme of Work 2014–2019, from Programme Budget 2014–2015 onwards;
- annual progress reports and notes, previous evaluations/assessments/reviews available at the different levels of WHO or externally, and administrative data maintained by project or programme managers;

- data for developing standards of comparison (possibly including routine reporting systems, surveys, policy analysis and research studies at national, regional and global levels), records or evaluations of similar programmes in different contexts, and reports and publications by donors, universities, research institutions, etc.

As a second step, the minimum amount of new information needed to answer the evaluation questions must be determined. Considerations of cost, time, feasibility and usefulness require that there should be a careful decision as to which data to collect. The evaluation team must ensure that the essential elements are present when planning an evaluation. This can be done by taking the following steps:

- Design a data collection plan, including which indicators to use to measure progress or assess effectiveness. Ideally, indicators should be built into the project or programme design and should be regularly tracked by monitoring. If no indicators are clearly stated, the evaluation must assess which indicators can be used as a proxy or benchmark, and must decide on the evaluability of the project or programme.
- Assess the extent to which indicators will enable the evaluation to judge progress, typically by comparing actual progress with original objectives. Comparisons may also be made with past performance, country-level targets, baseline data, similar services or programmes to help assess whether progress has been sufficient.

3.3.5 Quantitative and qualitative methods

The evaluation commissioner may require the reasons for programme success or failure to be addressed. In this case, the evaluation terms of reference need to make explicit the standard for measuring the programme's evolution. The terms of reference are developed in consultation with the evaluation team and must indicate the appropriate mix of quantitative and qualitative data collection methods needed.

- Quantitative data collection methods use indicators that are specific and measurable, and can be expressed as percentages, rates or ratios. They include surveys, research studies, etc.
- Qualitative data collection methods use techniques for obtaining in-depth responses about what people think and how they feel, and enable managers to gain insights into attitudes, beliefs, motives and behaviours. Qualitative methods have their particular strength in

addressing questions of why and how, enabling evaluators to come up with proposed solutions. They include interviews, SWOT (strengths, weaknesses, opportunities and threats) analysis, group discussions and observation.

Qualitative and quantitative methods should be used in a manner that is interrelated and complementary, whereby quantitative data may measure “what happened” and qualitative data may analyse “why and how” it happened; evaluations may also use a combination of quantitative and qualitative information to cross-check and balance findings.

3.3.6 Assessing impact

The OECD/DAC definition of impact is the “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended” (OECD, 2010b). The UNEG Impact Evaluation Task Force (IETF) refined this definition as follows: “Impact evaluation focuses on the lasting and significant changes that occurred in the short- or long-term, direct or indirect, produced by an intervention or a body of work, or to which the same has contributed” (UNEG, 2013). In the WHO results-based management framework and the Twelfth General Programme of Work, impact refers to the sustainable change in the health of populations to which the secretariat and countries contribute.

The issue of impact has been the subject of intense discussions in the international evaluation community and represents a particular challenge. The OECD/DAC Network on Development Evaluation, the Evaluation Cooperation Group, UNEG, and the European Evaluation Society have discussed appropriate ways and means to address the impact of interventions. Evaluation networks and associations such as the Network of Networks for Impact Evaluation (NONIE), and in particular the International Initiative for Impact Evaluation have been formed to focus on impact evaluation (Leeuw & Vaessen, 2009).

WHO remains engaged in the international debate and research initiatives related to impact evaluations through its continued active participation in the Evaluation Cooperation Group, NONIE, UNEG and other evaluation platforms.

Each WHO department/unit must ascertain the appropriate methodological approach and the most appropriate mix of quantitative and qualitative methods needed to assess impact, depending on the nature, complexity and target beneficiaries of its programmes.

Attribution

Impact evaluations focus on effects caused by an intervention, i.e. “attribution”. This means going beyond describing what has happened to look at causality. Evaluation of impact will, therefore, often require a counterfactual – i.e. an

assessment of the effects the intervention has had compared with what would have happened in the absence of the intervention.

However, interest in attributing results does not mean that a single set of analytical methods should be used in preference to all others in all situations. In fact, the NONIE guidance on impact evaluation underlines that no single method is best for addressing the variety of evaluation questions and aspects that might be part of impact evaluations. Different methods or perspectives complement each other, providing a more complete picture of impact. The most appropriate and useful methodology should be selected on the basis of specific questions or objectives of a given evaluation.

It is rarely possible to attribute the impact of projects/programmes on society rigorously to specific factors or causes. On the one hand, some researchers call for a rigorous assessment of causality through quantitative measures of impact. They advocate the use of randomized control trials and other experimental and quasi-experimental approaches as the gold standard of impact evaluation (Annex 11). On the other hand, a vast amount of literature has demonstrated that these approaches have severe limitations in complex and volatile environments (Patton, 2011).

Impact evaluations are usually based on a combination of counterfactual analysis (e.g. using control groups), before-and-after techniques and triangulation methods. Random sampling is used to select beneficiaries for one-on-one and focus-group discussions, as well as to identify project sites for direct observation purposes. The use of such techniques lays the groundwork for the surveys and case-studies that are then commissioned to collect primary data, especially in cases where the dearth of monitoring and evaluation data acts as a constraint on efforts to arrive at an in-depth appraisal of project impact. Annex 11 presents commonly used methodological approaches to impact evaluation.

Evaluation of the impact of normative work

UNEG defines normative work as:

the support to the development of norms and standards in conventions, declarations, regulatory frameworks, agreements, guidelines, codes of practice and other standard setting instruments, at global, regional and national levels. Normative work may also include support to the implementation of these instruments at the policy level, i.e. their integration into legislation, policies and development plans (UNEG, 2012a).

This concept of normative work also applies to the scientific and technical norms and guidelines produced by WHO at global level and to their application at country level. The amorphous nature of normative work makes the evaluation of its impact seemingly elusive. However, UNEG has developed guidance material

to help UN evaluators and the evaluation community at large to conceptualize, design, plan and conduct impact evaluations of the normative and institutional support work of the United Nations.

The notion of the counterfactual is not meaningful in the context of normative work as the impact of normative work at the macro level occurs in interaction with the activities of others (Van den Berg & Todd, 2011). UNEG stresses the relevance of using the theory of change:

A theory of change, also often referred to as the programme theory, results chain, programme logic model, intervention or attribution logic, is a model that explains how an intervention is expected to lead to intended or observed impacts. It illustrates, generally in graphical form, the series of assumptions and links underpinning the presumed causal relationships between inputs, outputs, outcomes and impacts at various levels (UNEG, 2012a).

There are five stages in developing a theory of change (CTC, 2013):

- identifying long-term goals and the assumptions behind them;
- backwards mapping to work out all the requirements necessary to achieve the goal (outcomes/preconditions);
- identifying the interventions necessary to achieve the desired outcomes;
- developing indicators to measure progress on outcomes and to assess performance;
- writing a narrative to explain the logic of the initiative.

The UNEG guidance material stresses the need to take into full account the complex nature of normative work, which typically involves long-term causality chains, where impact most likely occurs indirectly, involving interaction with the work of other actors and with a variety of other factors. Accordingly, and more than in other types of evaluation, it is important to design an explicit, overarching methodological framework which enables individual methods to be brought together to produce a meaningful overall analysis that can assess the contribution of an intervention, rather than list a set of methods and seek to attribute causality to an intervention.

This approach is not unique to impact evaluation of normative work and is applied to the analysis of public policy in general and to any work of WHO in particular. It should vary for each specific evaluation when assessing the evaluability of the subject item in question. Normative work, however, is often of a complex nature and assessing its impact may be more costly and challenging than carrying out other types of evaluation. In this regard, such evaluations may require evaluators with the experience and skills to work on complex situations,

broad strategies and policies and the evaluators have the experience and skills to interact with senior officials and political leaders.

3.4 Estimating resources

When preparing terms of reference for an evaluation, the commissioner should estimate total financial requirements and ensure that the necessary funding is available. Typically, funds come from the budget that has been allocated to the department, unit, programme or project, and the evaluation would be treated as a task in the annual or biennial operational workplan.

The following factors need to be considered in estimating the budget for an evaluation.

- The timing of the evaluation, determined by its purpose. An evaluation conducted early in implementation, which focuses on design issues rather than outcomes, tends to be less complex and smaller in scope than a heavier exercise conducted at the end of a programme or project cycle that requires more data.
- The scope and the complexity of the evaluation and whether it is a process or outcome/impact evaluation. The time and amount of work needed by the evaluation team to collect and analyse data will affect the cost of the evaluation.
- The availability and accessibility of primary and secondary data and the data collection methods selected. If the data readily available are insufficient, the evaluators will need to spend time and resources to locate or generate information and the evaluation will be more costly.

When preparing the budget for an evaluation, the commissioner needs to take into consideration the estimated direct and indirect costs of the evaluation. These should be built into the evaluation workplan and shared by the different entities involved in the evaluation.

Box 1

Specific issues to consider in estimating the direct cost of an evaluation

1. Institutional or consultancy fees (evaluation consultants and expert advisory panel members, if any):
 - One evaluator or a team? How many in a team? What is the composition (national or international)?
 - How many days will be required for each consultant and adviser?
 - Are the advisory panel members paid (daily fees, honorarium)?
 - What would be the daily rate range for each one of them?
 - What cost is associated with hiring?

Box 1 *continued***2. Travel and logistics**

- How much travel will be required of the evaluation team for briefings at WHO offices, interviews with stakeholders, data collection activities, meetings, etc?
- What will be the mode of travel (air, WHO or project vehicle)? Are there any particular considerations concerning accessibility or security issues?
- For how many days and what are the allowances?
- Any incidentals?
- Requirements for consultations with stakeholders. Are there regular meetings with the steering committee members to discuss progress of the evaluation? Is there a meeting with wider stakeholders to discuss evaluation findings and recommendations? Who will be invited to attend? What is the cost of organizing a meeting (renting venue, travel expenses for participants, refreshments, etc.)?
- Data collection and analysis tools and methods. What are the data collection methods? If surveys and/or questionnaires are used, what is the target population and area to be covered? What resources are required? Are there any particular research needs to complement a detailed analysis of the data collected?
- Are any supplies (office supplies, computer software for data analysis, etc.) needed?

3. Report printing and dissemination

- Publication and dissemination of evaluation reports and other products, including translation costs.

4. Communications

- What are the telephone, Internet and fax usage requirements?
- If surveys or questionnaires are conducted, how will they be administered (online, by mail, by telephone, etc.)?

In the case of a joint evaluation, the commissioner of the evaluation should agree on resourcing modalities with potential donors/agencies or government counterparts (Annex 8).

3.4.2 **Indirect costs**

It is less straightforward to estimate other costs associated with the evaluation. At times these costs can be considerable and in many cases they may exceed the direct costs. They typically include overheads such as:

- internal programme and project staff time (meetings, briefings, interviews, support);
- facilities and office space;
- secretarial support;
- participants' time (e.g. cost of responding to surveys, interviews and review deliverables).

3.5 Determining the evaluation management structure

A clearly defined organization and management structure should be decided upon by the evaluation commissioner at an early stage.

3.5.1 The evaluation commissioner

The evaluation commissioner is the owner of the evaluation. In some partnerships such as the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP) or the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), the commissioner can be the programme's Executive Board or a subcommittee of it. As such, the commissioner provides the general framework within which the evaluation exercise will be conducted. Specifically, the commissioner is responsible for:

- determining which outcomes and impacts of the projects will be evaluated and when;
- identifying the key questions that will frame the evaluation exercise;
- choosing an evaluation manager from among staff to liaise with the evaluation team and take over the day-to-day responsibility for managing the evaluation (see below);
- providing clear advice to the evaluation manager at the outset on how the findings will be used;
- convening an ad hoc evaluation management group where applicable (see below);
- safeguarding the independence of the exercise;
- allocating adequate funding and human resources;
- clearing the inception and final reports;
- responding to the evaluation by preparing a management response;
- implementing the recommendations of the evaluation in a timely fashion.

In the case of smaller evaluations, where it may not be necessary or time/cost-efficient to appoint an evaluation manager or to convene an ad hoc evaluation management group, the evaluation commissioner takes on their roles with regard to the selection and management of the evaluation team and the clearance of the evaluation workplan.

3.5.2 The evaluation manager

Evaluations often involve several institutional levels, countries and administrative settings. It is therefore advised that, for larger evaluations, the evaluation

commissioner appoint a WHO staff member to act as the evaluation manager who will liaise between the evaluation commissioner and the evaluation team leader. In smaller settings, it may not be necessary to appoint an evaluation manager.

The evaluation manager will assume the day-to-day responsibility for managing the evaluation and will serve as a central person connecting other key players. The evaluation team should be able to reach the evaluation manager at any time regarding operational or technical aspects of the evaluation. This will contribute to ensuring that communication remains effective, timely, collegial and efficient.

With the support of the evaluation commissioner and key stakeholders, the evaluation manager plays a central role in:

- developing the terms of reference and the evaluation plan;
- ensuring the selection of the evaluation team;
- managing the contractual arrangements, the budget and the personnel involved in the evaluation;
- organizing the briefing of the evaluation team;
- providing administrative and logistic support to the evaluation team;
- gathering basic documentation for the evaluation team;
- liaising with and responding to the commissioners (and co-commissioners, as applicable);
- liaising between the evaluation team, the ad hoc evaluation management group, the evaluation commissioner and other stakeholders;
- ensuring that the evaluation progresses according to the schedule fixed by the terms of reference;
- reviewing the evaluation workplan and the inception report;
- compiling comments to the evaluation team on the draft report;
- ensuring that the final draft meets quality standards;
- drafting a management response to the final report;
- overseeing final administrative and financial matters, including payments.

The designated evaluation manager should work closely with relevant staff in the department, office, programme or project and, whenever possible, should have experience in evaluation or monitoring and evaluation. The evaluation manager can seek advice from the GNE focal point in their area and from IOS, as appropriate.

3.5.3 The ad hoc evaluation management group

When warranted by the size and complexity of the evaluation, an ad hoc evaluation management group should be assembled by the evaluation commissioner to assist in the conduct and quality control of the evaluation. The group may comprise external experts and/or WHO staff.

The ad hoc evaluation management group should comprise key stakeholders and work closely with the evaluation manager and the evaluation team leader to guide the process. In WHO, the ad hoc evaluation management group typically consists of at least three people selected by the evaluation commissioner at an early stage and before the terms of reference are developed.

In some cases, there is already an entity – such as a steering group, programme or project board or thematic group – that constitutes the group of evaluation stakeholders and from which members of the ad hoc evaluation management group can be drawn to ensure adequate stakeholder participation. In this case, attention should be paid to the potential conflict of interest or compromise of the independence and objectivity of the evaluation process. If such a group does not exist and must be established for the purposes of the evaluation, it is important to maintain the impartiality and validity of evaluation results by ensuring that representation is balanced and that no particular group of opinion dominates. Consideration should be given to gender, geographical coverage and programme and technical knowledge (Box 2).

Box 2

Selection of the ad hoc evaluation management group

The principal determinants in selecting the ad hoc evaluation management group are:

- the familiarity of the candidates with the subject matter being evaluated;
- their independence.

Since the main role of the group is to provide advice to the evaluation team on the subject matter, technical competency in the topic and in evaluation methodology is crucial. However, one risk that needs to be addressed, particularly in evaluations of public health issues, is the possibility that the members of the group are biased towards one particular school of thought and would influence the evaluation design in that direction. It is not always possible to fully ascertain such biases at the selection stage so the evaluation commissioner needs to be aware of that risk throughout the evaluation process. At the practical level, it may be difficult to establish ownership and proper utilization and follow-up of the evaluation report if the evaluatee perceives a bias in the design and management of the evaluation.

Box 2 *continued*

The composition of the group also needs to be balanced by two other factors.

- *The knowledge of the members regarding evaluation process and methodology, and their experience (number of years, relevant areas).* It is important not only that the ad hoc evaluation management group contains members who are familiar with the subject matter, but also that the group includes experts on methodological issues so that they can provide oversight on the rigour and acceptability of the process and methods of data collection and analysis. Including several subject matter specialists and at least one evaluation specialist in the ad hoc evaluation management group provides an ideal mix. The evaluation specialist helps to keep the evaluation process on track. If there are only technical experts, there is a risk that the evaluation may diverge from the workplan.
- *The geographical and gender balance of the group.* The perception that the management group is representative both geographically and in terms of gender can powerfully affect the acceptance and utilization of the evaluation product, particularly for certain programme areas. However, a note of caution is required when considering geographical diversity, as this can increase the budget required for the evaluation. The cost of involving members from all over the world needs to be considered from a value-for-money perspective. It may be possible to organize virtual meetings or use regular scheduled meetings to arrange back-to-back meetings at minimal additional cost.

The functions of the ad hoc evaluation management group include:

- defining or confirming the profile, competencies, and roles and responsibilities of the evaluation manager;
- participating in the drafting and review of the terms of reference;
- approving the selection of the evaluation team;
- approving the evaluation workplan;
- clearing the evaluation inception report;
- overseeing the progress and conduct of the evaluation;
- reviewing the draft evaluation report and ensuring that the final draft meets appropriate quality standards (Annex 15).

The ad hoc evaluation management group should be kept informed of progress by the evaluation manager and should be available to respond to queries from the evaluation team. As the evaluation process progresses, the ad hoc evaluation management group may refer additional ideas and provide suggestions to the evaluation team for consideration.

3.5.4 The evaluation team leader

The evaluation team leader is responsible for:

- implementing the evaluation throughout its life-cycle, including developing a workplan, preparing an inception report, draft and final reports, and briefing the evaluation manager and stakeholders on the progress and key findings and recommendations, as needed;
- supervising the work of the evaluation team;
- liaising with the evaluation manager and the ad hoc evaluation management group, as appropriate.

3.5.5 The evaluation team

Attention must also be given to the required qualifications and competencies of the evaluators. Technical competency in the subject matter is the basic requirement. However, as site visits cover diverse geographical and cultural areas, other “soft” skills are an added advantage. These soft-skill mixes include language proficiency, knowledge of the local context, and interpersonal and intercultural communication abilities. For reference, UNEG has developed guidance documents spelling out evaluators’ core competencies which include criteria such as knowledge of the United Nations context, technical and professional skills, interpersonal skills, personal attributes and management skills (UNEG, 2008b).

The following should be considered in the selection of the evaluation team members (Annex 12):

- technical and sectoral expertise;
- in-depth understanding and experience of quantitative and qualitative evaluation methodology;
- previous experience of conducting evaluations;
- demonstrated analytical and writing skills;
- credibility, impartiality and interpersonal skills.

The evaluation team selection process must ensure that the composition of the team is balanced in terms of opinion, background and gender. It is also necessary to ensure the impartiality and absence of conflicts of interest (see WHO eManual, section VI.2.4) of all members of the evaluation team.

The choice of the team that will carry out the evaluation is important for the quality of the evaluation. An evaluation team may be composed of internal or external evaluators or a combination of both. The number of evaluators in the team depends on a number of factors. Multifaceted evaluations need to be undertaken by multidisciplinary teams. The members selected must bring

different types of expertise and experience to the team. The ideal team should represent a balanced mix of knowledge of evaluation methodology required for that particular evaluation, of the subject to be evaluated, of the context in which the evaluation is taking place or familiarity with comparable situations, and of cross-cutting issues in evaluation such as gender.

There are three main considerations in deciding on the composition of the evaluation team, based on the specific requirements of each evaluation:

i. Internal or external evaluators

Internal evaluators fall into two groups: internal to the programme/location being evaluated; and internal to WHO but from other programmes/locations. External evaluators are national and/or international evaluators not related to the entity being evaluated. WHO may select external evaluators in accordance with the Organization's rules and regulations for procurement. In accordance with WHO's evaluation policy, a database of evaluation experts from which evaluators can be drawn will be established and maintained by IOS and updated on a regular basis.² In evaluations at the country level, the evaluation team should combine national members (who bring the local perspective and experience) and external members (who bring the outside perspective). There are advantages and disadvantages to selecting external evaluators over internal evaluators (Table 4).

Table 4

Advantages and disadvantages of internal and external evaluators

	Advantages	Disadvantages
Internal evaluators	<ul style="list-style-type: none"> • Internal evaluators know WHO, its programmes and operations; they understand and can interpret the behaviour and attitudes of WHO staff and partners; and they may possess important informal information. • They are known to staff, so may pose less threat of anxiety or disruption. • They can more easily accept and promote the use of evaluation results. 	<ul style="list-style-type: none"> • Internal evaluators may lack objectivity and thus reduce credibility of findings. • They tend to accept the position of the Organization. • They are usually too busy to participate fully. • They are part of the authority structure and may be constrained by organizational role conflict.

² The roster is expected to be operational from 2014.

