USAID
FROM THE AMERICAN PEOPLE

HEALTH SYSTEMS 20/20
better systems. better health

FRAMEWORK AND GUIDELINE FOR THE ASSESSMENT AND EVALUATION OF HEALTH SYSTEMS STRENGTHENING PROGRAMS

This document was developed with inputs from a broad group of technical experts as an initial step to support the Global Fund in evaluating its health system strengthening (HSS) grants. The Framework was piloted for evaluating an HSS national strategy in South Sudan in July-August 2012. The Global Fund, in collaboration with technical partners, is in the process of revising its investment framework in health and community systems strengthening. Upon finalizing the investment framework, the present evaluation framework and guideline will be revised to ensure necessary consistency between the investment and evaluation frameworks.

The USAID-funded Health Systems 20/20 project provided technical support for development of this document; however, the views expressed do not necessarily represent those of USAID, nor has it endorsed this approach for use in evaluating USAID-supported HSS programs.

September 2012

This publication was produced for review by the United States Agency for International Development. It was prepared by Itamar Katz, Grace Chee, Alexandra Hulme, and Sayaka Koseki the for Health Systems 20/20 Project.

Health Systems 20/20 is USAID's flagship project for strengthening health systems worldwide. By supporting countries to improve their health financing, governance, operations, and institutional capacities, Health Systems 20/20 helps eliminate barriers to the delivery and use of priority health care, such as HIV/AIDS services, tuberculosis treatment, reproductive health services, and maternal and child health care.

September 2012

For additional copies of this report, please email info@healthsystems2020.org or visit our website at www.healthsystems2020.org

Cooperative Agreement No.: GHS-A-00-06-00010-00

Submitted to: Scott Stewart, AOTR
Health Systems Division
Office of Health, Infectious Disease and Nutrition
Bureau for Global Health
United States Agency for International Development

Recommended Citation: Katz, Itamar, Grace Chee, Alexandra Hulme and Sayaka Koseki. September 2012. *Framework and Guideline for the Assessment and Evaluation of Health Systems Strengthening Program*. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.

Abt Associates Inc. | 4550 Montgomery Avenue | Suite 800 North
| Bethesda, Maryland 20814 | P: 301.347.5000 | F: 301.913.9061
| www.healthsystems2020.org | www.abtassociates.com

In collaboration with:
| Aga Khan Foundation | Bitrán y Asociados | BRAC University | Broad Branch Associates
| Deloitte Consulting, LLP | Forum One Communications | RTI International
| Training Resources Group | Tulane University School of Public Health and Tropical Medicine

FRAMEWORK AND GUIDELINE FOR THE ASSESSMENT AND EVALUATION OF HEALTH SYSTEMS STRENGTHENING PROGRAMS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ART	Antiretroviral Therapy
DHS	Demographic Health Survey
DOTS	Directly Observed Therapy, Short-Course
HIS	Health Information Systems
HIV	Human Immunodeficiency Virus
HRH	Human Resources for Health
HSS	Health Systems Strengthening
M&E	Monitoring and Evaluation
MESST	Monitoring and Evaluation Systems Strengthening Tool
MoH	Ministry of Health
NSP	National Strategic Plans
OSDV	On-site Data Verification
RHM	Rural Health Motivator
RSQA	Rapid Services Quality Assessment
SDA	Service Delivery Area
TB	Tuberculosis
ToR	Terms of Reference
UNGASS	United Nations General Assembly Special Session (on HIV/AIDS)
USAID	United States Agency for International Development
WHO	World Health Organization

PREFACE

The framework and guideline for the assessment and evaluation of health systems strengthening (HSS) programs are designed to provide insight into the effects of ongoing interventions that aim to strengthen specific elements of a health system.

This document was developed to support the Global Fund to Fight AIDS, Tuberculosis and Malaria (the Global Fund) in assessing its HSS grants, yet it is designed also to effectively assess non-Global Fund supported programs. The present framework and guideline were piloted for evaluating HSS investments in South Sudan in July and August 2012. The Global Fund, in collaboration with technical partners, is in the process of revising its investment framework in health and community systems strengthening. Upon finalizing the investment framework, the present framework and guideline for HSS program evaluation will be revised to ensure necessary consistency between the investment and evaluation frameworks.

The framework evaluates HSS programs by assessing four major areas:

1. Strengths and challenges of the management of the HSS program. This assessment addresses questions such as the following:
    a. How well is the program performing programmatically and financially?
    b. Does the program have the capacity to manage the implementation of the program's interventions in an effective and efficient manner?
    c. Is the monitoring and evaluation (M&E) system sufficiently robust?
2. Effects of the HSS program on system and health outcomes by evaluating the performance of the outcome indicators targeted in the program's results framework. Importantly, a results framework encompasses what the HSS program was designed to achieve, and can be part of a national strategy plan, an operational plan, or a grant agreement. This assessment is done by determining whether or not the outcome indicators were achievable in light of the interventions prescribed by the HSS program, and if so, whether the program's goals were achieved. This assessment answers questions such as the following:
    a. How did the systems and health outcome indicators of results framework perform? Were there any potential confounding factors?
    b. Are the systems and health outcome indicators within the results framework relevant to the interventions of the HSS program? If not, should the outcome indicator be modified and/or the HSS program reprogrammed?
    c. Are the data sources for measuring the performance of the outcome indicators adequate in term of timeliness, relevance, and robustness?
3. Any positive or negative effects the HSS program may have had on the health system. This assessment aims to determine whether the HSS program strengthened or weakened an array of health system elements relevant to the program.
4. Whether or not the HSS program's intervention strategies need to be modified due to changes in the country's HSS environment. This assessment uses a formative evaluation, which determines whether the HSS program requires reprogramming due to evolution of the country's health sector priorities, the availability of new technical guidelines, and/or the implementation of new donor programs that may require harmonization with the HSS program under evaluation.

This approach is recommendation oriented and can provide concrete recommendations on the following:

1. How to reprogram the HSS interventions to better respond to the aim of the results framework, mitigate negative effects, strengthen positive effects, and adjust to new developments.
2. How to strengthen the results framework to measure more accurately the HSS interventions and their desired effects.

The approach is intended to be implemented over a period of several weeks by a small group of HSS and M&E experts and data collectors. It includes desk reviews, interviews with a wide range of key informants, and visits to a small sample of delivery sites. To reduce subjectivity, the evaluation triangulates findings from different data sources. Furthermore, for certain elements of the evaluation, it requires input from the M&E officer of the institution commissioning the evaluation. However, it must be noted that the framework's evaluation is not experimental in design, and though it provides insight into HSS programs, it is not intended to identify causal linkages or prove attribution between interventions and observed effects.