Table 4 *continued*

	Advantages	Disadvantages
	<ul style="list-style-type: none"> • They are often less expensive and their recruitment does not require time-consuming negotiations. • They contribute to strengthening evaluation capability in WHO. 	<ul style="list-style-type: none"> • They may not be sufficiently knowledgeable or experienced to design and implement an evaluation. • They may not have expertise in the special subject matter.
External evaluators	<ul style="list-style-type: none"> • External evaluators may be more objective and find it easier to formulate recommendations. • They may be free from organizational bias. • They may offer new perspectives and additional insights. • They may offer greater evaluation skills and technical expertise. • They are able to dedicate their full time to the evaluation. • They can serve as arbitrators or facilitators between parties. • They can bring the Organization into contact with additional technical resources. 	<ul style="list-style-type: none"> • External evaluators may not know the Organization, its policies, procedures and personalities; and they may be unaware of constraints affecting the feasibility of recommendations. • They may not be familiar with the local political, cultural and economic environment. • They may tend to produce very theoretical evaluation results (if from an academic institution), and may be perceived as adversaries, causing unnecessary anxiety. • They may be costly; they may require more time for contract negotiations, orientation and monitoring; and they may be hoping for further contracts (thus influencing their impartiality).

Source: adapted from UNICEF, 1991.

ii. Institutional or individual evaluators

The cost of hiring individuals to carry out the evaluation is generally less than that of hiring institutions; however, the value added by the branding effect and credibility of institutions also needs to be considered. In most cases, it is the resources available that determine whether institutions can be considered. In public health evaluations, again subject to the availability of resources, the larger evaluations with a global scope tend to be performed by public health academic institutions. Table 5 gives an overview of the advantages and disadvantages of using institutions or individuals.

Table 5
Advantages and disadvantages of individual versus institutional evaluators

	Advantages	Disadvantages
Individual evaluators	<ul style="list-style-type: none"> • Individuals may bring specialized expertise and many years of experience in particular subjects. • The variety of backgrounds of individual team members contributes to debate and discussion that can enrich the exercise. • Individuals may be less expensive than institutions. • Individuals may also be more amenable to last-minute changes in the terms of reference or other arrangements. • Especially for nationals, the evaluation process may provide an opportunity for capacity-development and learning among individual experts. 	<ul style="list-style-type: none"> • Identification of individual consultants is time-consuming and there are risks in selecting evaluation team members solely on the basis of claims made in their applications. • A team of professionals who have never worked together can have difficulty developing a sense of cohesiveness and coherence in their work, and internal conflicts can affect progress. • Changes in the schedule can result in additional costs in fees, per diem and travel arrangements. • Logistics must be provided by the country office.
Institutional evaluators	<ul style="list-style-type: none"> • Fees are agreed as a package that is unlikely to vary, unless there is a change in the terms of reference. • Members of the team are used to working together. • The institution assures the quality of the products. • A multidisciplinary approach is guaranteed (only if required in the contract). • Hiring procedures, although they can be longer than for an individual, are usually easier. • The institution develops the methodology or proposal for the evaluation. • In the event of sudden unavailability (e.g. illness) of an evaluator, the institution is responsible for providing a substitute. 	<ul style="list-style-type: none"> • Fees may be higher, as the institution's overheads will be included. • If the institution has been overexposed to the topic or the Organization, the credibility of the exercise can be compromised. • Team members tend to have similar approaches and perspectives, thereby losing some of the richness of different positions. • Bidding procedures can be lengthy and cumbersome. • Institutions may have difficulty in supplying a mixture of nationals and internationals.

Source: adapted from UNDP, 2009.

iii. Sole sourcing or competitive bidding

WHO financial rules for contracting determine which process to follow. If the evaluation budget exceeds the established threshold (WHO, 2012), competitive bidding procedures have to be followed. An adjudication report justifying the choice of a supplier and the cost is necessary in any case. A full-scale request for proposal or a request for quotations can be considered.

3.6 Managing conflicts of interest

WHO defines a conflict of interest as “any interest declared by an expert that may affect or reasonably be perceived to affect the expert’s objectivity and independence in providing advice to WHO” (WHO, 2011b). As outlined in the WHO evaluation policy, independence can be addressed at the organizational, functional and behavioural levels to reduce the potential for conflicts of interest.

The evaluation commissioner needs to be aware of any dynamics whereby the evaluation team leader may have other objectives for the report (e.g. a scholarly document targeted at the evaluation community) in addition to meeting the requirements of the commissioning organization. This potential source of conflict needs to be addressed adequately, starting as early as possible in the evaluation process.

Evaluators must inform WHO and stakeholders of any potential or actual conflict of interest. External evaluators are expected to sign a Declaration of Interests form. WHO staff must abide by the WHO eManual and the *Ethical principles and conduct of staff: compilation of WHO policies and practices* (WHO, 2009a). WHO staff must inform the evaluation manager of any conflict of interest, in accordance with WHO’s guidelines (WHO, 2011b). In addition, evaluators must follow the requirements of the ethical principles expressed in the UNEG ethical guidelines for evaluators by signing the Code of Conduct for Evaluators in the United Nations System (UNEG, 2008). The evaluation workplan should address any potential or actual conflict of interest and indicate measures put in place to mitigate its negative consequences.

If a conflict of interest is uncovered or arises during the evaluation, the evaluation manager should determine whether the evaluator should be removed and replaced. If the nature of the conflict of interest is such that the evaluation is compromised, the evaluation commissioner should decide whether the evaluation needs to be terminated.

3.7 Establishing an evaluation workplan

The evaluation team should refine the evaluation questions and methodologies, and should specify the schedule of the work to be undertaken in a workplan.

As a first step, the evaluation objectives and questions should be reviewed and should be grouped in a logical manner in the workplan, by subject area, by the data needed to address them logically, by output, outcome or impact, or by other criteria. The workplan should then outline the data that will be collected and how the information gathered will relate to each evaluation question. A schedule is also expected to guide progress of the work. The main objectives of an evaluation workplan are:

- to provide an opportunity for evaluators to build on the initial ideas and parameters set out in the terms of reference to identify what is feasible, suggest refinements and provide elaboration;
- to inform the evaluation by identifying what process is to be followed, who is to do what, what the cost is and when tasks are to be completed;
- to serve as the key reference for managing delivery throughout the performance of the evaluation work.

It is important that the evaluation team and the evaluation commissioner initiate the conduct of the evaluation exercise with a clear understanding of how it is to be carried out. The evaluation workplan should be approved by the ad hoc evaluation management group. The approved workplan functions as an agreement between the evaluation commissioner and the evaluation team, establishing the best approach for meeting the evaluation objectives.

Annex 13 provides an example of a template for an evaluation workplan, specifying objectives, activities, data sources, timeframe and person responsible in the evaluation team.

3.8 Preparing the inception report

For more complex evaluations, the inception report is a useful step for validating the workplan and providing a roadmap for its implementation. The inception report is usually prepared on the basis of the terms of reference, workplan, initial meetings and desk review, to illustrate the evaluation team's understanding of what is being evaluated, including strategies, framework analysis, activities, outputs, expected outcomes and their interrelationships. The inception report should assess the validity of:

- the purpose and scope of the evaluation, clearly stating the objectives and the main elements to be examined;
- the evaluation criteria and questions that the evaluation will use to assess performance and rationale;

- the evaluation methodology, describing the data collection methods and data sources to be used, including the rationale for their selection and their limitations, data collection tools, instruments and protocols, and discussion of their reliability and validity for the evaluation, and the sampling plan, as applicable;
- the evaluation workplan, identifying the key evaluation questions and how they will be answered by the methods selected;
- a revised schedule of key milestones, deliverables and responsibilities;
- detailed resource requirements, linked to the evaluation activities and deliverables detailed in the workplan.

The inception report provides an early opportunity to ensure that the process is taking place as expected on the basis of a common understanding on the part of the evaluation team and the evaluation commissioner, and to refine the terms of reference as needed. To ensure the quality and subsequent acceptability of an evaluation, it is important that the inception report be reviewed as thoroughly as the draft report by the evaluation manager and evaluation commissioner, and by the ad hoc evaluation management group.

Chapter 4. Conducting the evaluation

This chapter outlines the necessary steps to ensure that an evaluation is implemented in accordance with its terms of reference. It describes how to identify information needs, select data collection tools and provide adequate support to the evaluation team. It also describes WHO's quality assurance and control system for evaluation.

4.1 Identifying information needs and data collection methods

4.1.1 Data collection

The evaluation will need to select data collection methods that match its purposes. Table 6 shows the data collection methods most commonly used in evaluation and, for each method described, presents its advantages and challenges.

The most commonly used methods are documentary reviews, direct observation and interviews. While interviews are at the heart of evaluations, evaluators must seek additional sources of information and evidence for issues that will be included in conclusions or recommendations. It is important to differentiate the value that interviews have, depending on the level of expertise or information that they represent; in practice, the opinion of some interviewees is simply more important or better informed than that of others. The interviews can be structured and ask the same questions of all interviewees in the same way. Other interviews follow a snowball method, whereby the observed patterns that emerge after 5–10 interviews are tested with the following interviewees, thus enriching the discussions and interviews. See the typology of in-depth interviews outlined in Annex 14.

The evaluation team needs to consider the following factors in data collection:

- methodological rigour;
- cost–effectiveness;
- validity, reliability and credibility.



Table 6
Summary of common data collection methods used in evaluation

Method	Description	Advantages	Challenges
Monitoring and evaluation systems	<ul style="list-style-type: none"> This is a composite of routine, sentinel surveys and operational research. This is the in-built evaluation system that is described, planned and budgeted for projects, programmes and organizations. Uses performance indicators to measure progress, particularly actual results against expected results. 	<ul style="list-style-type: none"> Can be a reliable, cost-efficient, objective method to assess progress of outputs and outcomes. 	<ul style="list-style-type: none"> Dependent on viable monitoring and evaluation systems that have established baseline indicators and targets and have collected reliable data in relation to targets over time as well as data relating to outcome indicators.
Existing reports and documents	<ul style="list-style-type: none"> Existing documentation, including quantitative and descriptive information about the initiative/project, outputs and outcomes. 	<ul style="list-style-type: none"> Cost-efficient. 	<ul style="list-style-type: none"> Documentary evidence can be difficult to code and analyse in response to questions. Difficult to verify reliability and validity of data.
Questionnaires	<ul style="list-style-type: none"> Provide a standardized approach to obtaining information on a wide range of topics from a large number or diversity of stakeholders to learn about their attitudes, opinions, perceptions and level of satisfaction. 	<ul style="list-style-type: none"> Good for gathering descriptive data on a wide range of topics quickly at relatively low cost. Easy to analyse. Gives anonymity to respondents. 	<ul style="list-style-type: none"> Self-reporting may lead to biased reporting. Data may provide a general picture but may lack depth. May not provide adequate information on context. Subject to sampling bias.

Table 6 continued

Method	Description	Advantages	Challenges
Interviews	<ul style="list-style-type: none"> Solicit person-to-person responses to predetermined questions designed to obtain in-depth information about a person's impressions or experiences, or to learn more about their answers to questionnaires or surveys. 	<ul style="list-style-type: none"> Facilitates fuller coverage, range and depth of information on a topic. 	<ul style="list-style-type: none"> Can be time-consuming. Can be difficult to analyse. Can be costly. Potential for interviewer to bias client's responses. Perceptions/triangulation requirement.
On-site observation	<ul style="list-style-type: none"> Entails use of a detailed observation form to record accurate information about how a programme operated (ongoing activities, processes, discussions, social interactions and observable results as directly observed during the course of an initiative). 	<ul style="list-style-type: none"> Can see operations of a programme as they are occurring. Can adapt to events as they occur. 	<ul style="list-style-type: none"> Can be difficult to categorize or interpret observed behaviours. Can be expensive. Subject to (site) selection bias.
Group interviews	<ul style="list-style-type: none"> A small group of 6–8 people are interviewed together to explore in-depth stakeholder opinions, similar or divergent points of view, or judgements, as well as information about their behaviours, understanding and perceptions of an initiative or to collect information concerning tangible and intangible changes resulting from an initiative. 	<ul style="list-style-type: none"> Quick, reliable way to obtain common impressions from diverse stakeholders. Efficient way to obtain a broad range and depth of information in a short time. 	<ul style="list-style-type: none"> Can be hard to analyse responses. Requires trained facilitator. May be difficult to schedule. Perceptions/triangulation requirement.

Table 6 continued

Method	Description	Advantages	Challenges
Key informants	<ul style="list-style-type: none"> Qualitative in-depth interviews, often one-on-one, with a wide range of stakeholders who have first-hand knowledge about the initiative operations and context. These community experts can provide particular knowledge and understanding of problems and can recommend solutions. 	<ul style="list-style-type: none"> Can provide insight on the nature of problems and give recommendations for solutions. Can provide different perspectives on a single issue or on several issues. 	<ul style="list-style-type: none"> Subject to sampling bias. Must have some means to verify or corroborate information.
Expert panels	<ul style="list-style-type: none"> A peer review, or reference group, composed of external experts to provide input on technical or other substantive topics covered by the evaluation. 	<ul style="list-style-type: none"> Adds credibility. Can serve as added (expert) source of information that can provide greater depth. Can verify or substantiate information and results in topic area. 	<ul style="list-style-type: none"> Cost of consultancy and related expenses, if any. Must ensure impartiality and that there are no conflicts of interest.
Case studies	<ul style="list-style-type: none"> Involves comprehensive examination through cross-comparison of cases to obtain in-depth information with the goal of fully understand the operational dynamics, activities, outputs, outcomes and interactions of a project or programme. 	<ul style="list-style-type: none"> Useful to fully explore factors that contribute to outputs and outcomes. 	<ul style="list-style-type: none"> Requires considerable time and resources not usually available for commissioned evaluations. Can be difficult to analyse.

Source: UNDP, 2009.

4.1.2 Data quality

Two main criteria determine the quality of data (Bamberger, Rugh & Mabry, 2006):

- Reliability refers to consistency of measurement (e.g. ensuring that a particular data collection instrument such as a questionnaire will elicit the same or similar responses if administered under similar conditions).
- Validity refers to accuracy in measurement (e.g. ensuring that a particular data collection instrument actually measures what it was intended to measure). It also refers to the extent to which inferences or conclusions drawn from data are reasonable and justifiable.

There are three broad strategies to improve reliability and validity that an evaluation should address (UNDP, 2009):

- Improve the quality of sampling (to ensure greater representativeness).
- Improve the quality of data gathering (ensure that questionnaires, interview schedules, observation protocols or other data-gathering tools are tested, such as by a pilot approach, and that the evidence gathered is reviewed for accuracy and consistency).
- Use mixed methods of data collection and build in strategies (e.g. triangulating or multiple sources of data) to verify or cross-check data using several pieces of evidence, rather than relying on only one source.

Credibility concerns the extent to which the evaluation evidence and the results are perceived to be valid, reliable and impartial by the stakeholders, particularly the users of the evaluation results.

4.1.3 Analysis and synthesis of data

Data analysis is a systematic process that involves organizing and classifying the information collected, tabulating it, summarizing it, and comparing the results with other appropriate information to extract useful information that responds to the evaluation questions and fulfils the purposes of the evaluation. Data analysis seeks to detect patterns in evidence, either by isolating important findings or by combining sources of information to reach a broader understanding. It is the process of deciphering facts from a body of evidence by systematically coding and collating the data collected, thus ensuring their accuracy, conducting statistical analyses as needed and translating the data into usable formats or units of analysis related to each evaluation question.

Fig. 3 shows the different stages of data analysis and synthesis that build the evaluation process from the analysis plan, the interpretation of findings, to the drawing of conclusions and the formulation of recommendations and of lessons learned.

Fig. 3

Steps to data analysis and synthesis

Analysis plan

- The analysis plan should be built into the evaluation design and workplan detailed in the inception report. It is an essential evaluation tool that maps how the information collected will be organized, classified, interrelated, compared and displayed relative to the evaluation questions, including what will be done to integrate multiple sources, especially those that provide data in narrative form, and any statistical methods that will be used to integrate or present the data (e.g. calculations, sums, proportions, cost analysis, etc.). Possible challenges and limitations of the data analysis should be described. The analysis plan should be written in conjunction with selecting data collection methods rather than afterwards.

Interpreting the findings

- This is the process giving meaning to the evaluation findings derived from the analysis. It extracts from the summation and synthesis of information derived from the facts, statements, opinions and documents and turns findings from the data into judgements about results. Recommendations for future actions are made on the basis of those conclusions. Interpretation is the effort of determining what the findings mean, making sense of the evidence gathered in an evaluation and its practical applications for effectiveness.

Drawing conclusions

- A conclusion is a reasoned judgement based on a synthesis of empirical findings or factual statements corresponding to specific circumstances. Conclusions are not findings, they are interpretations that give meaning to the findings. Conclusions are considered valid and credible when they are directly linked to the evidence and can be justified on the basis of appropriate methods of analysis and synthesis to summarize findings.
- Conclusions should:
 - address the evaluation's stated objectives and provide answers to the evaluation questions;
 - consider alternative ways to compare results (such as comparison with programme objectives, a comparison group, national norms, past performance or needs);
 - generate alternative explanations for findings and indicate why these explanations should be discounted;
 - form the basis for recommending actions or decisions that are consistent with the conclusions;
 - be limited to situations, time periods, persons, contexts and purposes for which the findings are applicable.

Fig 3 *continued*

Making recommendations

- Recommendations are evidence-based proposals for action aimed at evaluation users. Recommendations should be based on conclusions. However, forming recommendations is a distinct element of evaluation that requires information beyond what is necessary to form conclusions. Developing recommendations involves weighing effective alternatives and policy, funding priorities, etc. within a broader context. It requires in-depth contextual knowledge, particularly about the organizational context within which policy and programme decisions will be made and the political, social and public health context in which the initiative will operate. Recommendations should be formulated in a way that will facilitate the development of a management response. They must be realistic and must reflect an understanding of the evaluation commissioner's organization and potential constraints to follow-up. Each recommendation should clearly identify its target group and stipulate the recommended action and rationale.

Lessons learned

- Lessons learned comprise the new knowledge gained from the particular circumstances (initiative, context outcomes and even evaluation methods) that is applicable to and useful in other similar contexts. Frequently, lessons learned highlight strengths or weaknesses in preparation, design and implementation that affect performance, outcome and impact.

Source: CDC (1999), UNDP (2009).

In the event that evaluators identify evidence of fraud, misconduct, abuse of power and/or violation of rights, they should confidentially refer the matter to the appropriate level of line management and/or Director IOS, in accordance with WHO's fraud prevention policy (WHO, 2005b). Evaluations should not substitute or be used for investigative purposes and decision-making in individual human resources matters.

4.2 Briefing and supporting the evaluation team

The success of an evaluation depends on the level of support and cooperation provided by the evaluation manager to the evaluation team. Supporting the evaluation team should not interfere with the evaluation process in ways that could jeopardize the evaluation's independence.

In particular for external evaluations, maintaining the relevance of the final report and especially its recommendations is a major concern. From the evaluation commissioner's perspective, proposing incremental progress may be more acceptable and effective than facing more radical change which may put at risk the entire programme management and affect the report's acceptability. Thus, there is the need to ensure that the report is not only accurate and complete but also relevant and effective for both the valuee and the evaluation commissioner.

There are risks of misunderstandings between the evaluation team and the programme management and implementers. Where programmes are carried out in difficult, or even dangerous, political and geographical situations, progress may be very limited but may nevertheless be better than in other programmes in the same location. In this situation, an insensitive report criticizing reduced programme achievements or non-achievement of expected results on time despite valid reasons may create disagreements.

4.2.1 **Managing the evaluation team**

In this regard, it is essential that the evaluation manager:

- organizes the briefing of the evaluation team on the purpose and scope of the evaluation and explains the expectations of the evaluation commissioner and the evaluation stakeholders in terms of standards of quality of the process and evaluation products (relevant evaluation policy guidelines and quality standards should be made available to them, and it is of particular importance that the evaluators should be requested to follow WHO (WHO, 2009a) and UNEG ethical principles (UNEG, 2008a);
- ensures that all information is made available to the evaluation team and provides support in case the team encounters difficulty in gathering the required data in the process of the evaluation;
- provides a preliminary list and contact information of stakeholders that the team should meet, as required by the evaluation team leader;
- introduces the evaluation team to the partners and stakeholders to facilitate initial contact;
- arranges meetings, interviews and field visits, as applicable, but does not participate in them as this could hinder the evaluation's independence;
- maintains communication through the evaluation assignment in order to be able to provide early troubleshooting in case difficulties are encountered by the evaluation team;
- provides comments and quality assurance on the workplan and the inception report prepared by the evaluation team;
- ensures security of consultants, stakeholders and other accompanying WHO staff as required;
- provides support in the planning of logistic arrangements for the evaluation team.

4.2.2 Operational support

Depending on the terms of the contract, in many cases it is the responsibility of the evaluation commissioner and/or evaluation manager to support the evaluation team with logistics.

Good logistics and administration will assist the evaluation team to meet the appropriate persons and to observe the required places and practices. In addition, any time spent by the evaluation team on logistics and administration may take time away from its central work.

Examples of logistic aspects to consider when planning for a field visit by the evaluation team include:

- informing the country office/evaluee about the evaluation and requirements, and obtaining their cooperation;
- providing lists of key stakeholders with their area of expertise and the extent of their collaboration;
- arranging for relevant WHO staff to brief the evaluation team on the local situation and conditions;
- arranging for a debriefing by the evaluation team before completing the field visit;
- working with the evaluation team on a selection of stakeholders to survey/interview;
- scheduling local meetings with key informants;
- providing travel (by air or other transportation) reservations;
- providing hotel reservations;
- obtaining visas, security clearances and letters of invitation;
- acting as back-up in case of any emergencies or unexpected developments.

4.3 Ensuring quality

WHO aims at a quality mechanism to ensure that:

- controls are in place to verify that individual evaluations undertaken at the different levels of the Organization comply with (i) professional quality standards (OECD, 2010a; UNEG, 2012b) while meeting the information needs of their intended users, and (ii) WHO's evaluation policy;
- assurance is provided that the evaluation policy is implemented effectively and efficiently across the Organization.

4.3.1 Quality control of individual evaluations

Compliance with professional quality standards

The evaluation process, methods and management structure described in this handbook are designed to confirm that the content and proceedings of individual evaluations match the professional evaluation standards and the specific requirements spelt out in the terms of reference. This control is exercised at different levels by:

- the evaluation team leader, who is responsible for the quality and relevance of the evaluation report in terms of meeting the objectives of the terms of reference and must spell out the quality mechanism that will guide the evaluation as part of the workplan;
- the evaluation manager and, where applicable, the ad hoc evaluation management group who review and clear the terms of reference, the evaluation workplan, and the inception, draft and final reports.

Quality control is a continuous process that is carried on throughout the evaluation process. The evaluation manager and the ad hoc evaluation management group must ensure that UNEG standards are adhered to, bearing in mind that the exact nature of quality assurance arrangements depends on the scope and complexity of evaluations and should be decided when organization and management for a particular evaluation are established.

Quality control is achieved when the following conditions are met (Danida, 2012):

- The evaluation plan and the terms of reference are coherent, to ensure a clear logic between rationale, purpose, objectives and resources available for a planned evaluation. If external consultants are hired, tender procedures stipulate standards for quality assurance and clearly state that these are part of the requirements of the tenderer. The quality assurance set-up and approach of the tenderer are also rated as part of the technical proposal.
- The principles of independence and impartiality of the evaluation team are adhered to, from selection to completion.
- The inception report is coherent, and the approach and methodology meet professional quality standards.
- The fieldwork applies robust methodologies – i.e. it uses methods that best answer the evaluation questions in order to ensure validity and reliability of findings and conclusions.

- The evaluation report addresses all evaluation questions listed in the terms of reference; evaluation findings are drawn up on the basis of solid evidence and high-quality and consistent analysis; and there is a clear link between findings, conclusions and recommendations.
- Relevant stakeholders comment on the draft report and sign off/ approve final versions of the inception report, workplan, progress reports and the evaluation report.
- Peer reviewers' comments are taken into consideration in finalizing the report, where applicable.

The evaluation manager and the ad hoc evaluation management group should complete the “Checklist for evaluation terms of reference” (Annex 10) when they are cleared and the “Checklist for evaluation reports” (Annex 15) as references to validate individual evaluation exercises. The completed checklists should be forwarded to the GNE focal point.

Compliance with WHO evaluation policy

Evaluations must also comply with WHO evaluation policy. The evaluation management structure is responsible for ensuring that evaluations are carried out in accordance with the policy.

In order to achieve this, the GNE will perform a quality check to review the compliance of individual evaluations with WHO's evaluation policy (Annex 4) and adherence to relevant policies on gender, equity and human rights.

4.3.2 Quality assurance of WHO's evaluation function

The evaluation policy and the corporate evaluation function provide the overall quality assurance framework for evaluations within WHO.

The GNE will develop a proposal for the periodic review (e.g. every three years) of the implementation of the evaluation policy and of the wider quality assurance system on evaluation throughout WHO. This proposal will be in line with other accountability approaches in WHO. It will include peer reviews of the evaluation material and products, meta-evaluations, and training on specific aspects that should be used uniformly across WHO to ensure the validity of the evaluation products and of the evaluation function. The evaluation policy will be updated accordingly.

Ultimately, the Organization makes all evaluation products (e.g. evaluation reports and follow-up documents) publicly available via the WHO evaluation website, in accordance with WHO's evaluation policy. The transparency of this mechanism gives all stakeholders the opportunity to access relevant evaluation documentation and contributes to WHO's accountability.

Chapter 5. Reporting

This chapter provides details on the requirements for developing high-quality evaluation reports. It describes the peer-review process established by WHO.

5.1 Preparing the draft evaluation report

A written report is the principal output of the evaluation process. The draft evaluation report should be logically structured and should contain evidence-based findings, conclusions, lessons learned and recommendations. In accordance with UNEG quality criteria, evaluation reports should:

- be well structured and complete;
- describe what is being evaluated and why;
- identify the questions of concern to users;
- explain the steps and the procedures used to answer those questions;
- present findings supported by credible evidence in response to the questions;
- acknowledge limitations;
- draw conclusions and lessons learned about findings based on evidence;
- propose concrete and usable recommendations derived from conclusions and lessons learned;
- bear in mind how the evaluation will be used.

The report elements presented in Fig. 4 compose a standard structure and should be considered for all evaluations.

Fig. 4

Evaluation report structure

Executive summary

- The executive summary is an essential part of the report for most stakeholders. It should be short and should provide a brief overview of the main conclusions, recommendations and lessons learned from the evaluation - i.e. purpose, context and coverage of the evaluation, methods, main findings, lessons and recommendations.

Introduction or background

- The introduction presents the scope of the evaluation and gives a brief overview of the evaluated project, programme or subject - i.e. logic and assumptions, status of activities, objectives of the evaluation and questions to be addressed.

Fig 4 *continued*

Methods: phases in data collection (desk review, field visits, etc.)

- This section of the report gives reasons for selecting the point in the life of the project, programme or subject when the evaluation took place, and explains why countries or case-studies were chosen for detailed examination.
- It reports on how information is collected (use of questionnaires, official data, interviews, focus groups and workshops).
- It also presents limitations of the method and describes problems encountered - such as key people not available for interview or documents not available - or limitations of indicators in the project design.

Findings

- Findings report on the data (what happened and why, what actual results were achieved in relation to those intended, what positive or negative intended or unintended impacts happened, and what the effects were on target groups and others). All findings should be supported by evidence.

Conclusions

- The conclusions give the evaluation's concluding assessments of the project, programme or subject in light of evaluation criteria and standards of performance. The conclusions provide answers to the evaluation's objectives and key questions.

Lessons

- This section presents general lessons that have the potential for wider application and use. Lessons may also be drawn from problems and mistakes. The context in which the lessons may be applied should be clearly specified.

Recommendations

- The recommendations should suggest actionable proposals for stakeholders in order to rectify poor existing situations, and should include recommendations concerning projects, programmes or subjects of a similar nature. Prior to each recommendation, the issue(s) or problem(s) to be addressed by the recommendation should be clearly stated. A high-quality recommendation is an actionable proposal that is:
 - feasible to implement within the timeframe and resources available;
 - commensurate with the available capacities of project or programme team and partners;
 - specific in terms of who would do what and when;
 - contains results-based language (i.e. measurable performance targets);
 - includes a trade-off analysis, whereby the implementation of the recommendation may require utilization of significant resources that would otherwise be used for other purposes.

Fig 4 *continued*

Annexes

- The annexes should include the evaluation terms of reference, list of interviewees, documents reviewed, etc. Dissident views or management responses to the evaluation findings may be appended later.

Source: UNEG, 2010.

Annex 15 presents a **quality checklist** for the evaluation report. This quality checklist must be completed by the evaluation manager or the evaluation management group. Once validated by the evaluation commissioner, the checklist should be submitted together with the evaluation report to the evaluation registry. In the particular case of evaluations of humanitarian programmes, the Active Learning Network for Accountability and Performance in Humanitarian Action has developed a pro forma checklist that WHO recommends for assessing the quality of the report (ALNAP, 2006).

5.2 The final evaluation report

The draft report is the last opportunity to provide feedback to the evaluation team before the final report is published. The evaluation manager and the evaluation commissioner (and, as applicable, the ad hoc evaluation management group) should review the quality of the draft evaluation report – i.e. provide comments on factual inaccuracies and, if applicable, verify that the recommendations are feasible. Comments should be limited to issues regarding the applied methodology, factual errors or omissions, in order to safeguard the independence of the evaluation exercise.

The evaluation commissioner may call on the GNE to assess the technical rigour of the evaluation.

The GNE is designed as a platform facilitating discussions on evaluation matters among peers. It is therefore possible to discuss any difficulty encountered in the course of an evaluation with peers in the network, and to reflect on possible options.

A high-quality final report should:

- be addressed to the right stakeholders (according to the terms of reference and in agreement with the evaluation commissioner);
- address all issues raised in the terms of reference;
- be based on an assessment of needs and demand for the product among targeted users to ensure relevance, effectiveness, usefulness and value of the product;

- designed for a specific audience, taking into account functional needs and technical levels;
- relevant to decision-making needs;
- timely;
- written in clear and easily understandable language;
- based on the evaluation information without bias;
- based on data presented in a clear manner;
- developed through a participatory process and validated through a quality review process with relevant stakeholders, to the extent that this is compatible with the methodology outlined in the terms of reference and agreed with the evaluation commissioner;
- easily accessible to the target audience through the most effective and efficient means;
- consistent in the presentation of products to enhance visibility and learning.

The evaluation team leader is responsible for finalizing the draft report on the basis of the comments received from the evaluation manager, evaluation commissioner and the ad hoc evaluation management group or other relevant stakeholders, as applicable.

Chapter 6. Communication, utilization and follow-up of evaluation results

This chapter describes how to utilize and follow up on the results of an evaluation to maximize the returns of the evaluation process.

This chapter details the criteria for ensuring adequate dissemination of the evaluation reports, the best practice for sharing findings and lessons learned, and the benefits of debriefing the evaluation team. It also outlines the requirements of a management response and the follow-up process established by WHO. Finally, it describes how evaluation informs WHO's programmatic cycle.

6.1 Communication

6.1.1 Debriefing

A formal or informal debriefing of the evaluation team leader and relevant team members with the evaluation commissioner, the evaluation manager and the ad hoc evaluation management group offers the opportunity to ensure that important points not included in the report are captured. Nuanced findings that may not come out clearly in the report can also be discussed. This debriefing also provides an opportunity to discuss areas that were not significant enough to be included in the report but should have further attention in later evaluations.

Evaluation team members often identify issues that need further attention but are not included in the evaluation report. Such issues can be mentioned in a debriefing meeting and may be captured in an end of evaluation report document such as a closing memorandum.

6.1.2 Disseminating evaluation reports

It is usually the responsibility of the evaluation commissioner to distribute the report. Evaluation terms of reference normally specify expectations in terms of dissemination. However, findings during the evaluation process may require modifications to the dissemination plan or additions to the list of recipients of the report.

While the main and most important recipients are the individuals with the power to act on the findings (usually senior management), it is good practice to share the report with the persons involved in the evaluation process as feedback on their inputs.

Common dissemination methods include printed reports (for relevant meetings), electronic copies of the evaluation products, postings on WHO web sites and through e-mail messages and list serves, and CD-ROMs. All evaluation products will be available on the WHO evaluation web site. The media, when used appropriately, can be powerful partners in disseminating findings, recommendations and lessons learned from the evaluation.

6.1.3 **Sharing findings and lessons learned**

Learning and actively using the knowledge generated from the evaluation are among the most important elements of the evaluation exercise. Time and resources required for effective follow-up and learning should be allocated at the outset of the programme and project design. While technical programmes share the results of their evaluations through presentations at technical meetings and through publications, the main dissemination channels of evaluation findings, conclusions and recommendations are briefings, presentations, the GNE, the WHO evaluation web site and annual reports to governing bodies and WHO senior management.

The GNE plays an important role in sharing the findings and lessons learned from evaluations. The virtual meetings of the GNE dedicate specific time to this purpose.

The GNE will assist in updating the registry process and the mapping of evaluations in WHO. The registry will be updated regularly by IOS. The registry will be posted on the WHO evaluation web site.

The WHO evaluation web site will provide access to the evaluation reports issued throughout the Organization, as well as generic information on evaluation processes and methodologies, including this handbook. This will ensure that evaluation-related documents are subject to the scrutiny of all stakeholders.

Reports should also be shared with all relevant stakeholders, as identified by the evaluation commissioner. It is advised that the list of intended recipients of the evaluation report be included in the annexes to the evaluation terms of reference.

6.2 **Utilization and follow-up of evaluation results**

6.2.1 **Drafting a management response**

Evaluation plays a key role as (i) a source of evidence on the achievement of planned outcome and impact (results), as well as on project, programme and institutional performance, thus supporting programme improvement and accountability, and (ii) an agent of change that contributes to building knowledge and organizational learning.

The value of an evaluation, however, is heavily dependent on the use that is ultimately made of its recommendations, which is determined by:

- its relevance in terms of timing, to ensure that its findings are available to inform key decisions;
- its credibility, which derives from the independence, impartiality, clear methodology and quality of the report;

- the level of acceptance of its recommendations, directly linked to the involvement of internal and external stakeholders and to the quality of the recommendations, which must be implementable;
- the appropriateness of the management response, and the dissemination and use of evaluation findings to enhance organizational knowledge.

Recommendations contained in the evaluation report constitute the synthesis of the value added by the evaluation process. Each evaluation should have an identified owner such as a responsible officer of a cluster, programme, office or project. Normally the evaluation commissioner is the identified owner of the evaluation.

The identified owner should ensure that an appropriate management response is issued in a timely manner to the appropriate head of country office, regional director, head of department, assistant director-general or the Director-General, as appropriate. It is recommended that a deadline for submission of the management response to an evaluation be agreed. The process of developing a management response should engage all key evaluation stakeholders in reflection on the key issues, findings and recommendations. In this regard, establishing an inclusive ad hoc evaluation management group from the outset is valuable. During this process, follow-up actions and those who should carry them out are identified and agreed upon.

The preparation of a management response is not a one-time activity. It should document learning that results from the evaluation exercise, and should feed it into the design of new programmes and projects or the definition of future outcomes.

A management response is typically prepared in the form of a matrix requiring feedback on each recommendation (e.g. accepted, not accepted, partially accepted) and a list of actions. It is the responsibility of the owner of the evaluation to develop an action plan that specifies a timeline for the implementation of the recommendations. For more details on respective roles and responsibilities in the drafting of management responses, see Annex 5.

The GNE can provide support by showing examples of a good management response and clarifying doubts in case the concerned managers lack experience in preparing such a response. The responsibility for the substance of a management response lies with the office concerned. However, the GNE will check the quality of the management response to ensure that the recommendations have been responded to and have a chance of being implemented.

6.2.2 Informing WHO's programme cycle

One of the main purposes of institutionalizing a follow-up process to evaluations is to influence the planning and implementation of strategies, programmes

and projects. Evaluation commissioners at all levels of the Organization should therefore consider the role that an evaluation will play in providing essential insights for subsequent phases of an intervention or policy by ensuring the following:

- The content of the planned evaluation addresses critical issues for the future planning of the intervention, policy or strategy at stake and informs subsequent phases or new interventions.
- The timing of the evaluation is adequate for providing a final report that can be considered in designing future interventions or policies.
- The methodologies applied are adequate for providing the right data to inform future planning.
- The right actors are involved to ensure their commitment to future interventions.
- The conclusions and recommendations contained in the final report provide realistic options for future developments.
- Follow-up reporting on evaluation recommendations takes place at intervals that allow alignment with the Organization's planning process.
- The implementation and follow-up processes clearly indicate how and when actions have been taken on the results of the evaluation to inform the programming cycle of the entity that was evaluated.

It is the responsibility of programme directors, under the guidance of PRP, to ensure that outputs/outcomes from the project and programme as defined in the operational plans are evaluable – i.e. they are based on an adequate SMART (specific, measurable, achievable, realistic and time-bound) set of objectives, performance indicators and related baselines, targets and timelines that can be used to measure progress towards an organizational objective.

The use of a logical framework provides a systematic planning procedure for project cycle management which includes the performance framework of planned activities with indicators, outputs, outcomes and impacts. The framework should highlight the project success criteria and list the major underlying assumptions and risks.³ The logical framework approach is problem-solving and takes into account the views of all stakeholders. Ensuring that WHO interventions address the issues raised by the logical framework matrix, or a similar approach, will help support their evaluability.

³ Risk is an uncertain event or set of events which, if they occur, will have an effect on the achievement of an organizational objective. Risks are considered in light of the probability of a threat or opportunity occurring and of the potential impact.

The knowledge generated by evaluations at WHO provides input into biennial operational planning, the programme budget process and the strategic planning of the General Programme of Work. The GNE plays a critical role in disseminating evaluation results across the Organization and ensuring that they also inform the programme cycle of individual programmes/projects at headquarters, regional and country levels. To this end, the GNE liaises on a regular basis with WHO's planning and country support networks to ensure that individual independent evaluations complement the performance assessment cycle and that evaluations are embedded in the planning and performance assessment as an integral part of the programme budget process.

6.2.3 Following up

Evaluation commissioners are ultimately responsible for the implementation of the evaluation recommendations. The GNE monitors the follow-up of the implementation of evaluation recommendations in a systematic manner. To facilitate the process, the members of the GNE are available to discuss and help coordinate the preparation of the management response.

The management response constitutes the baseline for monitoring accepted recommendations and agreed actions, which in turn informs follow-up reports on the status of the implementation.

An electronic tool is envisaged to monitor the timely implementation of recommendations. IOS will issue, through the GNE, periodic status reports on progress in the implementation of recommendations to senior management and will also report annually to the Executive Board.

References

- Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) (2006). *Evaluating humanitarian action using the OECD-DAC criteria*. London. Overseas Development Institute.
- Bamberger M, Rugh J, Mabry L (2006). *Strengthening the evaluation design and the validity of the conclusions*. Thousand Oaks, CA, Sage Publications.
- CDC (1999). *A framework for programme evaluation*. Atlanta, GA, Centers for Disease Control and Prevention (<http://www.cdc.gov/eval/framework/index.htm>, accessed 18 July 2013).
- CTC (2013). *How does theory of change work?* New York, NY, ActKnowledge/Center for Theory of Change (<http://www.theoryofchange.org/what-is-theory-of-change/how-does-theory-of-change-work/>, accessed 18 September 2013).
- Danida (2012). *Danida evaluation guidelines*. Copenhagen, Ministry of Foreign Affairs of Denmark (<http://www.netpublikationer.dk/um/11121/index.htm>, accessed 14 August 2013).
- European Commission (2012). *EC evalsed: the resource for the evaluation of socio-economic development*. Brussels, European Commission/General Directorate for Regional Policy (http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide2012_evalsed.docm, accessed 16 September 2013).
- Leeuw F, Vaessen J (2009). *Impact evaluations and development. NONIE guidance on impact evaluation*. Washington, DC, Network of Networks for Impact Evaluation (http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf, accessed 14 August 2013).
- OECD (1998). *Best practice guidelines for evaluation* (PUMA Policy Brief No. 5). Paris, Organisation for Economic Co-operation and Development (<http://www.oecd.org/governance/budgeting/1902965.pdf>, accessed 13 August 2013).
- OECD (2010a). *DAC quality standards for development evaluation*. Paris, Organisation for Economic Co-operation and Development (DAC Guidelines and Reference Series) (<http://www.oecd.org/dac/evaluation/qualitystandardsfordevelopmentevaluation.htm>, accessed 14 August 2013).
- OECD (2010b). *Glossary of key terms in evaluation and results based management*. Paris, Organisation for Economic Co-operation and Development (<http://www.oecd.org/development/peer-reviews/2754804.pdf>, accessed 13 September 2013).
- Patton MQ (2011). *The debate about randomized controls in evaluation: the gold standard question*. Copenhagen, Ministry of Foreign Affairs of Denmark (http://um.dk/en/~/_media/UM/Danish-site/Documents/Danida/Resultater/Eval/Patton_RCT_April_2011.pdf.jpg, accessed 13 September 2013).
- Trochim WMK (2006). Introduction to evaluation. In: *Research methods knowledge base*. New York, NY, Web Center for Social Research Methods (<http://www.socialresearchmethods.net/kb/intreval.php>, accessed 14 August 2013).
- UNDP (2009). *Handbook on planning, monitoring and evaluating for development results*. New York, NY, United Nations Development Group (<http://web.undp.org/evaluation/handbook/>, accessed 13 August 2013).
- UNEG (2008a). *UNEG code of conduct for evaluation in the UN system*. New York, NY, United Nations Evaluation Group (http://www.uneval.org/documentdownload?doc_id=100&file_id=547, accessed 13 August 2013).
- UNEG (2008b). *Core competencies for evaluators of the UN system*. New York, NY, United Nations Evaluation Group (<http://www.uneval.org/search/index.jsp?q=evaluators>, accessed 14 August 2013).
- UNEG (2010). *UNEG quality checklist for evaluation reports*. New York, NY, United Nations Evaluation Group (http://www.uneval.org/documentdownload?doc_id=1409&file_id=1851, accessed 28 February 2013).