The framework and guideline were developed with input from a broad group of technical experts. Some elements of the methodology have been tested in the Health System Assessment Approach and with the HIV/AIDS Program Sustainability Analysis Tool (HAPSAT), the former being implemented in over 20 countries, and the latter in 14 countries. In addition, the approach borrows questions from the Program Management Capacity Assessment Tool of the Global Fund.

ACKNOWLEDGMENTS

This document has been developed as part of collaboration between the Global Fund to Fight AIDS, Tuberculosis and Malaria and Health Systems 20/20, the USAID flagship project for strengthening health systems worldwide. USAID funded this activity.

Many people must be thanked for their contribution to this activity. They provided their insight through participation in a one-day consultation in December 2011, through interviews, and/or through providing feedback on the draft versions of this document: Ricardo Bitrán of Bitrán y Asociados; Ekhard Kleinau of CAMRIS; Margaret Kurk of Columbia University; Michael Hansen of GAVI Alliance; Eric Friedman of Georgetown University; George Shakarishvili, Perry Mwangala, and George Sakvarelidze of the Global Fund to Fight AIDS, Tuberculosis and Malaria; Wim Van Damme of the Institute of Tropical Medicine in Antwerp; Soumya Alva of ICF International; Peter Smith of Imperial College; Freddy Perez of PAHO; Linda Spink of TRG; Bob Emrey, Estelle Quain, Jodi Charles, Mai Hijazi, and Scott Stewart of USAID; Sarah Barber, Gerard Schmets, and Ties Boerma of World Health Organization; Gyuri Fritsche of the World Bank; and Ann Lion, Catherine Connor, Dawn Crosby, Jordan Tuchman, Julie Williams, Lisa Tarantino, Mursaleena Islam, Nancy Pielemeier, Nejmudin Bilal, Nicole Barcikowski, and Slavea Chankova of Abt Associates.

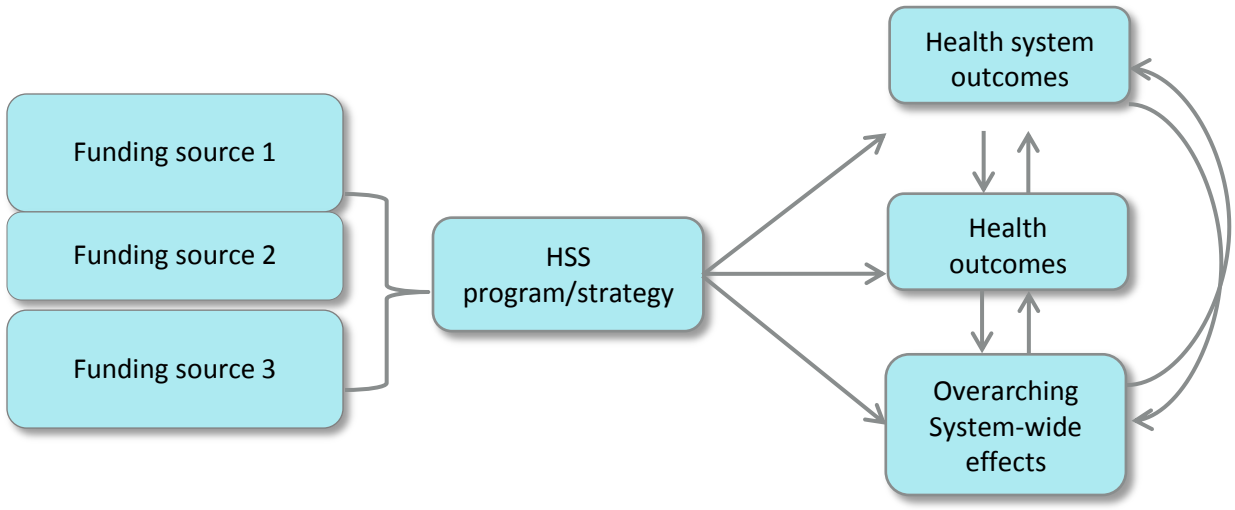
1. INTRODUCTION

A strong health system is a necessary prerequisite for maximizing the impact of disease control programs. As countries recognize the importance of well-performing health systems for improving health outcomes, both the demand for health systems strengthening (HSS) funding and the level of resource allocation by international donors have significantly increased in the last few years. With the scale up of HSS investments, the focus on evaluating and measuring the effect of HSS interventions on health and health systems outcomes has increased.

The evaluation framework and accompanying guideline detailed here provide a standardized methodology to assess the performance of HSS programs. This document describes the steps to conduct formative and process evaluations of HSS programs, as well as how to assess the system-wide effects of such programs on elements of the health system, which are beyond the scope of the HSS program, and the plausible effects that influence health outcomes, as defined in the program's results framework. In this document, the term HSS program refers to a compilation of interventions that aim to improve the capacity and performance of major functions of the health system, including service delivery, financing, governance, health information systems, human resources, and the supply chain management system. Such interventions typically aim to improve the health system's performance to benefit outcomes related to multiple diseases or conditions.

HSS programs are frequently financed by more than one funding source and are aligned with the goals of the country's national strategic plans (NSP). HSS programs target specific health and health system outcomes defined in their results framework, and are likely to have positive and/or negative overarching system-wide effects on sustainability, equity, and efficiency of the health system, which may facilitate or hinder the overall performance of the health system. As shown in Figure 1, the HSS strategy, system and health outcomes and system-wide effects are interlinked.

FIGURE 1: PATHWAY AND LINKAGES BETWEEN FUNDING SOURCES AND HSS PROGRAM, AND PROGRAM OUTCOMES AND EFFECTS



The approach presented in this document outline the assessment and evaluation purpose, methodology, and process for measuring HSS program performance, based on both quantitative and qualitative data, and culminate in recommendations for actionable improvements to the HSS program. Part A (chapters 2 through 6) describe the framework, while Part B (chapter 7) presents a guideline on the various steps in the implementation of the framework, including planning and preparation, data collection, analysis, and report writing. This document is supplemented by a preformatted Excel file to assist evaluators in the analysis of the data.