- UNEG (2011). *Integrating human rights and gender equality in evaluation – towards UNEG guidance*. New York, NY, United Nations Evaluation Group (http://www.unevaluation.org/HRGE_Guidance, accessed 13 August 2013).
- UNEG (2012a). *Impact evaluation of UN normative work, UNEG Task Force on Impact Evaluation (IETF)*. New York, NY, United Nations Evaluation Group.
- UNEG (2012b). *Norms for evaluation in the UN system*. New York, NY, United Nations Evaluation Group (http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=21, accessed 24 September 2013).
- UNEG (2013). *The role of impact evaluation in UN agency evaluation systems, UNEG Task Force on Impact Evaluation (IETF) (UNEG Guidance Note)*. New York, NY, United Nations Evaluation Group (<http://www.uneval.org/papersandpubs/documentall.jsp>, accessed 4 September 2013).
- UNICEF (1991). *A UNICEF guide for monitoring and evaluation: making a difference?* New York, NY, United Nations Children's Fund, (<http://preval.org/documentos/00473.pdf>, accessed 17 September 2013).
- UNICEF (2011). *How to design and manage equity-focused evaluations*. New York, NY, United Nations Children's Fund (http://mymande.org/sites/default/files/EWP5_Equity_focused_evaluations.pdf, accessed 17 September 2013).
- Van den Berg RD, Todd D (2011). The full road to impact: the experience of the Global Environment Facility Fourth Overall Performance Study. *Journal of Development Effectiveness*, 3:389–413.
- WHO (2005a). *Constitution of the World Health Organization*. Geneva, World Health Organization, 2005 (<http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf>, accessed 14 August 2013).
- WHO (2005b). *Fraud prevention policy and fraud awareness guidelines*. Geneva, World Health Organization (<http://intranet.who.int/homes/fnm/documents/fraudprevention.pdf>, accessed 22 August 2013).
- WHO (2007). Resolution WHA60.25. Strategy for integrating gender analysis and actions into the work of WHO. In: *World Health Assembly, First Special Session, Geneva, 9 November 2006, resolutions and decisions, annex. Sixtieth World Health Assembly, Geneva, 14–23 May 2007, resolutions and decisions, annexes*. Geneva, World Health Organization (WHASS1/2006–WHA60/2007/REC/1) (http://apps.who.int/gb/ebwha/pdf_files/WHASSA_WHA60-Rec1/E/reso-60-en.pdf, accessed 17 September 2013).
- WHO (2009a). *Ethical principles and conduct of staff. Compilation of WHO policies and practices*. Geneva, World Health Organization ([http://emanual.who.int/eM_RelCont_Lib/Ethical%20principles%20and%20conduct%20of%20staff\[1\].pdf](http://emanual.who.int/eM_RelCont_Lib/Ethical%20principles%20and%20conduct%20of%20staff[1].pdf), accessed 28 February 2013).
- WHO (2009b). *Strategy for integrating gender analysis and actions into the work of WHO*. Geneva, World Health Organization (<http://www.who.int/gender/documents/gender/9789241597708/en/index.html>, accessed 2 August 2013).
- WHO (2011a). *Gender mainstreaming for health managers: a practical approach*. Geneva, World Health Organization (http://whqlibdoc.who.int/publications/2011/9789241501064_eng.pdf, accessed 17 September 2013).
- WHO (2011b). *Guidelines for Declaration of Interests (WHO Experts)*. Geneva, World Health Organization, (http://intranet.who.int/homes/kms/documents/coi_guidelines_and_procedure_final.doc, accessed 17 September 2013).
- WHO (2012). *Procurement of services: revision of threshold for mandatory competitive bidding (Information Note 22/2012)*. Geneva, World Health Organization (<http://intranet.who.int/admin/infonotes/2012en.shtml>, accessed 17 September 2013).

Bibliography

Alkin MC, Ruskus JA. *Reflections on evaluation costs*. Los Angeles, CA, University of California Center for the Study of Evaluation, 1984.

Bamberger M, Clark M, Sartorius R. *Monitoring and evaluation for results: some tools, methods and approaches*. Washington, DC, World Bank. 2004, (<http://documents.worldbank.org/curated/en/2004/01/11528617/monitoring-evaluation-some-tools-methods-approaches>, accessed 16 September 2013).

Bamberger M, Segone M. *How to design and manage equity-focused evaluations*. New York, NY, United Nations Children's Fund, 2011 (<http://www.mymande.org/content/how-design-and-manage-equity-focused-evaluations>, accessed 12 September 2013).

Bridging the gap: the role of monitoring and evaluation in evidence-based policy making. New York, NY, United Nations Children's Fund, 2004.

CIDA evaluation guide. Ottawa, Canadian International Development Agency, 2004 ([http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview5/\\$file/english-e-guide.pdf](http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview5/$file/english-e-guide.pdf), accessed 10 September 2013).

Clinton T, Nunes-Neto B, Williams E. *Congress and program evaluation: an overview of randomized control trials (RCTs) and related issues*. Washington, DC, Congressional Research Service, Library of Congress, 2006.

Conducting quality impact evaluations under budget, time and data constraints. Washington, DC, World Bank, 2006 (http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/4585672-1251461875432/conduct_qual_impact.pdf, accessed 12 September 2013).

Consulting services manual 2006: a comprehensive guide to the selection of consultants. Washington, DC, World Bank, 2006 (<http://siteresources.worldbank.org/INTPROCUREMENT/Resources/2006ConsultantManual.pdf>, accessed 12 September 2013).

Evaluation guidelines for foreign assistance. Washington, DC, Office of the Director of the United States Foreign Assistance, 2009 (http://betterevaluation.org/resources/guide/usaid_foreign-assistance_guidelines, accessed 12 September 2013).

Evaluation manual: methodology and processes. Rome, International Fund for Agricultural Development Office of Evaluation, April 2009 (http://www.ifad.org/evaluation/process_methodology/doc/manual.pdf, accessed 2 August 2013).

Guidance for managing joint evaluations. Paris, Organisation for Economic Co-operation and Development, 2006 (DAC Evaluation Series) (<http://www.oecd.org/development/evaluation/37512030.pdf>, accessed 12 September 2013).

Guidance on evaluation and review for DFID staff. London, United Kingdom Department for International Development, 2005 (<http://webarchive.nationalarchives.gov.uk/+http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf>, accessed 10 September 2013).

Guidelines to avoid conflict of interest in independent evaluations. Manila, Asian Development Bank, 2012 (<http://www.adb.org/documents/guidelines-avoid-conflict-interest-independent-evaluations>, accessed 10 September 2013).

Hanberger A, Gisselberg K. *Sida's management response system*. Stockholm, Swedish International Development Cooperation Agency, 2006 (SIDA studies in evaluation 06/01) (<http://www.oecd.org/derec/sweden/37293078.pdf>, accessed 10 September 2013).

Handbook on planning, monitoring and evaluating for development results. New York, NY, United Nations Development Group. 2009. (<http://web.undp.org/evaluation/handbook/>, accessed 13 August 2013).



How to perform evaluations – evaluation workplans. Gatineau, Canadian International Development Agency, 2012 ([http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview3/\\$file/Eval_Workplans.pdf](http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview3/$file/Eval_Workplans.pdf), accessed 10 September 2013).

Khandker SR, Koolwal GB, Samad HA. *Handbook on impact evaluation: quantitative methods and practices.* Washington, DC, World Bank, 2010 (http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101Official0Use0Only1.pdf, accessed 12 September 2013).

Impact evaluation: methodological and operational issues. Manila, Asian Development Bank, 2006 (<http://www.adb.org/documents/impact-evaluation-methodological-and-operational-issues>, accessed 10 September 2013).

Improving evaluation practices: best practice guidelines for evaluation and background paper. Paris, Organisation for Economic Co-operation and Development, 1999 (PUMA/PAC(99)1) (http://ec.europa.eu/dgs/information_society/evaluation/data/pdf/lib_master/oecd_01e91637_improving_evaluation_practices.pdf, accessed 11 September 2013).

Inspection and evaluation manual: guidelines for the conduct of inspections and evaluations in the United Nations Office of Internal Oversight Services. New York, NY, United Nations Inspection and Evaluation Division, 2009 (http://www.un.org/Depts/oios/ied/ied_manual_v1_6.pdf, accessed 12 September 2013).

Leeuw F, Vaessen J. *Impact evaluations and development. NONIE guidance on impact evaluation.* Washington, DC, Network of Networks on Impact Evaluation, 2009 (http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf, accessed 10 September 2013).

Managing for results: a guide to using evaluation in the United Nations Secretariat. New York, NY, United Nations, 2005 (http://www.un.org/Depts/oios/pages/manage_results.pdf, accessed 12 September 2013).

Monitoring and evaluation plan: guidance for submission of an M&E plan for Global Fund grants. Geneva, The Global Fund to Fight AIDS, Tuberculosis and Malaria, 2010 (<http://www.theglobalfund.org/en/me/documents/planguidelines/>, accessed 10 September 2013).

Montague S, Young G., Montague C. *Using circles to tell the performance story.* Ottawa, Canadian Government Executive, 2003 (<http://www.pmn.net/wp-content/uploads/Using-Circles-to-Tell-the-Performance-Story.pdf>, accessed 19 September 2013).

National AIDS councils: monitoring and evaluation operations manual. Geneva, Joint United Nations Programme on HIV/AIDS, 2002 (UNAIDS/02.47E) (http://www.unaids.org/en/media/unaids/contentassets/dataimport/publications/irc-pub02/jc808-moneval_en.pdf, accessed 12 September 2013).

OECD DAC Network on Development Evaluation. *Evaluating development co-operation: summary of key norms and standards*, 2nd ed. Paris, Organisation for Economic Co-operation and Development, 2010 (<http://www.oecd.org/development/evaluation/dcdndep/41612905.pdf>, accessed 12 September 2013).

Performance monitoring and evaluation tips – conducting key informant interviews. Washington, DC, United States Agency for International Development Center for Development Information and Evaluation, 1996 (http://pdf.usaid.gov/pdf_docs/PNAB5541.pdf, accessed 19 September 2013).

Project evaluation. In: *Technical cooperation manual.* Geneva, International Labour Organization, 2012 (http://www.ilo.org/pardev/development-cooperation/evaluation/WCMS_172679/lang--en/index.htm, accessed 10 September 2013).

Quality checklist for evaluation terms of reference and inception reports. New York, NY, United Nations Evaluation Group, 2012 (<http://www.uneval.org/search/index.jsp?q=quality+checklist>, accessed 12 September 2013).

Ravallion M. *The mystery of the vanishing benefits: Ms speedy analyst's introduction to evaluation.* Washington, DC, World Bank (Working Paper No. 2153), 1999.

Results-oriented monitoring and evaluation: a handbook for programme managers. New York, NY, United Nations Development Programme Office of Evaluation and Strategic Planning, 1997 (OESP Handbook Series) (<http://web.undp.org/evaluation/documents/mae-toc.htm>, accessed 12 September 2013).

Sanders JR. *Program evaluation standards: how to assess evaluations of educational programs*, 2nd edition. Joint Committee on Standards for Educational Evaluation. Thousand Oaks, CA, Sage Publications. 1994.

The program manager's guide to evaluation, 2nd ed. Washington, DC, United States Department of Health and Human Services, Office of Planning, Research and Evaluation, 2010 (http://www.acf.hhs.gov/sites/default/files/opre/program_managers_guide_to_eval2010.pdf, accessed 12 September 2013).

The role of evaluation in results-based management. New York, NY, United Nations Evaluation Group, 2007, updated 2012 (http://www.unevaluation.org/papersandpubs/documentdetail.jsp?doc_id=87, accessed 12 September 2013).

Toolkits: a practical guide to planning, monitoring, evaluation and impact assessment, 2nd ed. London, Save the Children UK, 2003.

UNEP evaluation manual. Nairobi, United Nations Environment Programme, 2008 (<http://www.unep.org/eou/StandardsPolicyandPractices/UNEPEvaluationManual/tabid/2314/Default.aspx>, accessed 19 September 2013).

UNICEF evaluation report standards. New York, NY, United Nations Children's Fund, 2010 (http://www.unicef.org/evaluation/files/UNEG_UNICEF_Eval_Report_Standards.pdf accessed 12 September 2013).

WFP's evaluation policy. In: *World Food Programme, Executive Board, Second Regular Session, Rome, 27–30 October 2008.* Rome, World Food Programme, 2008 (<http://one.wfp.org/eb/docs/2008/wfp187763~2.pdf>, accessed 19 September 2013).

Wimbush E, Montague S, Mulherin T. The Applications of Contribution Analysis: Strengthening Outcomes Thinking, Practice & Collaborative Capacity. *Evaluation*, 2012, 18(3) 310–329.

W. K. Kellogg Foundation evaluation handbook: philosophy and expectations. Battle Creek, MI, W.K. Kellogg Foundation, 1998 (www.epa.gov/evaluate/pdf/eval-guides/evaluation-handbook.pdf, accessed 19 September 2013).

Writing a good executive summary. New York, NY, United Nations Children's Fund, 2002.

Zukoski A, Luluquisen M. Participatory evaluation. What is it? Why do it? What are the challenges? *Community-based Public Health Policy and Practice*, 2002, No. 5 (http://depts.washington.edu/ccph/pdf_files/Evaluation.pdf, accessed 10 September 2013).

Annex 1

WHO Evaluation policy¹

I. Purpose

1. The purpose of this policy is to define the overall framework for evaluation at WHO, to foster the 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.
2. The accountability framework of WHO includes several types of assessments. WHO considers that all are crucial to programme development and institutional learning. The current policy addresses only the assessments qualifying as “Evaluation” and excludes other forms of assessments conducted in WHO, such as monitoring, performance assessment, surveys, and audit.

II. Policy statement

3. Evaluation is an essential function at WHO, carried out at all levels of the Organization. It ensures accountability and oversight for performance and results, and reinforces organizational learning in order to inform policy for decision-makers and support individual learning.

III. Evaluation definition

4. “An evaluation is an assessment, as systematic and impartial as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area institutional performance (...).”²
 - (a) It focuses on expected and achieved accomplishments, examining the results chain, processes, contextual factors and causality, in order to understand achievements or the lack thereof.
 - (b) It aims at determining the relevance, impact, effectiveness, efficiency, and sustainability of the interventions and contributions of the Organization.

¹ Reproduced from *Evaluation policy*. Geneva, World Health Organization, 2012 (Information Note 28/2012).

² As defined in the *Norms for evaluation in the UN system*. Geneva, United Nations Evaluation Group, 2005 (UNEG/FN/Norms (2005)).

- (c) It provides evidence-based information that is credible, reliable and useful, enabling the timely incorporation of findings, recommendations and lessons learnt into the decision-making processes of the Organization.
- (d) It is an integral part of each stage of the programming cycle and not only an end-of-programme activity.

IV. Principles and norms³

- 5. This policy provides a framework for the evaluation function and evaluation processes to ensure the systematic application of the key principles for evaluation in WHO. The key principles set out below are interrelated and underpin the approach to evaluation in WHO.

A. Impartiality

- 6. Impartiality is the absence of bias in due process; it requires methodological rigour, and the objective consideration and presentation of achievements and challenges. Impartiality contributes to the credibility of evaluation and reduces bias in the data gathering, analysis, formulation of findings, conclusions and recommendations.
- 7. All evaluations shall be conducted in an impartial manner at all stages of the evaluation process. An evaluation management group will be established for each evaluation to ensure oversight of the evaluation process.

B. Independence

- 8. Independence is the freedom from the control, or undue influence, of others. Independence provides legitimacy to evaluation and reduces the potential for conflicts of interest that could arise if policy-makers and managers were solely responsible for the evaluation of their own activities.
- 9. Independence must be ensured at organizational, functional, and behavioural levels. At the **organizational level**, the evaluation function must be separated from those responsible for the design and implementation of the programmes and operations being evaluated. At the **functional level**, there must be mechanisms that ensure independence in the planning,

³ See *Norms for evaluation in the UN system* (Geneva, United Nations Evaluation Group, 2005 (UNEG/FN/ Norms (2005)), and *DAC principles for evaluation of development assistance* (Paris, Organisation for Economic Co-operation and Development Development Assistance Committee, 1991, reprinted 2008 (OCDE/GD(91)208)).

funding, and reporting of evaluations. At the **behavioural level**, there must be a code of conduct that is ethics-based. This code of conduct will seek to prevent or appropriately manage conflicts of interest.

10. Evaluators shall not be directly responsible for the policy, design, or overall management of the subject under review. WHO staff performing evaluations shall abide by the ethical principles and conduct of staff.⁴ External contractors shall abide by the WHO requirements for external contractual agreements. Evaluators must maintain the highest standards of professional and personal integrity during the entire evaluation process. They are expected to ensure that evaluations address gender and equity; and be sensitive to contextual factors, such as the beliefs, manners and customs of the social and cultural environments evaluated.
11. The whistleblower policy and other relevant policies will protect staff participating in evaluations from retaliation or repercussions.

C. Utility

12. Utility relates to the impact of the evaluation on decision-making and requires that evaluation findings be relevant and useful, presented in a clear and concise way, and monitored for implementation. The utility of an evaluation depends on its timeliness, relevance to the needs of the programme and stakeholders, the credibility of the process and products, and the accessibility of reports.
13. Utility will be ensured through the systematic prioritizing of the evaluation agenda based on established criteria and consultation with relevant stakeholders; the systematic follow-up of recommendations; public access to the evaluation products; and alignment with the results-based management framework.

D. Quality

14. Quality relates to the appropriate and accurate use of evaluation criteria, impartial presentation and analysis of evidence, and coherence between findings, conclusions and recommendations.
15. Quality will be assured through (a) the continuous adherence to WHO evaluation methodology, applicable guidelines and the norms and standards for evaluation of the United Nations Evaluation Group; (b) oversight by the

⁴ WHO Code of Ethics.

evaluation management group; and (c) peer-review of the evaluation report when justified. Other mechanisms such as periodic meta-evaluations will also be considered.

E. Transparency

16. To achieve transparency, stakeholders should be aware of the reason for the evaluation, the selection criteria, and the purposes for which the findings will be used. Transparency of process is also important, as is the accessibility of evaluation materials and products.
17. Transparency will be ensured through the approaches described below. The commissioner of the evaluation will ensure a continuous consultation process with relevant stakeholders at all stages of the evaluation process. The evaluation report shall contain details of evaluation methodologies, approaches, sources of information and costs incurred. In accordance with the WHO disclosure policy, evaluation plans, reports, management responses and follow-up reports will be made public on the WHO evaluation web site.

V. Types of evaluation

18. The WHO Secretariat commissions the following main types of evaluation:
 - (a) **Thematic evaluations** focus on selected topics, such as a new way of working, a strategy, a cross-cutting theme or core function; or they address an emerging issue of corporate institutional interest. Thematic evaluations provide insight into relevance, effectiveness, sustainability and broader applicability. They require an in-depth analysis of a topic and cut across organizational structures. The scope of these evaluations may range from the entire Organization to a single WHO office.
 - (b) **Programmatic evaluations** focus on a specific programme. This type of evaluation provides an in-depth understanding of how and why results and outcomes have been achieved over several years and examines their relevance, effectiveness, sustainability, and efficiency. Programmatic evaluations address achievements in relation to WHO's results chain, and require a systematic analysis of the programme under review. The scope of programmatic evaluations may range from a country to interregional or global levels.
 - (c) **Office-specific evaluations** focus on the work of the Organization in a country, region or at headquarters in respect of WHO's objectives and commitments.

19. The Executive Board may, at its discretion, also commission an evaluation of any aspects of WHO.

VI. External evaluations

20. Evaluations may be commissioned by the governing bodies to be conducted by external evaluators independent of the Secretariat. Other stakeholders, such as Member States, donors or partners, may also commission external evaluations of the work of WHO for the purpose of assessing performance and accountability or prior to placing reliance on the work of the Organization.
21. The Secretariat will fully cooperate in external evaluations through a process of disclosure of appropriate information and facilitation of their performance. The results of external evaluations, when made available, will be disclosed on the WHO evaluation web site.

VII. Planning and prioritization of evaluations

22. WHO will develop a biennial, Organization-wide evaluation workplan as part of the Organization's planning and budgeting cycle.
23. The workplan shall be established in consultation with senior management at headquarters and regions and with Heads of WHO Offices in countries, areas and territories, based on established criteria. The biennial workplan will be updated annually on the basis of the annual report to the Programme Budget and Administration Committee and the Executive Board. The workplan shall be submitted to the Executive Board for approval through the Programme Budget and Administration Committee.
24. The following categories shall be considered in the development of criteria⁵ for the selection of topics for evaluation:
 - 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 achievement of expected results.

⁵ Refer to the main text for further guidance on detailed selection criteria.

- 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.

VIII. Evaluation methodology

25. The following are the main components of an evaluation process.⁶

A. Design

26. Terms of reference for an evaluation shall include detailed information on the following elements:

- (a) context of the evaluation;
- (b) purpose and objectives of the evaluation;
- (c) scope and linkage to the Programme Budget and the General Programme of Work (outlining what is, and what is not, covered by the evaluation);
- (d) evaluation criteria (inter alia, relevance, impact, efficiency, effectiveness and sustainability) and key evaluation questions;
- (e) users (owner and audience) of the evaluation results;
- (f) methodology (approach for data collection and analysis, and involvement of stakeholders);
- (g) evaluation team (size, knowledge, skills and qualifications);
- (h) a detailed workplan (including a timetable, organization and budget);
- (i) deliverables (including report, distribution strategy and follow-up);
- (j) ad hoc evaluation management group (including technical staff requirements).

B. Ad hoc evaluation management group

27. When warranted by the size and complexity of the evaluation, an ad hoc evaluation management group shall be assembled by the evaluation commissioner to assist in the conduct and quality control of the evaluation. The group may comprise external experts and/or WHO staff. The functions of this ad hoc group include reviewing, and commenting on, the terms of reference and the draft report. The group shall be kept informed of progress and should be available to respond to queries from the evaluation team and provide suggestions for consideration.

⁶ Refer to the main text for further guidance on evaluation.

C. Team selection

28. The following should be considered in the selection of the evaluation team members:
 - (a) technical and sectoral expertise;
 - (b) in-depth understanding and experience of quantitative and qualitative evaluation methodology;
 - (c) previous experience of conducting reviews and evaluations.
29. The team selection process must ensure that no member of the evaluation team has a conflict of interest.
30. The evaluation team leader shall be responsible for interactions among the evaluation team members and have overall responsibility for the evaluation outputs.

D. Report

31. A written report is an essential requirement of the evaluation process. The final evaluation report shall be logically structured and contain evidence-based findings, conclusions, lessons learnt and recommendations.
32. The report must:
 - (a) include only information relevant to the overall purpose and objectives of the evaluation;
 - (b) describe the purpose of the evaluation and attach the terms of reference;
 - (c) answer the key questions detailed in the terms of reference;
 - (d) describe the methodology used to collect and analyse the information;
 - (e) indicate any limitations of the evaluation; and
 - (f) include the evidence on which the conclusions, lessons learnt, and recommendations are based.

IX. Financing of evaluation

33. The Director-General shall ensure that there are adequate resources to implement the Organization-wide evaluation workplan.
34. Regional Directors, Assistant Directors-General, Directors and Heads of WHO Country Offices must ensure that resources are adequate to implement their respective components of the Organization-wide evaluation workplan. An appropriate evaluation budget must be an integral

part of the operational workplan of a programme, and shall be discussed as necessary with stakeholders during the planning phase of each project/programme/initiative.

35. In determining the amount required to finance evaluation in WHO, estimations provided by other organizations have been considered. According to these, the overall programme budget might contain, as an integral part, a figure for evaluation that is equivalent to between 3% and 5% of that budget.

X. Accountability and oversight

36. The accountability framework defines from whom, and to whom, authority flows and for what purpose. It further defines the accountability of those with authority and their responsibility in exercising that authority. This section defines the roles and responsibilities⁷ for the main actors in the evaluation process as well as the monitoring mechanism used to implement the evaluation policy.

A. Roles and responsibilities

37. **The Executive Board of WHO**⁸ shall:
 - (a) determine the evaluation policy and subsequent amendments, as needed;
 - (b) provide oversight of the evaluation function within the Organization;
 - (c) encourage the performance of evaluations as an input to planning and decision-making;
 - (d) provide input to the biennial Organization-wide evaluation workplan on the items of specific interest to Member States;
 - (e) approve the biennial Organization-wide evaluation workplan;
 - (f) consider and take note of the annual report of the implementation of the biennial Organization-wide evaluation workplan;
 - (g) periodically revise the evaluation policy, as necessary.
38. **The Office of Internal Oversight Services** is the custodian of the evaluation function. IOS reports directly to the Director-General, and annually in a report for consideration by the Executive Board, on matters relating to evaluation at WHO. IOS is responsible for the following functions related to evaluation:

⁷ Refer to the main text for further details on the individual roles and responsibilities for evaluation.

⁸ WHO Executive Board and its subsidiary organ, the Programme, Budget and Administration Committee.

- (a) leading the development of a biennial Organization-wide evaluation workplan;
- (b) informing senior management on evaluation-related issues of Organization-wide importance;
- (c) facilitating the input of evaluation findings and lessons learnt for programme planning;
- (d) coordinating the implementation of the framework for evaluation across the three levels of the Organization;
- (e) maintaining a system to track management responses to evaluations;
- (f) maintaining an online inventory of evaluations performed across WHO;
- (g) maintaining a roster of experts with evaluation experience;
- (h) providing guidance material and advice for the preparation, conduct and follow-up of evaluations;
- (i) reviewing evaluation reports for compliance with the requirements of the policy;
- (j) strengthening capacities in evaluation among WHO staff (for example, making available standardized methodologies or training on evaluation);
- (k) submitting an annual report on evaluation activities to the Executive Board through the Director-General;
- (l) supporting the periodic review and updates to the policy as needed.

XI. Use of evaluation findings

A. Utilization and follow-up of recommendations

39. Recommendations contained in evaluation reports reflect the value added by the evaluation process. Each evaluation shall have an identified owner, such as the responsible officer of a cluster, programme, office or project. It is the responsibility of the owner to utilize the findings of the evaluation and develop an action plan and timeline for the implementation of the recommendations.
40. The evaluation owner shall ensure that an appropriate management response is issued in a timely manner to the appropriate assistant director-general at headquarters, or to the regional director in the regions and countries.
41. The Office of Internal Oversight Services shall monitor the follow-up of the implementation of evaluation recommendations in a systematic manner, coordinating efforts with the evaluation owners. IOS shall issue periodic

status reports on progress in the implementation of the recommendations to senior management and report annually to the Executive Board through the Programme, Budget and Administration Committee.

B. Disclosure and dissemination of evaluation reports

42. WHO shall make evaluation reports available in accordance with the Organization's disclosure policy.
43. Lessons learnt from evaluations shall be distilled, reported and disseminated as appropriate.

Annex 2

Typology of assessments other than evaluation conducted at WHO

Monitoring

Monitoring is a continuous management function that provides regular information on progress, or lack thereof, in the achievement of intended results. It is carried out in two different forms.

- (a) Performance assessment under the Results Based Management Framework. This refers only to programme monitoring within the Results-Based Management Framework and includes the mid-term review (MTR) and the end-of-biennium (EOB) performance assessment reports that all WHO programmes must complete as part of their work.
- (b) Routine assessment work of programme activities. This category includes the routine collection and analysis of data that units or programmes undertake with regard to their own activities and country programme progress, as well as the assessments conducted for specific donor reporting purposes, in addition to the routine performance assessment. This assessment work is performed internally and includes a form of time-bound annual reporting completed by countries on achievements during the year. Units or programmes use these analyses to assess performance and to reorient or guide their future activities. Special cases within this subcategory are the annual reports that technical programmes produce. These annual reports may include extensive analysis of activities or of programme progress. Many programmes consider these annual reports as multipurpose, serving as tools for both advocacy and resource mobilization, rather than as purely programmatic assessments.

Global surveys

Global surveys include ad hoc exercises completed by technical units or programmes, less frequently than on an annual basis, to collect information from countries to inform and improve the global programmes. Technical programmes use these global surveys as part of their programme development process and as internal and external advocacy tools.

Ad hoc consultations

Ad hoc consultations include a broad range of mechanisms through which technical programmes build evidence for their policies and strategies, and obtain feedback on performance. Examples of such mechanisms include meetings of expert committees (including technical advisory groups), informal technical consultations on technical or managerial issues, and the Strategic Advisory Group of Experts (SAGE) on Immunization.

Programme reviews

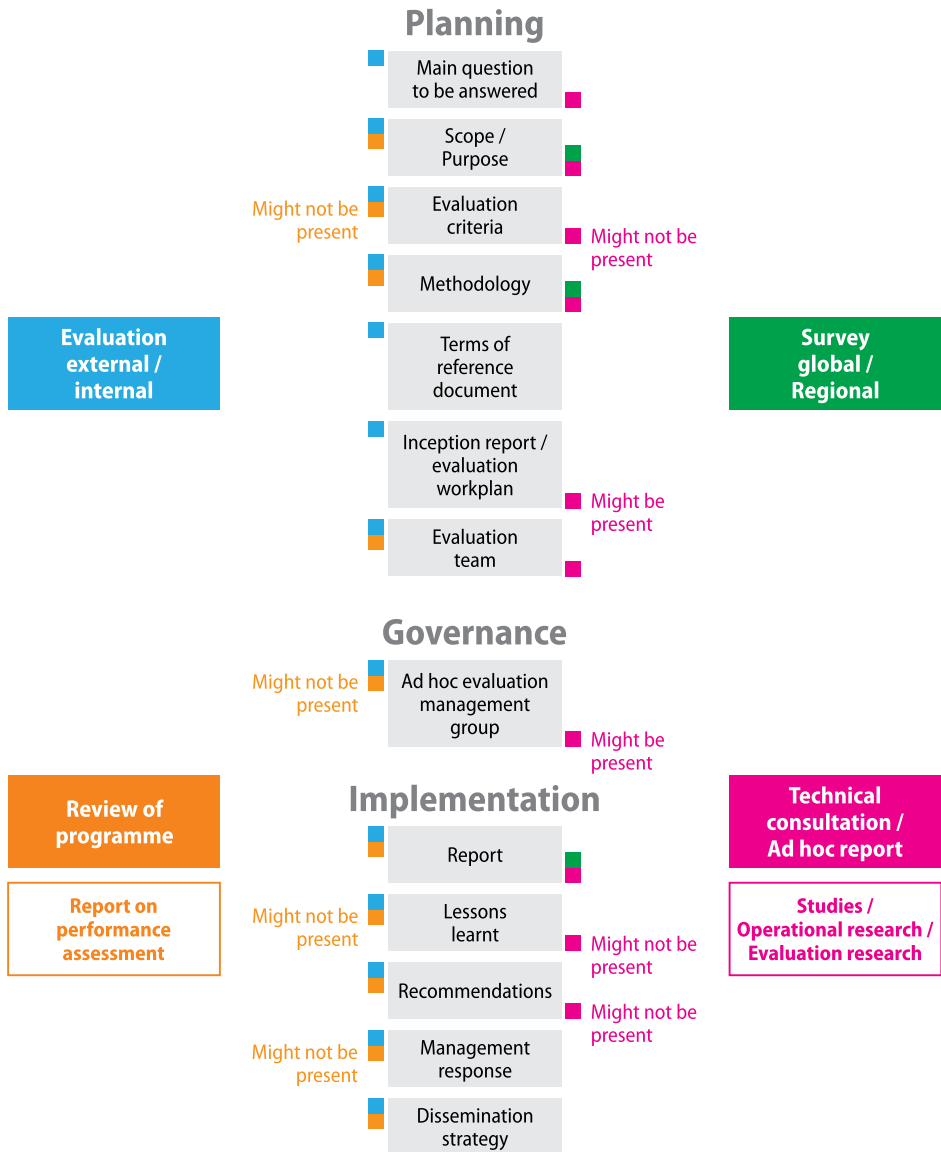
A programme review is the periodic assessment of the performance of an intervention. This category includes structured and periodic exercises, following specific terms of reference – or equivalent detailed guidelines – that examine technical and managerial issues of a programme with a view to identifying what needs to be improved in the short and medium term. Most of these reviews concern programmes in countries. In most cases, a programme review does not apply the methodology of an evaluation. However, these reviews inform evaluations and are part of the development process of the programme.

Audits

An audit assesses the adequacy of management controls to ensure the economical and efficient use of resources; the safeguarding of assets; the reliability of information; compliance with rules, regulations and established policies; the effectiveness of risk management; and the adequacy of organizational structures, systems, processes and internal controls. An audit focuses on compliance while evaluation focuses on results and on understanding what works, why and how. Integrated audits blend the compliance assessment with the analysis of the organizational setting and the achievement of results within the workplan, and the contribution that they make at the beneficiary level.

Annex 3

Basic components of the different types of assessment in WHO*



* Excluding monitoring and audit.

Annex 4

Checklist for compliance with the WHO evaluation policy

All evaluations conducted at WHO shall be carried out in accordance with UNEG norms and standards, as adapted to reflect the specificities of WHO. WHO evaluations shall follow the principles of impartiality, independence, utility, quality and transparency.

Reference	Item	Yes/No	Comments
Terms of reference	<p>The evaluation is based on the terms of reference.</p> <p>The terms of reference specify:</p> <ul style="list-style-type: none"> • the purpose and objectives of the evaluation • context of the evaluation • scope, and linkage to the Programme Budget and the General Programme of Work • evaluation criteria, e.g. relevance, effectiveness, efficiency, impact and sustainability • key evaluation questions • users (owners and audience) of the evaluation results • methodology (approach for data collection and analysis, and involvement of stakeholders) • evaluation team (size, knowledge, skills and required qualifications of evaluators) • a detailed workplan, including a timetable, organization and budget • adherence to WHO cross-cutting strategies on gender, equity and human rights • deliverables (including timing of inception, draft and final report, distribution strategy and follow-up) • as applicable, composition of the ad hoc evaluation management group (including technical staff requirements). 		

Table *continued*

Reference	Item	Yes/No	Comments
	<p>The terms of reference have been made available to major stakeholders and cleared by the ad hoc management group where applicable.</p> <p>The professional and personal integrity of the evaluation team has been assessed for possible conflict of interest.</p> <p>The inception report (as applicable) has been shared with stakeholders and cleared by the ad hoc management group.</p>		
Report	<p>The draft report has been revised to incorporate comments from the evaluation commissioner, the evaluation manager and, where relevant, the ad hoc management group.</p> <p>The final report is structured according to the content specified in the terms of reference.</p> <p>The conclusions of the final report provide answers to the questions listed in the terms of reference.</p> <p>The final report has been delivered in a timely manner.</p> <p>The final report has been accepted by the evaluation commissioner.</p> <p>The final report has been made available to relevant stakeholders and shared with the Global Network on Evaluation.</p>		

Annex 5

Roles and responsibilities – management responses to evaluations

Evaluation recommendation 1.				
Management response:				
Key action(s)	Timeframe	Responsible unit(s)	Tracking	
			Comments	Comments
1.1				
1.2				
1.3				

Evaluation recommendation 2.				
Management response:				
Key action(s)	Timeframe	Responsible unit(s)	Tracking	
			Comments	Comments
2.1				
2.2				
2.3				

Evaluation recommendation 3.				
Management response:				
Key action(s)	Timeframe	Responsible unit(s)	Tracking	
			Comments	Comments
3.1				
3.2				
3.3				

INSTRUCTIONS ON HOW TO USE THE TEMPLATES

Clearance routing

All parties involved in preparing and clearing the management response are requested to enter their name(s), position and units. All management responses should be reviewed by the relevant ADG/DPM office before transfer to IOS.

- Prepared by: include the name of the person preparing the matrix.
- Contributors: include the names and units that contributed actions to the response. At the minimum, this should include all responsible units.
- Cleared by: enter the name and position of the most senior person in the unit who cleared the draft response on behalf of management.

Management responses to evaluations should be clear and comprehensive, and should consist of the following elements.

- Key conclusions and recommendations: are the conclusions and recommendations relevant and acceptable? (The management response should address all recommendations.)
- Key actions: what are the concrete proposed actions? Who are the key partners in carrying out the actions?
- Implementation of actions: what are the responsible units? What is the timeframe for implementation?

Annex 6

Terms of reference of the Global Network on Evaluation

Introduction

Strengthening the evaluation culture across all levels of WHO calls for participatory approaches to evaluation, as outlined in the WHO evaluation policy. Thus, there is a need to establish and maintain a global network for the institutionalization and promotion of evaluation as a means to improve programme performance and results at the beneficiary level through lessons learned and evidence-based planning.

Purpose

The Global Network on Evaluation is an internal network of staff acting to promote the culture of evaluation, facilitate the sharing of information and knowledge management, and strengthen the practice of evaluation at WHO by:

- participating in the preparation of the biennial Organization-wide evaluation workplan and its annual update;
- submitting relevant evaluation reports to the evaluation inventory;
- following up on the status of management responses to evaluation recommendations;
- acting as focal points for evaluation in their respective areas;
- advising programmes across WHO on evaluation issues, as needed.

Membership

Chair

The GNE is chaired by the Executive Director of the Director-General's Office and IOS will provide the support structure for the network.

Composition

The GNE is composed of 23 staff members acting as focal points on evaluation matters at country, regional, headquarters and global levels, as follows:

- country level – one country office representative per region (6);
- regional level – one regional office representative per region (6);
- headquarters – one representative per cluster at headquarters (11);

- global – one representative from each of the seven departments addressing cross-cutting issues of particular relevance to the implementation of the evaluation policy (7).
The departments are: Country Collaboration; Communications, Gender, Equity and Human Rights; Internal Oversight Services; Knowledge Management and Sharing; Information Technology; and Planning, Resource Coordination, and Performance Management.

Nomination

To ensure an inclusive level of representativeness, the following nominations will be made.

- Each regional director will nominate a country-level focal point and a regional focal point.
- Each assistant director-general will nominate a focal point to represent each cluster. If the option of categories is chosen, the focal points will be chosen in consultation with the categories' leaders.
- Each director of the departments representing cross-cutting issues at the global level will nominate a focal point.

Profile of focal points

The following is the suggested profile of the focal points:

- country office level – head of WHO country office with a strong background in evaluation, who has the capacity to champion evaluation issues at the country level within the region;
- regional level – staff members working at regional level (ideally in the office of the director of programme management, assistant regional director or deputy regional director) whose current functions include monitoring and evaluation;
- headquarters level – staff members with responsibilities for monitoring and evaluation within their clusters;
- global level – staff members working on monitoring and evaluation within the departments addressing cross-cutting issues of special relevance to evaluation in WHO.

Expected commitment of each focal point

At present and until the GNE is fully operational, it is expected that each focal point would be able to commit to participating in:

- two annual meetings of the GNE (following the establishment of the network, a general meeting will agree on the identified plan of action with respect to the deliverables, the detailed method of work and the composition of ad hoc working groups);
- specific ad hoc working group(s) dealing with matters such as the quality control approach, consolidation of emerging technical issues that affect the evaluation policy in WHO, and selection criteria for prioritization of individual evaluations;
- other activities of the GNE such as assessment of evaluation material, capacity-building or discussion on matters pertaining to the network.

The current estimated commitment is 5–10% of the professional time and effort of each focal point. Focal points are expected to discuss with their supervisors the appropriate reflection of their role as focal points to the GNE in the Performance Management Development System (PMDS).

Methods of work

The GNE will perform its task virtually through electronic communications (messaging, teleconferences) for its regular business. However, it will consider physical meetings when circumstances permit, such as taking advantage of meetings of other networks (e.g. those of the networks of planning officers or country support).

The GNE may decide to establish ad hoc working groups on specific issues dedicated to the preparatory work to be submitted to the network for consideration, decision and action within its terms of reference.

The GNE secretariat is the responsibility of IOS. IOS ensures the smooth functioning of the GNE by providing the following.

- *Logistics for the regular business of the GNE.* This includes managing the GNE agenda and ensuring that the deliverables are achieved on time, in particular proposing the timing of the meetings and ensuring their calling, identifying agenda items, drafting minutes and following up on what has been agreed. IOS support also includes proposing modalities to address various issues, such as the process for choosing chairs, and products for the subgroups. For each deliverable, IOS will propose a plan to the GNE aligned with the requirements and commitments outlined in the evaluation policy.
- *Administration of the work of the GNE.* In particular, this relates to administration of the web site on evaluation, and management of the evaluation inventory and the database of experts.

- *Technical backup* as needed on evaluation issues. This includes ensuring the linkage with other networks such as UNEG.
- *Dissemination of information* on the work of the GNE and evaluation issues in accordance with the WHO evaluation policy.

Communication within the GNE remains internal unless the network decides otherwise and agrees on the information dissemination approach to the specific topic considered.

Deliverables

Key deliverables

The implementation of the WHO evaluation policy considers several interrelated products that constitute the minimal outputs of the GNE. These deliverables will be submitted to WHO governing bodies in accordance with the evaluation policy.

- *Organization-wide evaluation workplan.* The GNE assists with the identification of the biennial Organization-wide evaluation workplan, which will be updated annually. The evaluation policy outlines the principle criteria to be used for the selection of evaluation items across WHO. However, there is a need to further refine these criteria to make them more specific, and to agree on the weighting to be attached to each criterion to prioritize the areas/topics to be evaluated.
- *Annual evaluation report.* The GNE provides input to the report, including the annual update on the Organization-wide evaluation workplan.
- *Evaluation registry.* The GNE is responsible for identifying, collating and submitting the evaluative work, qualifying as the working definition of evaluation within the WHO evaluation policy, to the WHO evaluation inventory. IOS will support the maintenance of the inventory.
- *Quality control and quality assurance system.* The role of the GNE in relation to the quality assurance system is twofold. On the one hand, the GNE needs to agree on the quality control mechanism to ensure good-quality evaluations and appropriate follow-up of their recommendations across WHO. This includes the establishment of the checklists and standards to be used by staff involved in evaluations, to ensure that evaluations are of the highest quality. Checklists and guidelines will be used by the GNE as quality control tools as needed. On the other hand, the GNE needs to develop a proposal

for the periodic review (e.g. every three years) of the wider quality assurance system on evaluation across WHO. This proposal needs to be in line with other accountability approaches in WHO and is a mid-term deliverable that will be proposed to WHO senior management for action. Some of the components will include peer reviews of the evaluation material and products, meta-evaluations and training on specific aspects that should be used uniformly across WHO to ensure internal and external validity of the evaluation products and of the evaluation function. The GNE will take advice from the focal point of the Department on Gender, Equity and Human Rights to ensure that all WHO evaluations adhere to the relevant policies on gender and human rights.

Other deliverables

The GNE acts as a think tank on the critical issues in relation to evaluation across the Organization. This includes ensuring the minimum competencies of staff to implement the WHO evaluation policy, sensitization on specific evaluation aspects relevant to WHO and contributing to a pool of evaluation resources.