1.1 OBJECTIVES

This HSS program assessment and evaluation framework and the accompanying guideline outline the evaluation methods and the process for assessing the following areas:

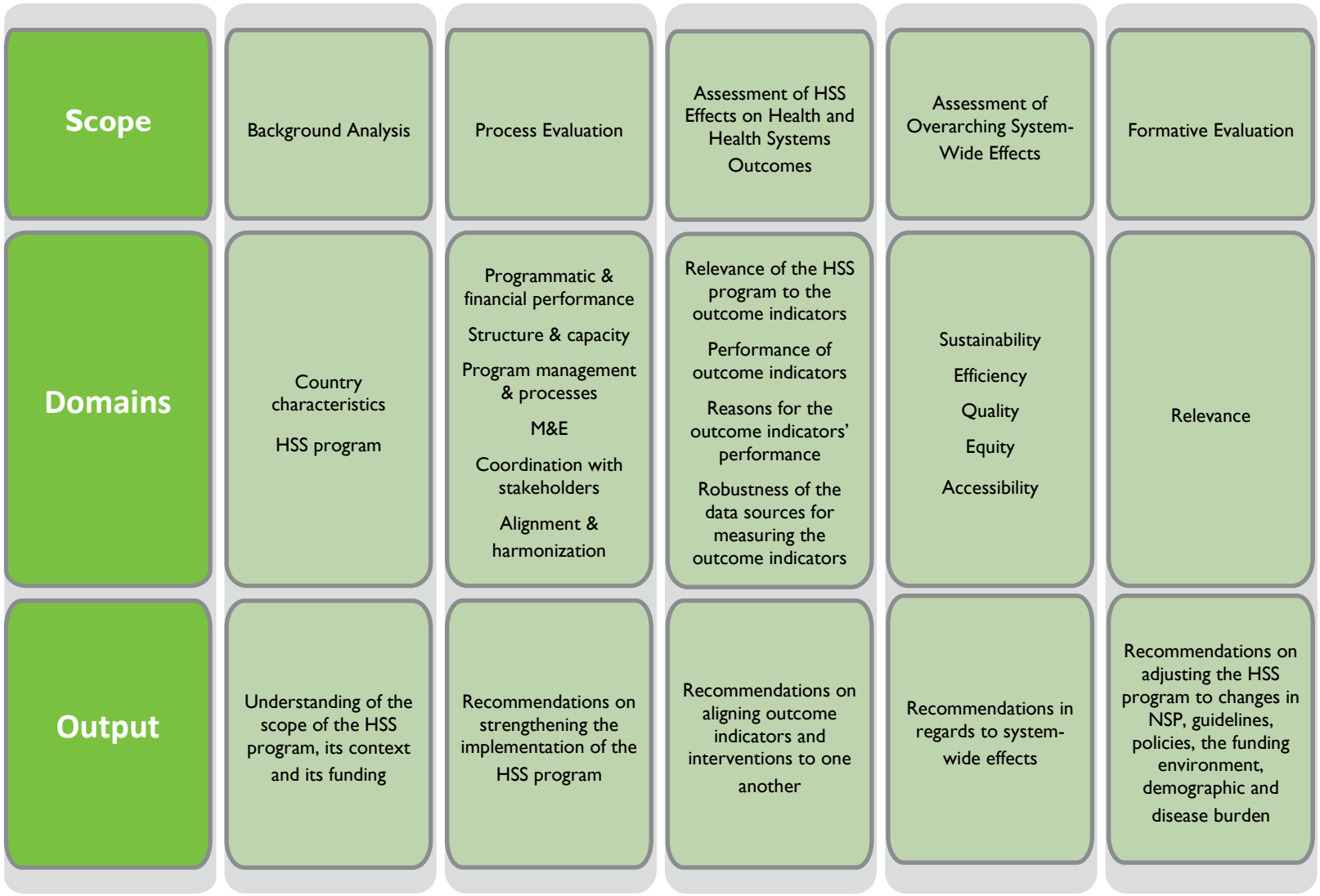
- **Capacity of the implementing entity to efficiently and effectively implement the programs.** The process evaluation examines financial and programmatic performance, examines management and implementation issues, and identifies challenges that interfere with effective and efficient program delivery.
- **Program's effect on health and health systems outcomes.** The assessment of the extent to which the HSS programs could plausibly have had an effect on the outcome indicators of the results framework of the HSS program, by examining the pathway, scale, and timing of the interventions in relation to the outcome indicators. The performance of those outcome indicators is further reviewed and then assessed against the quality of the HSS interventions and potential confounding factors. Recommendations are formulated to enhance the interventions and/or to modify the results framework, to better adjust the outcome indicators to the interventions.
- **Overarching system-wide effect of the HSS program.** The assessment of the system-wide effects on sustainability, equity, and efficiency examines a range of potential effects (positive or negative) that an HSS program might have had on the broader health system, beyond the immediate scope of the HSS program in question.
- **Relevance of the HSS program.** The formative evaluation assesses the compatibility of specific HSS interventions against changes in NSP, normative policies and guidelines, funding environment, and the disease and demographic burden.

This approach also includes instructions on the process to develop recommendations for improving the design and implementation of the HSS program, based on the findings from the evaluation.

1.2 CONCEPTUAL FRAMEWORK

The evaluation framework, summarized in Figure 2, is organized into five sections: background analysis, process evaluation, assessment of HSS effect on health and health systems outcomes, assessment of overarching system-wide effects, and formative evaluation.

FIGURE 2: FRAMEWORK FOR HSS PROGRAM EVALUATION



1.2.1 BACKGROUND ANALYSIS

Definition: The background analysis is designed for scoping and understanding the assessed HSS program in its country context.

Dimensions: The background analysis will look at the following:

1. Country characteristics.
2. HSS program. Within this element, the analysis will further examine funding sources of the HSS program, its share of the total health expenditure, and its scope.

Methodology: Desk review of HSS program and country-related documentation.

Output: Summary of the HSS program, its funding, and its rationale/objective in the country context.

1.2.2 PROCESS EVALUATION

Definition: Process evaluation examines program performance and management and implementation issues, and identifies challenges that interfere with effective and efficient program delivery.

Dimensions: The questions on the process evaluation are categorized into the following domains:

1. Summary of existing reviews of the programmatic and financial performance, data quality, and overall performance.
2. Structure and capacity of the program-implementing entity.
3. Program management systems and processes.
4. Monitoring and evaluation (M&E).
5. Coordination with stakeholders.
6. Alignment and harmonization.

Methodology: Triangulation of information obtained from HSS program documentation and interviews with key informants.

Outputs: An analysis of program performance, implementation processes and challenges, differences between program plans/design and implementation, achievement of output targets (e.g., number of health workers trained through the program), and the strengths and weaknesses of the program management entity to manage the HSS program effectively, as well as recommendations to enhance the program design and implementation.

1.2.3 ASSESSMENT OF THE HSS PROGRAM EFFECTS ON HEALTH AND HEALTH SYSTEM OUTCOMES

Definition: Applying rigorous impact evaluation methods to measure the impact on health outcome indicators directly attributable to HSS program interventions is not feasible in the context of most of the Global Fund to Fight AIDS, Tuberculosis and Malaria (the Global Fund) HSS programs and within a rapid assessment. Therefore, the evaluation suggested in this framework is to assess broadly the relevance of the outcome indicators of the HSS programs to its interventions and then measure the performance of the relevant outcome indicators against their targets in the results framework, followed by a review of potential confounding factors and potential data quality issues.