- *Strengthening capacity.* A crucial component of the evaluation culture is the strengthening of the capacity and practice of evaluation across WHO. With this perspective, the GNE will identify an agenda of activities geared to ensuring that a sufficient capacity is established and maintained to implement the evaluation policy in WHO. The GNE will identify a road map to achieve or support this capacity-building, including developing proposals for submission to the Global Learning Committee, Staff Development Fund.
- *Guidance on specific issues.* The GNE will consider specific guidance on issues related to evaluation in WHO as necessary. Some of these issues include the costs of evaluations, resourcing of the implementation of the WHO evaluation policy, relations between centralized and decentralized functions and the evaluation of impact in the WHO context.
- *Database of evaluation experts.* WHO will use the database format available at UNEG to ensure compatibility of the database content and to foster its use by and beyond WHO. The content of the database will remain internal to WHO. IOS will support the maintenance of the database, based on inputs from the GNE. However, each member of the GNE is responsible for its content and for raising issues to ensure its overall quality.

- *Lessons learned.* The GNE will produce a synthesis of the results of the evaluation in order to provide a composite body of information that managers across WHO can utilize in their planning and implementation of programmes. The executive summary of evaluation reports should form the basis of such a synthesis document.

Information dissemination approach

The GNE will use several channels to communicate information depending on its target audience.

- *Electronic means through WHO web sites dedicated to evaluation.* The Intranet site will provide all staff across WHO and, as appropriate, the public in general (via the Internet site), with access to the Organization-wide evaluation workplan, evaluation inventory and the capacity-building agenda; guidance on specific issues; and links to the evaluation expert database and to external sites of evaluation resource networks.
- *Briefings to WHO senior management.* The GNE will provide briefings on specific issues related to its work for the consideration of WHO senior management, as appropriate.
- *Capacity-building activities.* The GNE will take advice from the focal points of the Department of Knowledge Management and Sharing and that of Global Learning and Performance Management and identify the calendar of activities and the related delivery mechanisms. These could include lunchtime seminars, webinars, presentations and work through other existing networks. Examples of networks considered are the network for planning officers or the country support network, given that the focal points in the evaluation GNE also address evaluation issues at the regional level.

Annex 7

Advantages, limitations and methodologies of participatory evaluation

Fig. A7.1
Advantages of participatory evaluations

Identify relevant evaluation questions

- Participatory evaluation ensures that the evaluation focuses on questions relevant to the needs of programme planners and beneficiaries. Participatory approaches allow local stakeholders to determine the most important evaluation questions that will affect and improve their work.

Improve programme performance

- Participatory evaluation is reflexive and action-oriented. It provides stakeholders, including beneficiaries, with the opportunity to reflect on project progress and to generate knowledge that results in the ability to apply the lessons learnt. It provides opportunities for groups to take corrective action and make mid-course improvements.

Empower participants

- A participatory approach is empowering because it claims the right for stakeholders to control and own the process of making evaluation decisions and implementing them. Participating in an evaluation from start to finish can give stakeholders a sense of ownership of the results. Recognizing local capacities and expertise builds confidence in the community and among participants.

Build capacity

- Conducting a participatory evaluation promotes participant learning and is an opportunity to introduce and strengthen evaluation skills. Active participation by stakeholders can result in new knowledge and a better understanding of their environment. This, in turn, enables groups to identify action steps and to advocate for policy changes. It can provide participants with tools to transform their environments.

Develop leaders and build teams

- Participatory evaluation builds teams and participant commitment through collaborative enquiry. Inviting a broad range of stakeholders to participate and lead different parts of the process can develop and acknowledge stakeholders' leadership skills. It can lead to stronger, more organized groups, strengthening the community's resources and networks.

Fig A7.1 *continued*

Sustain organizational learning and growth

- A participatory evaluation is not just interested in findings; it is focused on creating a learning process. It creates a knowledge base among stakeholders, which can be applied to other programmes and projects. The techniques and skills acquired can lead to self-sustained action.

Box A7.1

Limitations of participatory evaluations

Such evaluations involve active participation of multiple stakeholders, which include beneficiaries, the implementing organization and the operating unit at each phase of the evaluation process (planning, data collection, analysis, reporting, dissemination and follow-up actions). A common modality involves collecting background material and circulating it among the stakeholders. These stakeholders analyse the material and explore its implications in a workshop or a series of workshops. Findings and recommendations are formulated by a panel. These workshops enable managers of operating units to listen and respond to stakeholders. Face-to-face interactions facilitate better understanding of the workings of a project or programme and its achievements and problems. Participants often come up with new ideas for solving problems or improving performance. As managers themselves participate in the evaluation process, they are inclined to use resulting information and recommendations.

However, participatory evaluations have many limitations. Such evaluations tend to be less objective because participants have vested interests which they articulate and defend in such workshops. Moreover, they are less useful in addressing complex technical issues, which may require specialized technical expertise. Yet another limitation is that, although they may generate useful information, their credibility is limited because of their less formal nature.

Source: USAID (2009). *Evaluation guidelines for foreign assistance*. Washington, DC, Office of the Director of United States Foreign Assistance (http://betterevaluation.org/resources/guide/usaid_foreign-assistance_guidlines, accessed 12 September 2013).

Box A7.2

Methods commonly used in participatory evaluations

The participatory approach to evaluation is aimed at promoting action and community-level change. It tends to overlap more with qualitative than with quantitative methods. However, not all qualitative methods are participatory and, inversely, many participatory techniques can be quantified.

As with qualitative methods, participatory evaluation ensures that the perspectives and insights of all stakeholders and beneficiaries, as well as project implementers, are taken into consideration. However, the participatory approach is very action-oriented. The stakeholders themselves are responsible for collecting and analysing the information and for generating recommendations for change.

Box A7.2 *continued*

The role of an outside evaluator is to facilitate and support this learning process. Participatory monitoring and evaluation develops ownership by placing a strong emphasis on building the capacity and commitment of all stakeholders to reflect, analyse and take responsibility for implementing any changes they recommend.

Typically, participatory methods have been used to learn about local conditions and local people's perspectives and priorities during project appraisal. However, one can go further and use participatory methods not only at the project formulation stage but throughout the duration of the project, and especially for evaluating how the participants perceived the benefits from the project. Participatory monitoring and evaluation is an important management tool that provides task managers with quick feedback on project effectiveness during implementation. This has become increasingly important as development interventions move away from "blueprint projects" towards the more flexible planning that enables projects to learn and adapt on the ground.

There are many different participatory information collection and analysis tools. Most of these are not inherently monitoring and evaluation tools but can be used for a variety of purposes ranging from project planning and community mobilization to monitoring and evaluation, depending on the way they are employed. As with all participatory approaches, the key to success is to be flexible and innovative in the use of appropriate tools and methods, and to be willing to adapt to local circumstances.

Participatory methodologies and the associated tools and techniques which are commonly used in participatory monitoring and evaluation include beneficiary assessment; participatory rural appraisal; and self-esteem, associative strength, resourcefulness, action planning and responsibility (SARAR).

- **Beneficiary assessment.** This is a consultative methodology used in evaluations (and other stages of the project cycle) to gain insights into the perceptions of beneficiaries regarding a project or policy. The overall objective of a beneficiary assessment is to make the voices of beneficiaries and other local stakeholders heard by those managing a project or formulating policy. The focus of beneficiary assessments is on obtaining systematic qualitative information, including subjective opinions, to complement the data from quantitative evaluations. Wherever possible, beneficiary assessment results are quantified and tabulated. Moreover, sample sizes are selected with credibility in mind. Although beneficiary assessment results are not usually conducive to statistical analysis, they are based on more than just anecdotal information. The systematic nature of beneficiary assessments also enhances the reliability of the findings through the combination of techniques used to gather information. Such techniques allow for cross-checking of responses, and a reasonable assessment of the extent to which opinions expressed by respondents represent widely held views in their community. However, the actual techniques used and the beneficiary assessment process itself will depend on the topic and circumstances of the work.

Box A7.2 *continued*

In addition to generating descriptive information, beneficiary assessments are designed to produce recommendations, as suggested by those consulted, for changes to the current or planned policies and programmes. This action-oriented nature of beneficiary assessment work requires that the results be produced with a minimum of delay after completion of fieldwork so that the necessary adjustments to projects or policies can then be identified and undertaken.

The most common application of beneficiary assessment techniques has been in projects with a service delivery component where it is especially important to gauge user demand and satisfaction. During implementation, beneficiary assessments can provide feedback for monitoring purposes and for reorientation of the project. Towards the end of the project, beneficiary assessments can also complement technical and financial evaluations, as well as survey-based impact evaluations with the views of the beneficiaries themselves.

The primary audiences of beneficiary assessment findings are decision-makers and managers of the development activity. For this reason, special efforts are made to seek the involvement of these decision-makers in the beneficiary assessment process from the design stage to the review and final presentation of the results.

Beneficiary assessments usually make use of three qualitative methods of information gathering, namely: semi-structured individual interviews, focus group discussions, and participant observation. Semi-structured interviews provide the bulk of the findings. They are meant to be quantified, and hence the sample must be large enough and representative. Focus group interviews and participant observation are done primarily for illustration and contextual background and need not conform to the same standards of representativity.

The quality and effectiveness of beneficiary assessments depend heavily on the training and preparedness of the field workers and the appropriate supervision and monitoring of their work. Where field workers are unclear about the kind of information required for the evaluation, the common tendency is to collect lengthy, descriptive and very detailed information on individual cases, rather than focusing only on the relevant topics. For this reason, there should be at least one opportunity to review the preliminary findings and methods, preferably midway through the fieldwork, so that this kind of problem can be addressed in time to reorient the field workers' work.

Another limitation seen in some beneficiary assessments is the failure to ensure active participation by key decision-makers throughout the process. In this case, even if the findings are of good quality and highly relevant, they are unlikely to generate much impact. Without a sense of ownership, decision-makers may not accept the findings, particularly if they are somewhat controversial and critical of the project or policy concerned. This caveat applies to all evaluation work regardless of the type of approach or technique used.

Box A7.2 *continued*

- **Participatory rural appraisal.** This comprises a set of techniques aimed at shared learning between local people and outsiders. The term itself is misleading because participatory rural assessment is increasingly being used not only in rural settings, and not only for project appraisal, but throughout the project cycle as well as for research studies. Indeed, the term “participatory rural assessment” is one of many labels for similar participatory assessment approaches, the methodologies of which overlap considerably. It is probably more useful to consider the key principles behind participatory rural assessment and its associated techniques, rather than the name as such, when assessing its appropriateness to a particular situation. There are five key principles that form the basis of any participatory rural assessment activity no matter what the objectives or setting.
- *Participation.* Participatory rural assessment relies heavily on participation by communities, as the method is designed to enable local people to be involved, not only as sources of information, but also as partners with the participatory rural assessment team in gathering and analysing the information.
 - *Flexibility.* The combination of techniques that is appropriate in a particular development context will be determined by such variables as the size and skill mix of the participatory rural assessment team, the time and resources available and the topic and location of the work.
 - *Teamwork.* Generally, a participatory rural assessment is best conducted by a local team (speaking the local languages) with a few outsiders present, a significant representation of women and a mix of sector specialists and social scientists, according to the topic.
 - *Optimal ignorance.* To be efficient in terms of both time and money, participatory rural assessment work is aimed at gathering just enough information to make the necessary recommendations and decisions.
 - *Systematic.* As data generated by participatory rural assessments are seldom conducive to statistical analysis (given the largely qualitative nature and relatively small sample size), alternative ways have been developed to ensure the validity and reliability of the findings. These include sampling based on approximate stratification of the community by geographical location or relative wealth and cross-checking – i.e. using a number of techniques to investigate views on a single topic (including through a final community meeting to discuss the findings and correct inconsistencies).

Participatory rural assessment offers a “basket of techniques” from which those most appropriate for the project context can be selected. The central part of any participatory rural assessment is semi-structured interviewing. While sensitive topics are often better addressed in interviews with individuals, other topics of more general concern are amenable to focus group discussions and community meetings.

Box A7.2 *continued*

During these interviews and discussions, several diagrammatic techniques are frequently used to stimulate debate and record the results. Many of these visuals are not drawn on paper but on the ground with sticks, stones, seeds and other local materials, and then transferred to paper for a permanent record.

Key diagrammatic techniques of participatory rural assessment include mapping techniques, ranking exercises and trend analysis. Visual-based techniques are important tools for enhancing a shared understanding between outsiders and insiders, but may hide important differences of opinion and perspective when drawn in group settings and may not reveal culture-based information and beliefs adequately. They therefore need to be complemented by other techniques, such as careful interviewing and observation, to cross-check and supplement the results of diagramming.

Participatory rural assessment involves some risks and limitations. Many are not unique to this method but are inherent in any research method that aims to investigate local conditions. One of the main problems is the risk of raising expectations. This may be impossible to avoid, but can be minimized with careful and repeated clarification of the purpose of the participatory rural assessment and the role of the team in relation to the project, or government, at the start of every interview and meeting. Trying to use participatory rural assessment as a standard survey to gather primarily quantitative data, using large sample sizes and a questionnaire approach, could greatly compromise the quality of the work and the insights produced. Also, if the participatory rural assessment team is not adequately trained in the methodology before the work begins, there is often a tendency to use too many different techniques, some of which are not relevant to the topic at hand. In general, when a training element is involved, there will be a trade-off between the long-term objective of building the capacity of the participatory rural assessment team and getting good-quality results in their first experience of using the methodology.

Furthermore, one common problem is that insufficient time is allowed for the team to spend with the local people, to listen to them and to learn about the more sensitive issues under consideration. Rushing will also often mean missing the views of the poorest and least articulate members of the communities visited. The translation of participatory rural assessment results into a standard evaluation report poses considerable challenges, and individuals unfamiliar with participatory research methods may raise questions about the credibility of the findings.

- **Self-esteem, associative strength, resourcefulness, action planning and responsibility (SARAR).** This is an education/training methodology for working with stakeholders at different levels to engage their creative capacities in planning, problem-solving and evaluation. The acronym SARAR stands for the five attributes and capacities that are considered the minimum essentials for participation to be a dynamic and self-sustaining process:

Box A7.2 *continued*

- *self-esteem*: a sense of self-worth as a person as well as a valuable resource for development;
- *associative strength*: the capacity to define and work towards a common vision through mutual respect, trust and collaborative effort;
- *resourcefulness*: the capacity to visualize new solutions to problems even against the odds, and the willingness to be challenged and take risks;
- *action planning*: combining critical thinking and creativity to come up with new, effective and reality-based plans in which each participant has a useful and fulfilling role;
- *responsibility*: for follow-through until the commitments made are fully discharged and the hoped-for benefits achieved.

SARAR is based on the principle of fostering and strengthening these five attributes among the stakeholders involved in the evaluation. Such a process will enable the development of those people's own capacities for self-direction and management, and will enhance the quality of participation among all stakeholders. The various SARAR techniques can be grouped into five categories according to how they are most commonly used. While there is no set order in which these techniques are used, the five types of technique are often applied progressively, having a cumulative effect.

- Creative techniques involve the use of open-ended visual tools such as mapping and non-serial posters to encourage participants to break out of conventional ideas and routine ways of thinking.
- Investigative techniques such as pocket charts are designed to help participants do their own needs assessment by collecting and compiling data on problems and situations in their community.
- Analytical techniques, including three pile-sorting and gender-analysis tools, enable participants to prioritize problems and opportunities and to examine a problem in depth, allowing them to better understand its causes and identify alternative solutions.
- Planning techniques are used to simplify the planning process so that decisions can be made, not only by the more prestigious and articulate participants (such as community leaders or senior staff), but also by the less powerful, including non-literate community members.
- Informative techniques help in gathering information and using it for better decision-making.

At the outset, participants are involved in using their creativity to look at situations in new ways and to build their capacity for self-expression. Then they gain tools for investigating and analysing reality in more detail. Finally, they develop skills in gathering information, making decisions and planning initiatives.

Box A7.2 *continued*

Less successful applications of SARAR have usually been traced to insufficient training of the SARAR facilitators. Without adequate preparation, facilitators will not feel comfortable experimenting with the different techniques and may be more inclined to adopt a blueprint approach – i.e. always using the same set of techniques in a predetermined way and not being responsive to the differences among communities or the various groups of stakeholders. In other cases, problems have arisen when the use of SARAR techniques has been considered an end in itself, rather than a means to support the development and implementation of project activities. This problem can occur when SARAR activities are not linked to concrete follow-up activities. In such cases communities eventually see no benefit in being involved in the SARAR sessions and the whole process begins to break down.

The effectiveness of SARAR, like that of similar participatory techniques, can also be limited by a general resistance – usually by higher-level managers and decision-makers rather than field workers or community members – to the use of qualitative, informal and visual-based techniques. This can lead to problems if these sceptics obstruct the SARAR process by dismissing the results as unscientific or the participatory process itself as inefficient.

These three methods can be used alone or can be combined in a single evaluation. They represent only a small sample of the vast range of participatory techniques that can be used for monitoring and evaluation. It should be noted that none of these participatory methods is intended to be a replacement for good-quality survey work. Indeed, they are often used in conjunction with other methods. For example, the findings from a preliminary study using participatory approaches can usefully give direction and focus to subsequent survey-based evaluations. In turn, the survey can verify and quantify the qualitative findings from participatory evaluations and be applied on a larger scale. Participatory evaluations done after quantitative surveys can verify or challenge survey findings, and can go some way toward explaining the information collected by the quantitative survey-based evaluations.

Annex 8

Key elements of the joint evaluation process

The planning and conduct of a joint evaluation are generally similar to any other well-managed evaluation. However, there are a number of specific issues that need addressing. In particular, it is important to assess whether the programme/project warrants a joint evaluation and to discuss the purpose of the evaluation by asking the following questions.

- Is the focus of the programme/project an outcome that reaches across sectors and agencies?
- Is the programme/project co-financed by multiple partners?
- Is the topic a contentious issue, thus calling for a balanced approach?

It is essential to determine the partners at an early stage to ensure their involvement and ownership. One way to identify key partners is to focus on where the financing comes from, who the implementing partners are, or which other agencies or institutional partners may contribute to the overall programme/project's goal or outcome. It is also important to assess the potential contributions of partners and discuss the level of objectivity that they may or may not have to ensure that the evaluation is independent and free from strong biases.

Choosing an effective management structure and strong communications system is critical to the evaluation process. To manage the conduct of the evaluation, a two-tiered structure can be established with a management group that oversees the process and a smaller management group to ensure that implementation goes smoothly. This ad hoc evaluation management group would normally include a representative from each partner organization and government entity and would meet at specific times to approve the terms of reference and the evaluation team, ensure oversight of the evaluation, introduce balance in the final evaluation judgements, and take responsibility for the use of results. Depending on the scope of the evaluation, the ad hoc evaluation management group, bringing together technical representatives from concerned organizations or entities, could be responsible for daily management tasks such as approving an evaluation manager to deal with the recruitment and management of the evaluation team. It is extremely important to agree early on decision-making arrangements and the division of labour with the other partners. This includes deciding who among the management group will take the lead role in each of the subsequent steps in the evaluation. A conflict resolution process should be determined to deal with any problems that may arise.



Evaluation partners need to agree on the scope of the evaluations, the issues to be covered and the timeframe of the exercise. This implies discussing proposed terms of reference and determining which agency's procedures will be followed. It is important to allow flexibility to adapt and additional time to accommodate delays due to the different approaches to evaluation that different organizations may have. There are two ways to manage this, either:

- to agree that the evaluation will be managed using the systems and procedures from one agency; or
- to split the evaluation into components and agree whose systems will be used to manage which components.

When WHO takes the lead, the preferred approach to funding should be for partners' financial support to be pooled in a fund that is administered by one agency and covers all costs related to the exercise. The second option, where individual partners finance certain components of the evaluation while WHO covers others, increases transaction and coordination costs.

Regarding the selection of the evaluation teams, there are also two options: either tasking one of the partners with recruiting the evaluation team in consultation with the other partners, or asking each partner to contribute its own experts. All parties involved should agree on an evaluation team leader, or delegate to a particular partner the recruitment of the team leader, and make clear to the evaluation team that its independence will be respected.

Finally, partners need to agree on the report and dissemination strategy. They should agree that they all have the opportunity to correct factual errors in the report and, if it is impossible to resolve differences on the findings and conclusions, to request that dissenting views be included in the report. Sometimes it may be necessary to allow for separate evaluation products to ensure that all partners' accountability or reporting requirements are fulfilled.

Follow-up may be difficult on a joint evaluation report as the internalization of the findings and implementation of the recommendations need to be done at the level of individual institutions and of the partnership between them. Therefore, partners need to agree on what to do individually and collectively and devise follow-up mechanisms to monitor the status of the changes. WHO may select recommendations that are pertinent to WHO and prepare a management response focusing on these recommendations.

Annex 9

Evaluation workplan: criteria for selection of evaluation topics

Introduction

1. The evaluation policy states that WHO will develop a biennial, Organization-wide evaluation workplan as part of the Organization's planning and budgeting cycle. This biennial workplan, to be updated annually, ensures accountability and oversight of performance and results, and reinforces organizational learning in a way that informs policy for decision-makers and supports individual professional development.
2. The evaluation workplan is one of the deliverables of the evaluation policy, and its identification is among the most critical contributions of the Global Network on Evaluation.
3. Evaluation workplans constitute the annual and biennial iteration of a broader multi-year Organization-wide evaluation agenda. The evaluation agenda includes a combination of:
 - evaluation of WHO products, entities and functions (projects, programmes, initiatives and offices) and of the WHO evaluation function;
 - evaluations across WHO under the centralized and the decentralized evaluation functions.

Identification of the evaluation workplan

Evaluation universe

4. For practical purposes, WHO will consider two types of boundaries when identifying the evaluation workplan.
 - a) Evaluation commissioner. Only evaluations that are commissioned by the WHO Secretariat or jointly with other stakeholders in the case of partnerships will be included in the workplan. Evaluations commissioned by WHO governing bodies, or other stakeholders, will be referred to when prioritizing what needs to be evaluated, since one of the criteria is the time since the last evaluation of any evaluation candidate.

- b) The evaluation universe comprises the following.
- **Office-specific evaluations** include all budget centres in WHO, such as WHO country offices and departments or units at headquarters or regional offices. The list of budget centres relates to the WHO Programme Budget and is available within the Secretariat.
 - **Programmatic evaluations** include all global programmes and initiatives when considering more than one budget centre, covering at least two levels within WHO – e.g. a global initiative or normative work being evaluated at headquarters and regional levels, or a regional strategy or programme being evaluated at regional and country levels. The provisional list of programmes/normative work, strategies, and initiatives potentially included for programmatic evaluations is available online in the Evaluation Registry and will be completed through discussion with WHO senior management and the Global Network on Evaluation before 1 October every year.
 - **Thematic evaluations** include any selected topic of corporate institutional interest such as a new way of working, a corporate strategy, a cross-cutting theme or core function, or an emerging issue. The full list of selected topics of corporate institutional interest will be completed through consultation with WHO senior management, the Global Network on Evaluation and IOS before 1 October every year.

Evaluation selection criteria

5. WHO evaluation policy outlines the three broad categories grouping the criteria for the selection of topics for evaluation, namely:
- **organizational requirements** relevant to global, international or regional commitments; specific agreements with stakeholders, partners or donors; and 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 achievement of expected results;
 - **organizational utility** relating to a cross-cutting issue, theme, programme or policy question; potential for staff or institutional learning (innovation); and degree of comparative advantage of WHO.

Box A9.1 provides further details of the specific criteria to be used for the identification of the workplan.

Box A9.1

Criteria for the identification of the biennial WHO-wide evaluation workplan**Organizational requirement***Global, international or regional commitments:*

- Millennium Development Goals
- disease eradication strategies
- disease elimination strategies
- International Health Regulations
- other areas subject to formal reporting to the World Health Assembly
- other areas subject to formal reporting to regional committees
- other areas subject to formal reporting to the United Nations General Assembly
- other areas subject to formal reporting to other global or international forums.

Specific agreements with stakeholders, partners or donors:

- inclusion of evaluation requirement in the collaborative agreement of any area, project, programme or initiative at global or headquarters level, and its timing;
- inclusion of evaluation requirement in the collaborative agreement of any area, project, programme or initiative at regional level, and its timing;
- inclusion of evaluation requirement in the collaborative agreement of any area, project, programme or initiative at country level, and its timing.

Requests from governing bodies:

- any specific evaluation request put forward by the governing bodies.

Organizational significance*Level of investment.**Inherent risks:*

- impact on reputational risks
- timing since the last evaluation
- complexity and associated inherent risks.

Performance issues or concerns in relation to achievements of expected results:

- recurrent issues identified through IOS work
- other issues identified through the Global Network on Evaluation.

Organizational utility*Cross-cutting issue, theme, programme or policy question:*

- potential for staff or institutional learning, including the potential for replication of innovative/catalytic initiatives;
- flagship programme or strategy for WHO Global Programme of Work;
- relevant to the WHO reform process.

Box A9.1 *continued*

Degree of comparative advantage of WHO:

- in relation to its core functions
- in relation to production of global public goods.

When applying the criteria, other related issues need to be considered. These include:

- the evaluability of the project (technical, operational)
- the utilization of the evaluative funding
- the existence of other evaluation mechanisms in place.

In addition, evaluations are mandatory for programmes and initiatives once in their life-cycle when at least one of the following conditions apply.

- WHO has agreed to a specific commitment with the related stakeholders over that life-cycle.
- The programme or initiative's life-cycle exceeds a cycle of the Global Programme of Work.
- The programme or initiative's cumulative investment size exceeds 2% of the Programme Budget.

Prioritization

6. Each specific criterion needs to be assigned a value, with a view to prioritizing the items to be included in the evaluation workplan. The value attached to each criterion is not fixed beforehand and needs to be agreed upon through a consultation process with the support of the Global Network on Evaluation before 1 October each year.

Annex 10

Checklist for evaluation terms of reference¹

Reference	Item	Yes/No	Comments
1. Evaluation purpose	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. specify the purpose of the evaluation and how it will be used. b. define the mandate for the conduct of the evaluation. c. clearly state why the evaluation is being done, including justification for why it is being done at this time. d. identify the primary and secondary audiences for the evaluation, and how the evaluation will be useful. 		
2. Evaluation objectives	<ul style="list-style-type: none"> a. The terms of reference include clearly defined, relevant and feasible objectives. b. The objective(s) clearly follow from the overall purpose of the evaluation. c. The objectives described in the terms of reference are realistic and achievable in light of the information that can be collected in the context of the undertaking. 		

¹ Source : adapted from UNEG (2012). *Standards for Evaluation in the UN System*. New York, United Nations Evaluation Group (http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=21, accessed 24 September 2013).

Table *continued*

Reference	Item	Yes/No	Comments
3. Evaluation context	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. include sufficient and relevant contextual information. b. adequately describe the particular political, programmatic and governance environment in which the evaluation will be taking place. For example, the most relevant aspects of the economic, social and political context are described. c. adequately describe the most relevant programmatic and/or thematic aspects relevant to the evaluation. 		
4. Evaluation scope	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. explicitly and clearly define what will and will not be covered, including the timeframe, phase in the project and/or geographical areas to be covered by the evaluation. b. establish the linkage between the subject of the evaluation and the General Programme of Work and Programme Budget. c. show that the scope of the evaluation is adequate to meet the stated objective(s). d. show that the scope of the evaluation is feasible, given resources and time considerations. 		

Table *continued*

Reference	Item	Yes/No	Comments
5. Evaluation criteria	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. specify the criteria that will be utilized to guide the evaluation. b. specify the evaluation criteria against which the subject to be evaluated will be assessed, including, for example, relevance, efficiency, effectiveness, impact and sustainability. c. spell out any additional criteria of relevance to the particular type of evaluation being undertaken, such as evaluations of development, humanitarian response, and normative programmes. 		
6. Key evaluation questions	<ul style="list-style-type: none"> a. The terms of reference include a comprehensive and tailored set of evaluation questions within the framework of the evaluation criteria. b. The terms of reference contain a set of evaluation questions that are directly related to both the objectives of the evaluation and the criteria against which the subject will be assessed. c. The set of evaluation questions adds further detail to the objectives and contributes to further defining the scope. d. The set of evaluation questions is comprehensive enough to raise the most pertinent evaluation questions, but also concise enough to provide users with a clear overview of the evaluation objectives. 		

Table *continued*

Reference	Item	Yes/No	Comments
7. Users	a. The terms of reference should identify who are the users (owners and audience) of the evaluation results. This could include responsible WHO staff, implementing partners, recipients of the intervention, policy-makers, and other stakeholders in the activity being evaluated.		
8. Methodology	<p>a. The terms of reference specify the methods for data collection and analysis, including information on the overall methodological design.</p> <p>b. The terms of reference contain a clear and accessible methodological plan – preferably a stand-alone section that is clearly differentiated from other information contained in the terms of reference.</p> <p>c. The terms of reference state the overall methodological approach and design for the evaluation. Examples of approaches include participatory, utilization-focused, theory-based and gender- and human rights-responsive. Examples of overall design include non-experimental, quasi-experimental and experimental.</p> <p>d. The data collection and analysis methods in the terms of reference are sufficiently rigorous to assess the subject of the evaluation and ensure a complete, fair and unbiased assessment. For example, there will be sufficient data to address all evaluation questions.</p>		

Table *continued*

Reference	Item	Yes/No	Comments
	<ul style="list-style-type: none"> e. The evaluation methodology includes multiple methods (triangulation), preferably with analysis of both quantitative and qualitative data and with a range of stakeholders covered by the data collection methods. f. Logical and explicit linkages are provided between data sources, data collection methods and analysis methods. For example, sampling plans are included. g. The evaluation methodology takes into account the overall purpose of the evaluation, as well as the needs of the users and other stakeholders. h. The evaluation methodology explicitly and clearly states the limitations of the chosen evaluation methods. i. The terms of reference specify that the evaluation will follow UNEG norms and standards for evaluations as well as ethical guidelines. 		
9. Evaluation team	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. include information on the size of the evaluation team, and identify the team leader. b. specify the required knowledge, skills and qualifications of evaluators. c. describe how the independence and objectivity of the team are ensured, and how conflicts of interest are addressed. 		

Table *continued*

Reference	Item	Yes/No	Comments
10. Evaluation workplan	<p>The terms of reference include a workplan for the evaluation. The workplan:</p> <ul style="list-style-type: none"> a. states the outputs that will be delivered by the evaluation team, including information on the degree to which the evaluation report will be accessible to stakeholders, including the public. b. describes the key stages of the evaluation process and the project timeline. c. establishes clear roles and responsibilities for evaluation team members, the commissioning organization and other stakeholders in the evaluation process. d. describes the quality assurance process. e. describes the process, if any, for obtaining and incorporating comments on a draft evaluation report. f. includes an evaluation project budget. 		
11. Gender, equity and human rights	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. specify how gender, equity and human rights aspects will be incorporated into the evaluation design. b. indicate both duty-bearers and rights-holders (particularly women and other groups subject to discrimination) as primary users of the evaluation and how they will be involved in the evaluation process. c. spell out the relevant instruments or policies on gender, equity and human rights that will guide the evaluation process. 		

Table *continued*

Reference	Item	Yes/No	Comments
	<ul style="list-style-type: none"> d. include an assessment of relevant gender, equity and human rights aspects through the selection of the evaluation criteria and questions. e. specify an evaluation approach and data collection and analysis methods that are gender-sensitive and human rights-based, and specify that evaluation data are to be disaggregated by sex, ethnicity, age, disability, etc. f. define the level of expertise on gender, equity and human rights needed in the evaluation team and the team's responsibilities in this regard, and call for a gender-balanced and culturally diverse team that makes use of national/regional evaluation expertise. 		
12. Deliverables	<p>The terms of reference:</p> <ul style="list-style-type: none"> a. identify the expected deliverables from the evaluation (inception, draft and final report). b. provide details of the timing of the inception report, draft and final report. c. outline the structure of the final report, e.g. the executive summary, the clarity of content, and suitability of format for the intended audience. d. state who will make inputs to the final report, and who has final control over the report's structure and content. e. specify the distribution list of the final report. f. describe the proposed distribution strategy of the final report. 		

Table *continued*

Reference	Item	Yes/No	Comments
13. Ad hoc evaluation management group	<p>If the size and complexity of the evaluation warrants an ad hoc evaluation management group, the terms of reference should:</p> <ul style="list-style-type: none"> a. provide details of the members of the group, including technical requirements. b. specify how the evaluation commissioner has ensured that there is no conflict of interest or compromise of the independence and objectivity of the evaluation process, in the selection of the ad hoc evaluation management group members. 		

Annex 11

Methodological approaches to impact evaluation

The following categories are used to classify evaluation methods. These categories are in practice often combined.

Randomization or experimental design

A randomized control trial (RCT) attempts to estimate a programme's impact on an outcome of interest. An outcome of interest is something – often a public policy goal – that one or more stakeholders care about (e.g. the unemployment rate, which many actors may wish to be lower). An impact is an estimated measurement of how an intervention affected the outcome of interest compared with what would have happened without the intervention. A simple RCT randomly assigns some subjects to one or more treatment groups (also sometimes called experimental or intervention groups) and others to a control group. The treatment group participates in the programme being evaluated and the control group does not. After the treatment group experiences the intervention, an RCT compares what happens to the two groups by measuring the difference between the two groups on the outcome of interest. This difference is considered an estimate of the programme's impact.^a

Propensity score matching

Propensity score matching methods are often used to control for bias when randomization is not possible. These methods were developed to ensure comparability between the treatment and comparison groups in terms of propensity to participate in the development programme. The first step involves estimating the likelihood (the propensity score) that, given certain characteristics, a person would have received the treatment or intervention. The propensity scores are then used to group observations that are close to each other. Comparisons of development results can be applied to different groups of observations that have the same propensity to participate, thus ensuring comparability.^b

^a Source: Clinton T, Nunes-Neto B, Williams E (2006). *Congress and program evaluation: an overview of randomized control trials (RCTs) and related issues*. Washington, DC, Congressional Research Service, Library of Congress.

^b Source: Ravallion M (1999). *The mystery of the vanishing benefits: Ms speedy analyst's introduction to evaluation*. Washington, DC, The World Bank (Working paper No. 2153).



Pipeline comparison

Pipeline comparison methods use those who have applied for and are eligible to receive the intervention in the future, but have not yet received it, as a comparison group. Their only difference from the current recipients is that they have not yet received the intervention.

Simulated counterfactual

Simulated counterfactual methods are used for interventions affecting the entire population, for which no comparison group can be identified. A counterfactual distribution of outcomes in the absence of the intervention is simulated on the basis of a theoretical model and information on the situation prior to the intervention.

Difference in means *or* single difference

Difference in means or single difference methods estimate the impact of an intervention by comparing the value of the indicator of interest for the recipients and nonrecipients.

Difference-in-difference *or* double difference

Difference-in-difference or double difference methods estimate impacts by comparing the value of the indicator of interest for the recipients and nonrecipients before (first difference) and after an intervention (second difference).

Instrumental variables

This method uses instrumental variables (which affect receipt of the intervention but not the outcomes of interest) to control for selection bias when intervention placement is not random.

Annex 12

Core competencies for evaluators

WHO has developed core competencies for evaluators based on the guidance developed by UNEG.¹ The main competencies needed for an evaluator to perform a high-quality evaluation can be categorized as follows.

1. Knowledge of the WHO context:
 - environment
 - policy level of work
 - institutional level of work
 - strategic level of work
 - activity level of work
 - project level of work
 - programme level of work
 - results-based management
 - human rights
 - gender
 - diversity.

2. Technical and professional skills:
 - planning for influential evaluations
 - evaluation design
 - data collection
 - data analysis (quantitative and qualitative)
 - reporting
 - follow-up on recommendations
 - best practices
 - lessons learned
 - dissemination and outreach.

¹ *Core competencies for evaluators of the UN system*. New York, NY, United Nations Evaluation Group, 2008 (http://www.uneval.org/documentdownload?doc_id=1408&file_id=1850, accessed 28 February 2013).



3. Interpersonal skills:

- communication skills (written and oral)
- cultural sensitivity
- negotiation
- facilitation.

4. Personal attributes:

- ethical behaviour
- judgement capacity
- education (evaluation and research)
- work experience (evaluation and research).

5. Management skills:

- managing evaluation process/projects
- team management
- coaching and training
- resource management.

In addition, the evaluation team leader should have the following competencies.

- Work experience: relevant evaluation experience in field work.
- Evaluation design: ability to develop evaluation terms of reference that address salient issues, identify potential impact and use-appropriate evaluation methodologies, including evaluability, at the outset.
- Data collection and analysis: knowledge of evaluation with quantitative and qualitative data collection and analysis tools, techniques and approaches.
- Reporting: ability to draft credible and compelling evaluation reports, with evidence-based findings and recommendations, for maximum impact.
- Managing the evaluation process/project: command of the management process of evaluation projects at various levels (e.g. activity, project and programme levels) as well as the management of evaluation teams.
- Ethics: knowledge of WHO values and ethical behaviour.

Annex 13

Evaluation workplan template

Activity	Timeline	Responsible unit/staff	Collaborating units/offices	Budget (US\$)	Source of funding	Link with relevant evaluation objectives and deliverables	Expected outcome/ key question answered
	Q1						
	Q2						
	Q3						
	Q4						

Instructions for completing the template

Activity

Describe all the evaluation activities to be carried out. Include the assumptions on which the budget is based.

Timeline

Specify the timeline for each evaluation phase/activity within the evaluation process.

Responsible unit/staff

Specify the entity primarily responsible for carrying out the activity and indicate the level of detail required.

Collaborating units/offices

Indicate any collaborating units/support from the WHO Secretariat and others.

Budget

Indicate the budget (in US\$) required for the implementation of the activity.

Source of funding

Indicate whether the budget is directly tied to the Organization's budget. If not, indicate the external source of funding. If funding is not yet secured, mark "not yet secured" and indicate the source from which funding will be sought.

Link with relevant evaluation objectives and deliverables

Provide a reference to the relevant action plan or other recommendations.

Expected outcome/key question answered

Indicate precisely which question is addressed and how it relates to the evaluation criteria.

Annex 14

Typology of in-depth interviews¹

In-depth interviewing entails asking questions, listening to and recording the answers, and then posing additional questions to clarify or expand on a particular issue. Questions are open-ended and respondents are encouraged to express their own perceptions in their own words. In-depth interviewing aims at understanding the beneficiaries' view of a programme, their terminology and judgements.

There are three basic approaches to in-depth interviewing which differ mainly in the extent to which the interview questions are determined and standardized beforehand: the informal conversational interview, the semi-structured interview and the standardized open-ended interview. Each approach serves a different purpose and has different preparation and instrumentation requirements.

Informal conversational interviews rely primarily on the spontaneous generation of questions in the natural flow of an interaction. This type of interview is appropriate when the evaluator wants to maintain maximum flexibility to pursue questioning in whatever direction appears to be appropriate, depending on the information that emerges from observing a particular setting or from talking to one or more individuals in that setting. Under these circumstances, it is not possible to have a predetermined set of questions. The strength of this approach is that the interviewer is flexible and highly responsive to individual differences, situational changes and emerging new information. The weakness is that it may generate less systematic data that are difficult and time-consuming to classify and analyse.

Semi-structured interviews involve the preparation of an interview guide that lists a predetermined set of questions or issues that are to be explored during an interview. This guide serves as a checklist during the interview and ensures that basically the same information is obtained from a number of people. Yet, there is a great deal of flexibility. The order and the actual working of the questions are not determined in advance. Moreover, within the list of topic or subject areas, the interviewer is free to pursue certain questions in greater depth. The advantage of this approach is that it makes interviewing of a number of different persons more systematic and comprehensive by delimiting the issues to be taken up in

¹ Reproduced from: *Qualitative methods*. Washington, DC, World Bank, 2011 (<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20190070~menuPK:412148~pagePK:148956~piPK:216618~theSitePK:384329,00.html>, accessed 27 August 2013).

the interview. Logical gaps in the data collected can be anticipated and closed, while the interviews remain fairly conversational and situational. The weakness is that it does not permit the interviewer to pursue topics or issues of interest that were not anticipated when the interview guide was elaborated. Also, interviewer flexibility in wording and sequencing questions may result in substantially different responses from different persons, thus reducing comparability.

Structured/standardized open-ended interviews consist of a set of open-ended questions carefully worded and arranged in advance. The interviewer asks each respondent the same questions with essentially the same words and in the same sequence. This type of interview may be particularly appropriate when there are several interviewers and the evaluator wants to minimize the variation in the questions they pose. It is also useful when it is desirable to have the same information from each interviewee at several points in time or when there are time constraints for data collection and analysis. Standardized open-ended interviews allow the evaluator to collect detailed data systematically and facilitate comparability among all respondents. The weakness of this approach is that it does not permit the interviewer to pursue topics or issues that were not anticipated when the interview instrument was elaborated. Also, standardized open-ended interviews limit the use of alternative lines of questioning with different people according to their particular experiences. This reduces the extent to which individual differences and circumstances can be fully incorporated in the evaluation. A particular case is the purpose-developed telephone survey using structured questionnaires.

Interviews with individual respondents

A common type of individual respondent interview is the *key informant interview*. A key informant is an individual who, as a result of his or her knowledge, previous experience or social status in a community, has access to information that is valuable for the evaluator – such as insights about the functioning of society, its problems and needs. Key informants are a source of information that can assist in understanding the context of a programme or project, or clarifying particular issues or problems. However, since the selection of key informants is not random, the issue of bias always arises. Another difficulty of this method lies in separating the informants' potential partiality to form a balanced view of the situation.

Group interviews

Interviews with a group of individuals can take many different forms depending on the purpose they serve, the structure of the questions, the role of the interviewer and the circumstances under which the group is convened. Some of the group interview types relevant to evaluation are: focus groups, community interviews and spontaneous group interviews.