Dimensions: This part of the evaluation will look at the following aspects:

1. Extent to which the relevance, in terms of logic, scope, and timing, of HSS program interventions can plausibly affect the outcome indicators.
2. Changes in the performance of the outcome indicators during the HSS program period.
3. Quality of the HSS program and confounding factors that might have affected the changes in performance of the outcome indicators.
4. Robustness of the data sources for measuring the outcome indicators.

Methodology: Qualitative assessment of the plausible effects of the interventions of the HSS program on health system and health outcome indicators in the results framework. The qualitative assessment includes a review of documentations, interviews with key informants, and site visits. This is followed by a quantitative calculation of the performance of the systems and health outcome indicators, and a qualitative assessment of the performance and the quality of the interventions and the confounding factors.

Outputs: The following are the outputs of this assessment component:

1. Analysis of the plausible effects: The evaluation team will conduct an analysis of the plausible effects of the interventions of the HSS program on the systems and health outcome indicators.
2. Assessment of the HSS program design: Per outcome indicator, the evaluation team will ask whether there are any relevant interventions in the HSS program, and, if not, should the outcome indicator be modified or should the HSS program be reprogrammed.
3. Assessment of the measurements: The evaluator will provide an assessment on the ability of the results framework to measure the outcome. Per outcome indicator, the timeliness, relevance, and robustness of the data source will be reviewed, as detailed above.

1.2.4 ASSESSMENT OF OVERARCHING SYSTEM-WIDE EFFECTS

Definition: The assessment of system-wide effects is an assessment of the positive and negative influences of the HSS program on the overarching constituents of the health system, such as equity, efficiency, and sustainability.

Dimensions: The questions concerning the system-wide effect of the HSS program are categorized into the following domains, which align with the criteria widely used for health system assessment approaches:

1. **Equity:** The degree to which HSS program resources and services are delivered to beneficiaries according to their needs, regardless of social, economic, geographic, gender, demographic, or other characteristics.
2. **Quality:** The degree to which services are implemented according to internationally and nationally recognized and evidence-based technical policies and guidelines[1].
3. **Efficiency:** The degree to which the health system maximizes outcomes relative to its level of resources or achieves a given program outcome with a minimum of resources[2]. Within the

[1] The Global Fund to Fight AIDS Tuberculosis and Malaria (2011). Local Fund Agent guidelines for on-site data verification (OSDV) and rapid services quality assessment (RSQA) implementation http://www.theglobalfund.org/en/me/documents/MEQualityServices/. Geneva.
[2] WHO (2010). Monitoring the Building Blocks for Health Systems: A Handbook of Indicators and their Measurement Strategies. Geneva, WHO.

context of this framework and guideline, efficiency is translated into the elimination of parallel systems by avoiding duplication of activities, increasing harmonization of activities, and developing integrated approaches.

4. **Accessibility:** The degree to which the existence or absence of financial, cultural, geographic, or other barriers determines how beneficiaries are able to utilize services supported by the HSS program(s).
5. **Sustainability:** The degree to which the health system is able to maintain programmatic outcomes and impacts beyond the completion of grant-supported activities. The focus here is on institutional sustainability, which is the capacity of the health system, if suitably financed, to assemble and manage the necessary nonfinancial resources to successfully carry on its normal activities in the future[3].

Methodology: Triangulation of information obtained from HSS program documentation and interviews with key informants.

Output: An analysis of the positive and negative effects of the program on strengthening the health system, the effects and recommendations in regards to system-wide effects, per each relevant domain.

1.2.5 FORMATIVE EVALUATION

Definition: The formative evaluation assesses the technical content of the HSS program (i.e., objectives and activities) and the relevance of the program as planned given changes in policies, national priorities, and epidemiological profile that might have occurred after the program planning stage. The periodic revision of objectives and activities is desirable given the dynamics of the broader context within which the HSS programs are implemented (e.g., changes in the country's health sector priorities, changes in the country's health policy, availability of new technical guidelines, initiation of other donor programs that may require harmonization – all of which may occur during the lifecycle of the program). Through this review, programmatic areas that may require revisions are identified.

Dimensions: The formative evaluation will formulate recommendations to improve the relevance of the HSS program. Relevance is the degree to which program goals, HSS program objectives, and implemented activities correspond to the country's epidemiologic, demographic, policy, economic, and social context. During the lifecycle of an HSS program the context may change, and, therefore, goals, objectives, and activities designed at the time of program inception may need to be revisited.

Methodology: Triangulation of information obtained from HSS program documentation and interviews with key informants.

Output: Recommendations on reprogramming HSS program objectives and activities.

[3] Health Systems 20/20 (2012). Health Systems Assessment Approach: A How-To Manual Version 2.0. Bethesda, MD, Abt Associates Inc.

1.3 THE GUIDELINE

Part B of this document elaborate on the guideline, namely the phases and steps involved in executing this approach. There are four phases including preparation and initiation, data collection and analysis, formulation of recommendations and evaluation completion, composed of 11 steps (Figure 3).