Focus group interviews are interviews with small groups of relatively homogeneous people with similar background and experience. Participants are asked to reflect on the questions asked by the interviewers, provide their own comments, listen to what others in the group have to say and react to their observations. The main purpose is to elicit ideas, insights and experiences in a social context where people stimulate each other and consider their own views along with the views of others. Typically, these interviews are conducted several times with different groups so that the evaluator can identify trends in the perceptions and opinions expressed. The interviewer acts as a facilitator, introducing the subject, guiding the discussion, cross-checking participants' comments and encouraging all members to express their opinions. One of the main advantages of this technique is that participant interaction helps identify false or extreme views, thus providing a quality control mechanism. However, a skilful facilitator is required to ensure balanced participation of all members.

Community interviews are conducted as public meetings in which the whole community is consulted. Typically, these interviews involve a set of factually based fairly closed-ended questions. Once the interviewers pose the question, the group will interact to obtain consensus around an answer. Interviewing the community as a whole can provide valuable information on how well a project is working. The major weakness of this method is that participation may be limited to a few high-status members of the community or that community leaders may use the forum to seek consensus on their own views and preferences.

Annex 15

Checklist for evaluation reports¹

WHO has developed a checklist to ensure that the final product of the evaluation – the evaluation report – meets the expected quality based on UNEG guidance. The checklist should also be shared as part of the terms of reference prior to the conduct of the evaluation, or after the report is finalized, to assess its quality.

Reference	Item	Yes/No	Comments
1. Report structure			
The report should be logically structured with clarity and coherence	a. Is the report well structured, logical, clear and complete (i.e. executive summary, introduction/background, methods, findings, conclusions, lessons learnt, recommendations, annexes)?		
	b. Is there key basic information in the title page and opening pages: <ul style="list-style-type: none"> • name of the evaluation object • timeframe of the evaluation and date of the report • location (country, region) of the evaluation object • names and/or organizations of evaluators • name of the organization commissioning the evaluation • table of contents, which also lists tables, graphs, figures and annexes • list of acronyms? 		

¹ Adapted from UNEG (2010). *UNEG quality checklist for evaluation reports*. New York, NY, United Nations Evaluation Group (http://www.uneval.org/documentdownload?doc_id=1409&file_id=1851, accessed 28 February 2013).

Table *continued*

Reference	Item	Yes/No	Comments
	c. Is there an executive summary that includes: <ul style="list-style-type: none"> • background to the evaluation • evaluation objectives and intended audience • evaluation methodology • most important findings and conclusions • main limitations • main recommendations? 		
2. Object of evaluation^a			
The report should present a clear and full description of the object of the evaluation	a. Is the logic model and/or the expected results chain (inputs, outputs and outcomes) of the object clearly described?		
	b. Is the context of key social, political, economic, demographic and institutional factors that have a direct bearing on the object described?		
	c. Are the scale and complexity of the object of the evaluation clearly described, including for example: <ul style="list-style-type: none"> • the number of components, if more than one, and the size of the population that each component is intended to serve, both directly and indirectly; • the geographical context and boundaries (such as the region, country, and/or landscape and challenges where relevant); • the purpose and goal, and organization/management of the object; 		

^a The "object" of the evaluation is the intervention (outcome, programme, project, group of projects, themes, soft assistance) that is the focus of the evaluation.

Table *continued*

Reference	Item	Yes/No	Comments
	<ul style="list-style-type: none"> • the total resources from all sources, including human resources and budget(s) (e.g. concerned agency, partner government and other donor contributions); • the implementation status of the object, including its phase of implementation and any significant changes (e.g. plans, strategies, logical frameworks) that have occurred over time and the implications of those changes for the evaluation? 		
	d. Are the key stakeholders involved in the object implementation identified, including the implementing agency(s), partners and other key stakeholders, and their roles described?		
3. Purpose			
The purpose, objectives and scope of the evaluation should be fully explained	a. Is the purpose of the evaluation clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information was needed, how the information will be used?		
	b. Does the report provide a clear explanation of the evaluation objectives and scope, including the main evaluation questions?		
	c. Does the report describe and justify what the evaluation did and did not cover?		
	d. Does the report describe and provide an explanation of the chosen evaluation criteria and performance standards?		

Table *continued*

Reference	Item	Yes/No	Comments
4. Methodology			
The report should present a transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve the stated purpose	a. Does the report describe the data collection methods and analysis, the rationale for selecting them and their limitations?		
	b. Are reference indicators and benchmarks included where relevant?		
	c. Does the report describe the data sources, the rationale for their selection and their limitations?		
	d. Does the report include discussion of how the mix of data sources was used to obtain a diversity of perspectives and ensure data accuracy and overcome data limits?		
	e. Does the report present evidence that adequate measures were taken to ensure data quality, including evidence supporting the reliability and validity of data collection tools (e.g. interview protocols, observation tools)?		
	f. Does the report describe the sampling frame: area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects and limitations of the sample?		
	g. Does the report give a complete description of the consultation process with stakeholders in the evaluation, including the rationale for selecting the particular level and activities for consultation?		

Table *continued*

Reference	Item	Yes/No	Comments
	h. Are the methods employed appropriate for the evaluation and for answering its questions?		
	i. Are the methods employed appropriate for analysing gender, equity and human rights issues identified in the evaluation scope?		
5. Findings			
Findings should respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report and are based on evidence derived from data collection and analysis methods described in the methodology section of the report	a. Do the reported findings reflect systematic and appropriate analysis and interpretation of the data?		
	b. Do the reported findings address the evaluation criteria (such as efficiency, effectiveness, sustainability, impact and relevance) and the questions defined in the evaluation scope?		
	c. Are the findings objectively reported based on the evidence?		
	d. Are gaps and limitations in the data reported and discussed?		
	e. Are unanticipated findings reported and discussed?		
	f. Are reasons for accomplishments and failures, especially continuing constraints, identified as far as possible?		
	g. Are overall findings presented with clarity, logic and coherence?		

Table *continued*

Reference	Item	Yes/No	Comments
6. Conclusions			
Conclusions should present reasonable judgements based on findings and sustained by evidence, and should provide insights pertinent to the object and purpose of the evaluation	a. Do the conclusions reflect reasonable evaluative judgements relating to key evaluation questions?		
	b. Are the conclusions well substantiated by the evidence presented?		
	c. Are the conclusions logically connected to evaluation findings?		
	d. Do conclusions provide insights into the identification of and/or solutions to important problems or issues pertinent to the prospective decisions and actions of evaluation users?		
	e. If applicable to the evaluation objectives, do the conclusions present the strengths and weaknesses of the object (policy, programmes, projects or other interventions) being evaluated, on the basis of the evidence presented and taking due account of the views of a diverse cross-section of stakeholders?		
7. Lessons			
Lessons should present remarks with potential for wider application and use	a. Are the lessons drawn from experience (achievements, problems, mistakes)?		
	b. Is the context in which the lessons may be applied clearly specified?		

Table *continued*

Reference	Item	Yes/No	Comments
8. Recommendations			
Recommendations should be relevant to the object and purposes of the evaluation, supported by evidence and conclusions, and developed with the involvement of relevant stakeholders	a. Does the report describe the process followed in developing the recommendations, including consultation with stakeholders?		
	b. Are the recommendations firmly based on evidence and conclusions?		
	c. Are the recommendations relevant to the object and purposes of the evaluation?		
	d. Do the recommendations clearly identify the target group of each recommendation?		
	e. Are the recommendations clearly stated with priorities for action made clear?		
	f. Are the recommendations actionable and do they reflect an understanding of the commissioning organization and potential constraints to implementation?		
	g. Do the recommendations include an implementation plan?		
9. Gender, equity and human rights			
The report should illustrate the extent to which the design and implementation of the object, the assessment of results and the evaluation process incorporate a gender, equity and human rights-based approach	a. Do the evaluation objectives and scope include questions that address issues of gender and human rights, as appropriate?		
	b. Does the report use gender-sensitive and human rights-based language throughout, including data disaggregated by sex, age, disability, etc.?		

Table *continued*

Reference	Item	Yes/No	Comments
	c. Are the evaluation approach and data collection and analysis methods appropriate for analysing the gender equity and human rights issues identified in the scope?		
	d. As well as noting the actual results on gender equality and human rights, does the report assess whether the design of the object was based on a sound gender analysis and human rights analysis, and was implementation for results monitored through gender and human rights frameworks?		
	e. Do reported findings, conclusions, recommendations and lessons provide adequate information on gender equality and human rights aspects?		
	f. Does the report consider how the recommendations may affect the different stakeholders of the object being evaluated?		

Annex 16

Glossary of key terms in evaluation¹

Accountability

Obligation to demonstrate that work has been conducted in compliance with agreed rules and standards or to report fairly and accurately on performance results vis-à-vis mandated roles and/or plans. This may require a careful, even legally defensible, demonstration that the work is consistent with the contract terms.

Note: Accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources. For evaluators, it connotes the responsibility to provide accurate, fair and credible monitoring reports and performance assessments. For public sector managers and policy-makers, accountability is to taxpayers/citizens.

Activity

Actions taken or work performed through which inputs – such as funds, technical assistance and other types of resources – are mobilized to produce specific outputs.

Related term: development intervention.

Analytical tools

Methods used to process and interpret information during an evaluation.

Appraisal

An overall assessment of the relevance, feasibility and potential sustainability of a development intervention prior to a decision on funding.

Note: In development agencies, banks, etc., the purpose of appraisal is to enable decision-makers to decide whether the activity represents an appropriate use of corporate resources.

Related term: ex-ante evaluation.

¹ Based on: *Glossary of key terms in evaluation and results-based management*. Paris, Organisation for Economic Co-operation and Development, Development Assistance Committee, 2010 (available at <http://www.oecd.org/dac/evaluation/18074294.pdf>).

Assumptions

Hypotheses about factors or risks that could affect the progress or success of a development intervention.

Note: Assumptions can also be understood as hypothesized conditions that bear on the validity of the evaluation itself (e.g. relating to the characteristics of the population when designing a sampling procedure for a survey). Assumptions are made explicit in theory-based evaluations where evaluation systematically tracks the anticipated results chain.

Attribution

The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.

Note: Attribution refers to that which is to be credited for the observed changes or results achieved. It represents the extent to which observed development effects can be attributed to a specific intervention or to the performance of one or more partners taking account of other interventions (anticipated or unanticipated), confounding factors or external shocks.

Audit

An independent, objective assurance activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to assess and improve the effectiveness of risk management, control and governance processes.

Note: A distinction is made between regularity (financial) auditing which focuses on compliance with applicable statutes and regulations, and performance auditing which is concerned with relevance, economy, efficiency and effectiveness. Internal auditing provides an assessment of internal controls undertaken by a unit reporting to management, while external auditing is conducted by an independent organization.

Baseline study

An analysis describing the situation prior to a development intervention, against which progress can be assessed or comparisons made.

Benchmark

Reference point or standard against which performance or achievements can be assessed.

Note: A benchmark refers to the performance that has been achieved in the recent past by other comparable organizations or what can be reasonably inferred to have been achieved in the circumstances.

Beneficiaries

The individuals, groups or organizations, whether targeted or not, that benefit, directly or indirectly, from the development intervention.

Related terms: reach, target group.

Cluster evaluation

An evaluation of a set of related activities, projects and/or programmes.

Conclusions

Conclusions point out the factors of success and failure of the evaluated intervention, with special attention paid to the intended and unintended results and impacts, and more generally to any other strength or weakness. A conclusion draws on data collection and analyses undertaken, through a transparent chain of arguments.

Counterfactual

The situation or condition that hypothetically may prevail for individuals, organizations or groups if there were there no development intervention.

Country programme evaluation/country assistance evaluation

Evaluation of one or more donors' or agencies' portfolios of development interventions, and the assistance strategy behind them, in a partner country.

Data collection tools

Methodologies used to identify information sources and collect information during an evaluation.

Examples include informal and formal surveys, direct and participatory observation, community interviews, focus groups, expert opinion, case-studies and literature searches.

Development intervention

An instrument for partner (donor and non-donor) support aimed to promote development.

Examples include policy advice, projects and programmes.

Development objective

Intended impact contributing to physical, financial, institutional, social, environmental or other benefits to a society, community or group of people via one or more development interventions.

Economy

Absence of waste for a given output.

Note: An activity is economical when the costs of the scarce resources used approximate to the minimum needed to achieve planned objectives.

Effect

Intended or unintended change due directly or indirectly to an intervention.

Related terms: results, outcome.

Effectiveness

The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.

Note: Also used as an aggregate measure of (or judgement about) the merit or worth of an activity – i.e. the extent to which an intervention has attained, or is expected to attain, its major relevant objectives efficiently in a sustainable fashion and with a positive institutional development impact.

Efficiency

A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.

Evaluability

The extent to which an activity or a programme can be evaluated in a reliable and credible fashion.

Note: Evaluability assessment calls for the early review of a proposed activity in order to ascertain whether its objectives are adequately defined and its results verifiable.

Evaluation

The systematic and objective assessment of an ongoing or completed project, programme or policy, its design, implementation and results. The aim is to

determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learnt into the decision-making process of both recipients and donors.

Evaluation also refers to the process of determining the worth or significance of an activity, policy or programme. It is an assessment, as systematic and objective as possible, of a planned, ongoing or completed development intervention.

Note: Evaluation in some instances involves the definition of appropriate standards, the examination of performance against those standards, an assessment of actual and expected results and the identification of relevant lessons.

Related term: review.

Ex-ante evaluation

An evaluation that is performed before implementation of a development intervention.

Related terms: appraisal, quality at entry.

Ex-post evaluation

Evaluation of a development intervention after it has been completed.

Note: It may be undertaken directly after or long after completion. The intention is to identify the factors of success or failure, to assess the sustainability of results and impacts, and to draw conclusions that may inform other interventions.

External evaluation

The evaluation of a development intervention conducted by entities and/or individuals outside the donor and implementing organizations.

Feedback

The transmission of findings generated through the evaluation process to parties for whom it is relevant and useful in order to facilitate learning. This may involve the collection and dissemination of findings, conclusions, recommendations and lessons from experience.

Finding

A factual statement based on evidence from one or more evaluations.

Formative evaluation

An evaluation intended to improve performance, most often conducted during the implementation phase of projects or programmes.

Note: Formative evaluations may also be conducted for other reasons such as compliance, legal requirements or as part of a larger evaluation initiative.

Related term: process evaluation.

Goal

The higher-order objective to which a development intervention is intended to contribute.

Related term: development objective.

Impacts

Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

Independent evaluation

An evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention.

Note: The credibility of an evaluation depends in part on how independently it has been carried out. Independence implies freedom from political influence and organizational pressure. It is characterized by full access to information and by full autonomy in carrying out investigations and reporting findings.

Indicator

A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor.

Inputs

The financial, human and material resources used for the development intervention.

Institutional development impact

The extent to which an intervention improves or weakens the ability of a country or region to make more efficient, equitable and sustainable use of its human,



financial and natural resources, for example through: better definition, stability, transparency, enforceability and predictability of institutional arrangements; and/or better alignment of the mission and capacity of an organization with its mandate, which derives from these institutional arrangements. Such impacts can include intended and unintended effects of an action.

Internal evaluation

Evaluation of a development intervention conducted by a unit and/or individuals reporting to the management of the donor, partner or implementing organization.

Related term: self-evaluation.

Joint evaluation

An evaluation in which different donor agencies and/or partners participate.

Note: There are various degrees of “jointness” depending on the extent to which individual partners collaborate in the evaluation process, merge their evaluation resources and combine their evaluation reporting. Joint evaluations can help overcome attribution problems in assessing the effectiveness of programmes and strategies, the complementarity of efforts supported by different partners, the quality of aid coordination, etc.

Lessons learnt

Generalizations based on evaluation experiences with projects, programmes or policies that abstract from the specific circumstances to broader situations. Frequently, lessons highlight strengths or weaknesses in preparation, design and implementation that affect performance, outcome and impact.

Logical framework (logframe)

A management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.

Related term: results-based management.

Meta-evaluation

The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or to assess the performance of the evaluators.

Mid-term evaluation

Evaluation performed towards the middle of the period of implementation of the intervention.

Related term: formative evaluation.

Monitoring

A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.

Related term: performance monitoring, indicator.

Outcome

The likely or achieved short-term and medium-term effects of an intervention's outputs.

Related terms: results, outputs, impacts, effect.

Outputs

The products, capital goods and services that result from a development intervention, which may also include changes resulting from the intervention that are relevant to the achievement of outcomes.

Participatory evaluation

Evaluation method in which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation.

Partners

The individuals and/or organizations that collaborate to achieve mutually agreed objectives.

Note: The concept of partnership connotes shared goals, common responsibility for outcomes, distinct accountabilities and reciprocal obligations. Partners may include governments, civil society, nongovernmental organizations, universities, professional and business associations, multilateral organizations, private companies, etc.

Performance

The degree to which a development intervention or a development partner operates according to specific criteria/standards/guidelines or achieves results in accordance with stated goals or plans.

Performance indicator

A variable that allows the verification of changes in the development intervention or shows results relative to what was planned.

Related terms: performance monitoring, performance measurement.

Performance measurement

A system for assessing performance of development interventions against stated goals.

Related terms: performance monitoring, performance indicator.

Performance monitoring

A continuous process of collecting and analysing data to compare how well a project, programme or policy is being implemented against expected results.

Process evaluation

An evaluation of the internal dynamics of implementing organizations, their policy instruments, their service delivery mechanisms, their management practices and the linkages among these.

Related term: formative evaluation.

Programme evaluation

Evaluation of a set of interventions, marshalled to attain specific global, regional, country or sector development objectives.

Note: A development programme is a time-bound intervention involving multiple activities that may cut across sectors, themes and/or geographical areas.

Related term: country programme/strategy evaluation.

Project evaluation

Evaluation of an individual development intervention designed to achieve specific objectives within specified resources and implementation schedules, often within the framework of a broader programme.

Note: Cost–benefit analysis is a major instrument of project evaluation for projects with measurable benefits. When benefits cannot be quantified, cost–effectiveness is a suitable approach.

Project or programme objective

The intended physical, financial, institutional, social, environmental or other development results to which a project or programme is expected to contribute.

Purpose

The publicly stated objectives of the development programme or project.

Quality assurance

Quality assurance encompasses any activity that is concerned with assessing and improving the merit or the worth of a development intervention or its compliance with given standards.

Note: Examples of quality assurance activities include appraisal, results-based management, reviews during implementation, evaluations, etc. Quality assurance may also refer to the assessment of the quality of a portfolio and its development effectiveness.

Reach

The beneficiaries and other stakeholders of a development intervention.

Related term: beneficiaries.

Recommendations

Proposals aimed at enhancing the effectiveness, quality or efficiency of a development intervention, at redesigning the objectives, and/or at the reallocation of resources. Recommendations should be linked to conclusions.

Relevance

The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies.

Note: Retrospectively, the question of relevance often becomes a question of whether the objectives of an intervention or its design are still appropriate given changed circumstances.

Reliability

Consistency or dependability of data and evaluation judgements, with reference to the quality of the instruments, procedures and analyses used to collect and interpret evaluation data.

Note: Evaluation information is reliable when repeated observations using similar instruments under similar conditions produce similar results.

Results

The output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention.

Related terms: outcome, effect, impacts.

Results chain

The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts and feedback. In some agencies, reach is part of the results chain.

Related terms: assumptions, results framework.

Results framework

The programme logic that explains how the development objective is to be achieved, including causal relationships and underlying assumptions.

Related terms: results chain, logical framework.

Results-based management

A management strategy focusing on performance and achievement of outputs, outcomes and impacts.

Related term: logical framework.

Review

An assessment of the performance of an intervention, periodically or on an ad hoc basis.

Note: Frequently “evaluation” is used for a more comprehensive and/or more in-depth assessment than “review”. Reviews tend to emphasize operational aspects. Sometimes the terms “review” and “evaluation” are used synonymously.

Related term: evaluation.

Risk analysis

An analysis or an assessment of factors (called assumptions in the logframe) that affect or are likely to affect the successful achievement of an intervention's objectives. A detailed examination of the potential unwanted and negative consequences to human life, health, property or the environment posed by development interventions; a systematic process to provide information regarding such undesirable consequences; and/or the process of quantification of the probabilities and expected impacts for identified risks.

Sector programme evaluation

Evaluation of a cluster of development interventions in a sector within one country or across countries, all of which contribute to the achievement of a specific development goal.

Note: A sector includes development activities commonly grouped together for the purpose of public action such as health, education, agriculture, transport.

Self-evaluation

An evaluation by those who are entrusted with the design and delivery of a development intervention.

Stakeholders

Agencies, organizations, groups or individuals who have a direct or indirect interest in the development intervention or its evaluation.

Summative evaluation

A study conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of the programme. Summative evaluations are also referred to as impact evaluations.

Sustainability

The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.

Target group

The specific individuals or organizations for whose benefit the development intervention is undertaken.

Terms of reference

Written document presenting the purpose and scope of the evaluation, the methods to be used, the standard against which performance is to be assessed or analyses are to be conducted, the resources and time allocated, and reporting requirements. Two other expressions sometimes used with the same meaning are “scope of work” and “evaluation mandate”.

Thematic evaluation

The evaluation of a selection of development interventions, all of which address a specific development priority that cuts across countries, regions and sectors.

Triangulation

The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment.

Note: By combining multiple data sources, methods, analyses or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observers or single theory studies.

Validity

The extent to which the data collection strategies and instruments measure what they purport to measure.



9 789241 548687