FIGURE 3: HSS PROGRAM EVALUATION'S PHASES AND STEPS

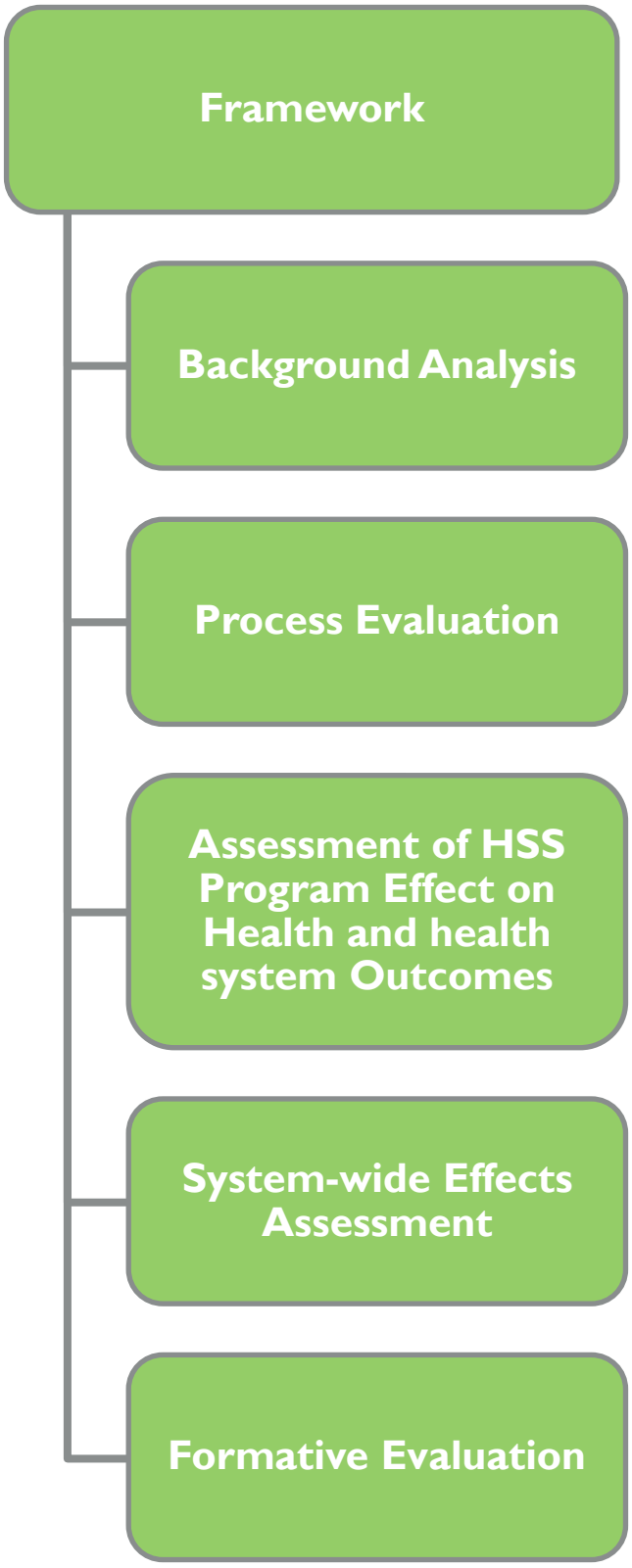
Phase 1	Phase 2	Phase 3	Phase 4
Preparation and initiation	**Data collection and analysis**	**Formulation of recommendations**	**Evaluation completion**
1. Assemble evaluation team and data collectors	5. Collect data	7. Present preliminary findings to stakeholders and formulate recommendations	8. Draft evaluation report
2. Notify program and obtain national authorizations	6. Analyze the findings		9. Review and collect feedback from country and commissioning organization
3. Request documentation			10. Finalize evaluation report
4. Formulate data collection plan			11. Initiate follow-up of recommended actions

1.4 DELIVERABLES

In conducting the evaluation, the evaluators will submit the following to the commissioning organization:

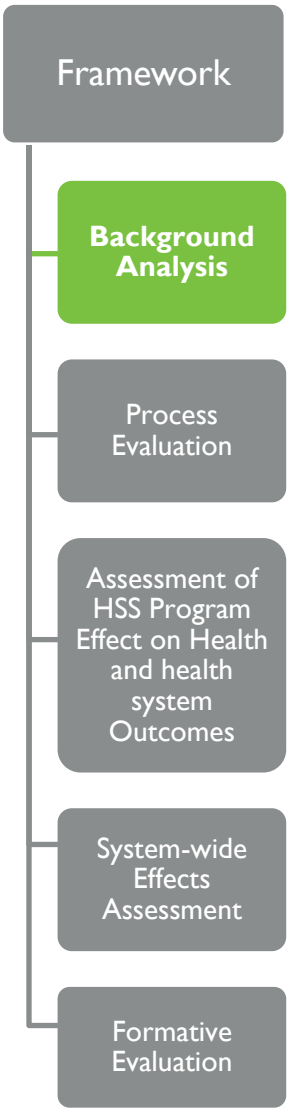
- **Evaluation questions and data collection plan to answer evaluation questions,** including facilities to be visited and key informants to be interviewed.
- **Draft evaluation report,** which includes a synthesis of the documents and observations, interviews, and conversations with key stakeholders within the implementing organizations, government, civil society organizations supporting health system activities, and program beneficiaries.
- **Completed template,** an Excel file which is integral part of this guideline.
- **Final evaluation report,** which addresses the feedback of the management of the HSS program on the draft evaluation report.

PART A: THE FRAMEWORK



2. BACKGROUND ANALYSIS

The background analysis aims to provide a scope and understanding of the HSS program. Based on a desk review of documentation received from the country and reputable on-line sources, the analysis reviews characteristics of the HSS program and the environment in which it is implemented.



2.1 CONTEXTUAL FACTORS

Evaluating background information will help the team understand the demographic, political, and economic context as well as the health profile of the country in which the HSS program is implemented. Potential background information to consider includes country-income level, child and maternal mortality, disease burden of the major epidemics and conditions that are targeted as part of the HSS program, special characteristics of the health systems, and special characteristics of the country (e.g., post-conflict). This will provide the evaluation team with an understanding of the environment in which the HSS program is implemented.

2.2 HSS PROGRAM CHARACTERISTICS

The HSS program characteristics are divided into three subcategories (see Figure 4), as follows:

1. **Sources of funding of the HSS program:** In this section, the funding sources and the funding amounts for the examined HSS program can be input by the evaluation team. A dropdown list of a wide range of domestic resources and donors is provided.
2. **Overall funding of the HSS program:** The overall funding is calculated based on figures input under the "funding sources" section. In addition, the duration of the investment(s) should be entered. The template will calculate the annual investment in the HSS program, as well as the total cost of the HSS program as a percentage of the country's total health expenditures. The larger the percentage, the higher the probability that the health system and health outcome indicators, which are the result of the HSS program, will be affected.
3. **Description of interventions:** This section should describe the HSS program and detail the geographical areas in which these interventions are implemented. The template is designed to organize the narrative by service delivery area (SDA)[4].

[4]**Service delivery areas (SDAs):** Thematic grouping of activities under each objective, for the purpose of harmonizing the Global Fund grant's programmatic and financial data and reporting.

FIGURE 4: ANNOTATED SCREEN SHOT OF HSS PROGRAM CHARACTERISTICS

HSS Program Characteristics		
Funding sources	**Amount (US$)**	**Notes**
Please select [Example: Domestic resources]		
Please select [Example: Global Fund]		
Please select		
Overall funding	**Amount (US$)**	**Notes**
Population		
Health expenditure per capita (in US$)		
Total health expenditures (in US$)		
Total investment in the HSS program (in US$)		
Total duration of the investment(s) (years)		
Annual investment in the HSS program (in US$)		
Proportion of HSS program from total health expenditures		
SDA of intervention	**Detailed description of intervention**	
Please select [Example: Health information systems (HIS) interventions]		
Please select [Facility management and organization]		
Please select		

3. PROCESS EVALUATION

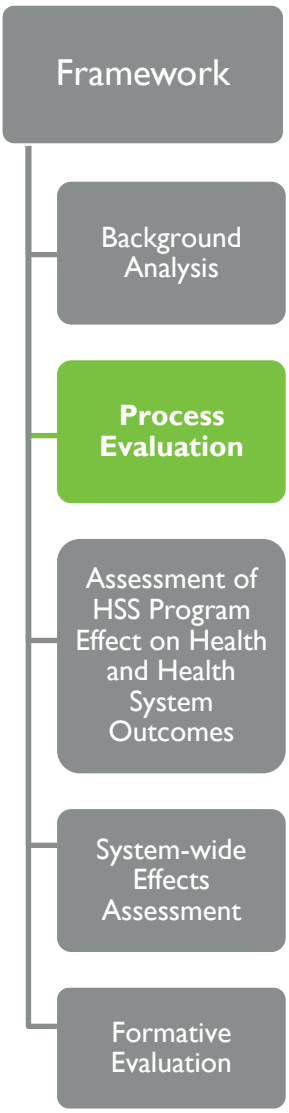
The process evaluation reviews the program-implementing entity's capacity to manage the HSS program, as well as relevant systems and processes by assessing the following:

1. Program performance
2. Structure and capacity
3. Program and financial management and processes
4. Monitoring and evaluation
5. Coordination with stakeholders
6. Alignment and harmonization

This section on HSS program performance provides a short description of the four elements related to the performance of the HSS program, based on existing reviews:

1. Summary of programmatic performance: The degree to which the HSS program is achieving its programmatic targets (if any) and the challenges it faced.
2. Summary of financial management: Summary of how the funding is being managed, including the burn rate (money expedited versus the planned budget).
3. Summary of data quality: Summary of data quality as reported by assessment reports.
4. Review of overall HSS program management: Based on the above, an overall assessment of the HSS program performance. The better the execution of the HSS program, the higher the likelihood to observe systems and health outcomes.

The above performance review should be based on existing documents, such as progress reports and annual reviews.



Annex A contains a list of questions, per element and sub-element for the remaining five areas covered by the process evaluation: structure and capacity, program and financial management and processes, monitoring and evaluation, coordination with stakeholders, and alignment and harmonization). For example, the sub-element "financial management system" has two questions:

1. Does the program implementer have a financial management system that can correctly record all transactions and balances?
2. Do the expenditures in the financial management system correspond to those reported to the funding sources?

It is important to note that these are guiding questions that aim to direct the evaluation team toward key strengths and challenges. There is no need to answer each and every question; rather, the questions are intended to guide the process evaluation toward its main goal: to provide the management of the HSS program and its stakeholders feasible, concrete recommendations to enhance the program.

Answers for the process evaluation questions are obtained through program documents and, in particular, through interviews with key stakeholders. Through this process, the evaluation team will conduct the following:

1. Identify a list of strengths and challenges.
2. Validate the findings by
    a. Discussing information gathered in interviews with other interviewees
    b. Supplementing findings, where possible, with supporting documentation from interviewees
    c. Modifying, or refining, recommendations raised by the evaluation team and the interviewees
    d. Through stakeholder engagement, further breaking down a subset of these recommendations into action items, each with timeframe and responsibilities, and if possible, an estimated cost.

The evaluation team will need to conduct interviews on identical topics with a wide range of key informants from various organizations, including the Ministry of Health (MoH), service providers, donors, and implementing partners. This **triangulation**, supplemented by documentations, will reduce the subjectivity of the collected information.

At the end of this process, the evaluation team should have a good understanding of the performance of the program and key strengths and challenges facing the program management entity and its capacity to execute the HSS program. The evaluation team should use this understanding to provide recommendation s for improving program implementation.

4. ASSESSMENT OF HSS PROGRAM EFFECT ON HEALTH AND HEALTH SYSTEMS OUTCOMES

This section details the methodology for assessing the performance of health system and health outcome indicators within the results framework of the HSS program.

Health system and health outcome indicators of the HSS program are defined as those that appear in the results framework of the strategy, operational plan, or grant agreement, and are the indicators and targets that the program was set to achieve.

The underlying assumption of the evaluation presented in this section is that each outcome indicator in the results framework needs to be linked to the interventions of the HSS program and that its targets are achievable within its scope. In HSS programs, health system indicators, as opposed to health indicators, will usually be more directly linked to HSS interventions.

An HSS program should have desired systems and health outcome objectives that it aims to achieve. For the evaluators to understand the linkages between HSS interventions and health objectives, they will review relevant health outcome indicators to assess how the HSS interventions are likely to influence these indicators. Outcome indicators for which there are no relevant interventions should be considered irrelevant, at least in relation to the reviewed HSS program, and their performance should not be measured.

Performance of the outcome indicators will be measured against targets and baselines (lack of baselines is a finding on its own and will need to be addressed). This quantitative performance review will be accompanied by a qualitative assessment, in which the performance of the interventions, the quality of their implementation, and the external confounding factors are assessed to determine if the planned interventions were likely to contribute to the performance of the outcome indicators. If the performance of the interventions is considered poor in relation to the quality of the implementation, and if there are possible confounding factors to explain any improvement in the performance of the outcome indicators, then it is impossible to attribute those improvements to the interventions of the examined program. The workflow is shown in Figure 5 and is described in detail below.

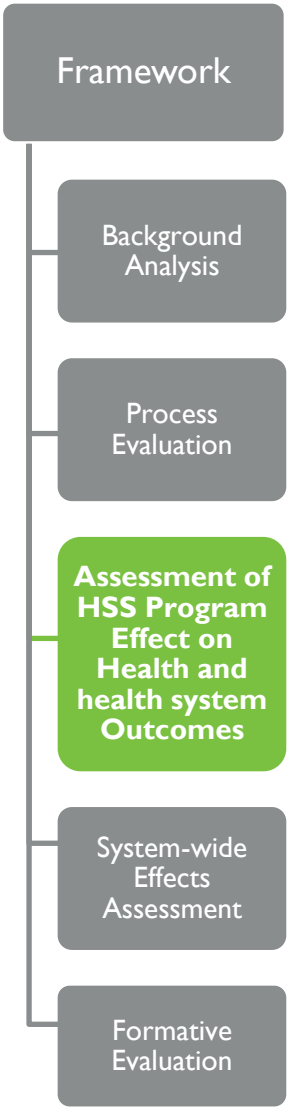
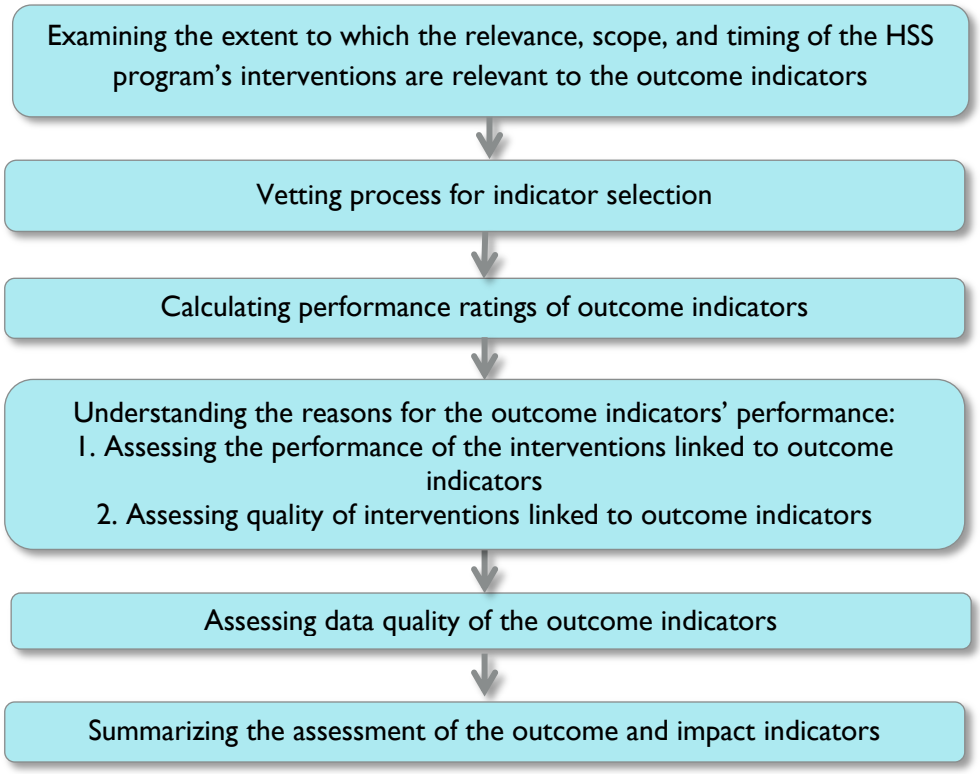


FIGURE 5: WORKFLOW OF THE OUTCOME ASSESSMENT



Box 1: Linking HSS Interventions to Health Outcome and Impact

This evaluation will be limited in scope and cannot include complex, scientifically designed studies such as randomized control trials or quasi-experimental methods. Given the limited scope of the evaluation**,** it may not be possible to examine a full causal chain between HSS activities and health outcomes and impacts. However, it is possible to gain a deeper understanding of the plausible effects of the HSS program.

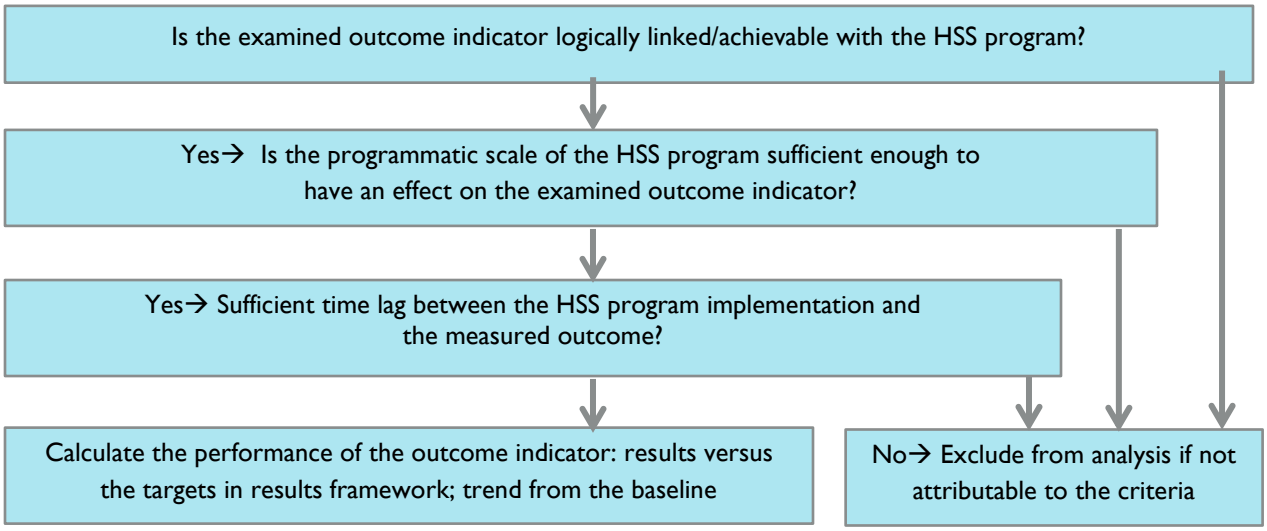
4.1 LINKING HSS INTERVENTIONS TO HEALTH AND HEALTH SYSTEMS OUTCOME INDICATORS

Prior to measuring the performance of outcome indicators, the evaluation team must ensure that the interventions of the HSS programs are relevant to its outcome indicators. This will indicate whether the desired, articulated outcome targets are achievable through the planned and implemented HSS interventions. The process suggested for establishing such linkage is as follows (also shown in Figure 6):

1. Is the examined outcome indicator logically linked and achievable with or by the HSS program (i.e. is the HSS program likely to improve or worsen those outcome indicators)? Examples of questions to raise regarding some common indicators targeted by HSS programs include the following:
    a. Under-five mortality: Can the services supported by the HSS program be expected to have a major effect on under-five mortality? For example, HIV and tuberculosis (TB) are not major causes of death among this age group, while outside Africa, malaria is usually not a major cause of death at any age group, so interventions focusing on these diseases may not be key drivers of improvements in (or worsening of) child mortality.
    b. Disease-specific outcome indicators: Do the HSS interventions have the potential to improve or worsen the performance of this indicator within the period of the HSS program? For example, strengthening training institutions may not result in sufficient numbers of trained health workers within a short time period to affect disease prevalence or treatment outcomes.
    c. Relevant services: For example, in "percentage of suspected malaria cases that have laboratory diagnosis," the question will be whether the examined HSS program contributed to strengthening lab services.
    d. Geographical overlap: Do the measured outcome indicators and the HSS program cover the same geographical area?
2. Is the planned (not actual result) programmatic scale of the HSS interventions sufficient enough to have an effect on the examined outcome indicator? If both the examined intervention and the measured outcome are at the same scale (e.g., national, district, etc.), the intervention would be relevant. For example, if the stock-outs of TB medicines is a major cause of incomplete treatment in a given country, and as part of an HSS program the entire supply management is revamped, then improvement in treatment completion might be a relevant health outcome indicator. However, in most cases, interventions are likely to have only a partial programmatic contribution; for example, training a subset of the total national workforce on an issue that is likely to affect a given national outcome indicator. In such cases, the intervention should have a substantial programmatic contribution, defined here as at least one-third (Note: this is a suggestive threshold; the evaluators can decide otherwise, if based on reasonable outcome). In the example above, training 40 percent of the total workforce would be considered substantial. In contrast, the under-five mortality indicator, usually measured at the national level and serving as proxy for health system strength, would not effectively reflect the effects of a small-scale HSS program. It is important to use the target rather than the actual result, as the latter is considered in the performance assessment.
3. Is the time lag between the intervention implementation and the measured outcome indicator sufficient to expect the intervention to have an effect on the outcomes? The sufficient time lag will vary across programs, but in general, the more direct the interaction between the intervention and the measured outcome, the shorter the lag time required for an effect to be realized and the fewer chances for confounding factors to occur.

The workflow to examine the HSS program relevance to an outcome indicator is summarized in Figure 6.

FIGURE 6: WORKFLOW TO EXAMINE THE RELEVANCE OF THE HSS PROGRAM ON TARGETED OUTCOME INDICATORS



Note: At times, outcome indicators can be linked directly to another outcome indicator rather than be a direct result of the interventions funded by the HSS program. For example, a decline in TB mortality might not be directly linked to any of the HSS programs, but linked to the default rate, which in turn is linked to an intervention to ensure TB patients are taking their medications throughout the treatment. In such cases, the above workflow will be done in two steps: first, the evaluator will go through the above workflow to assess the linkage between the intermediate outcome indicator(s) and the ultimate outcome indicator; second, if the indicators are sufficiently linked, then the workflow process will be repeated to assess the relevance of the intermediate outcome indicator(s) to the interventions.

In the example in Box 2, exploring the parameters suggested by the framework enabled the evaluators to suggest a plausible effect between the HSS program and the outcome indicator. However, in many cases either the scale or the quality of program implementation might be too inconclusive to determine the plausible effect of the HSS program on a given outcome result. For example, the training may have increased the RHMs nationally by 5 percent – a scale that is unlikely to generate a substantial effect on the national-level outcome indicator. Such caveats will need to be noted, and during the vetting process (see below), the evaluation team and the relevant M&E officer will decide whether to include this outcome indicator in the evaluation or not.

It is possible that an HSS program can be contributing only to strengthening specific aspects of the health system without direct linkage to health outcomes. For example, an HSS program that strengthens the health information system might contribute to better service delivery and supply management, yet the link to specific health outcome indicators might not be direct enough to suggest partially or fully plausible effects of the interventions on the outcome indicator. In such cases, the evaluation team should consider recommending using only system outcome indicators.

Box 2: Assessing Plausible Effects of the HSS Program on an Outcome Indicator – an Example

The workflow to plausible effects of the HSS program on an outcome indicator is illustrated through the example of a hypothetical HSS program which includes in its results framework the following outcome indicator: Percentage of general population that can correctly identify four or more malaria signs and symptoms.

Q: Is there a logical link between the HSS program and the examined outcome indicator?

The HSS program trained rural health motivators (RHMs) (the output indicator) to promote health, including malaria prevention, in their communities. The outcome indicator was to increase the community's awareness and knowledge of the signs and symptoms of malaria. Thus, there is logical link between the intervention and the outcome indicator.

Q: Is the intervention at a magnitude that can plausibly result in a change consistent with the target of the outcome indicator?

The HSS program aimed to add 2,500 RHMs to the existing 4,000 RHMs nationally – a 62-percent increase in the number of RHMs that enables achieving national coverage of RHMs. Therefore, an effect at the national level is reasonably expected.

Q: Is the time lag between intervention implementation and the measured indicator sufficient to expect an effect of the intervention?

In this case the knowledge of malaria signs and symptoms in the general population was measured one year after the training was conducted, enabling the RHMs to disseminate the messages and the community members to internalize these messages (if they were communicated correctly and effectively). Thus, this indicator is valid.

Given all these assessment results, the performance of this outcome indicator should be measured.

4.2 VETTING PROCESS FOR INDICATOR SELECTION

The extent to which the HSS program had a plausible effect on an outcome indicator will not always be clear. However, the decision as to whether there is a plausible effect between the HSS program and a given indicator will affect the evaluation's results. To ensure that this process is robust, the evaluators should select the indicators that appear to be relevant for HSS program interventions in the planning stage, and refine them in consultation with key informants. The selection and any revision of indicators and interventions need to be vetted by a designated M&E officer of the entity commissioning the evaluation.

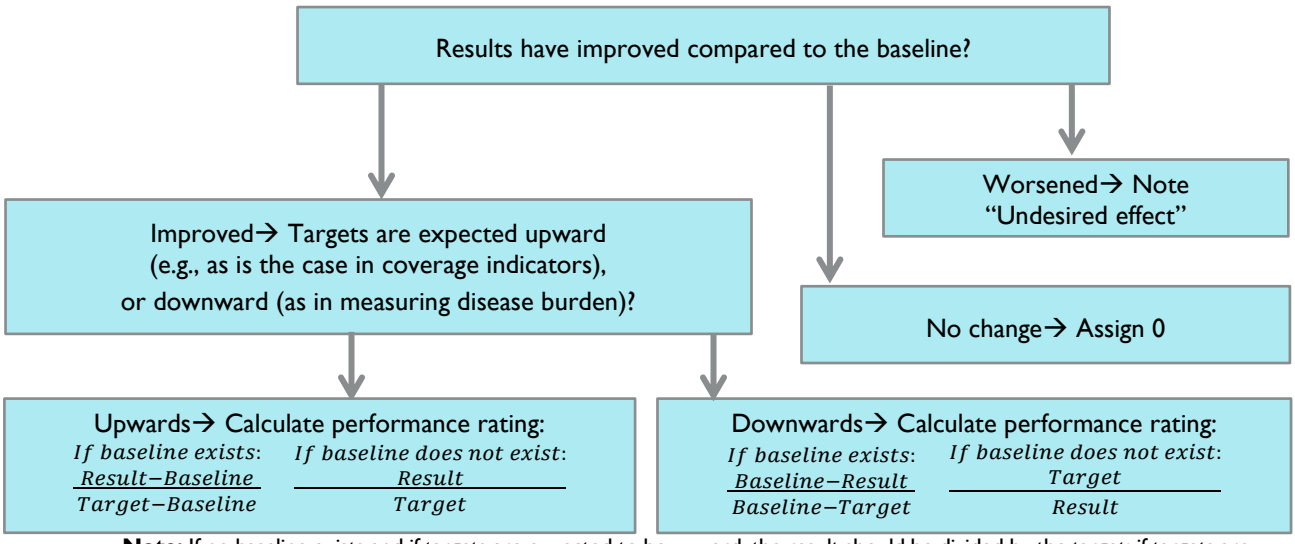
4.3 CALCULATING PERFORMANCE RATINGS OF OUTCOME INDICATORS

The existence of annual targets for outcome indicators within the results framework of an HSS program enables the evaluators to measure performance by comparing results versus targets. As such, each outcome indicator that the evaluation team concluded to be plausibly affected by the HSS program (as per the conclusions from the assessment step above) will have its performance calculated. The performance calculation will consist of measuring the results against the target of the outcome indicator, as defined in the results framework of the HSS program. There are two types of equations for performance, depending on whether a positive trend is upward, where a result is expected to increase

from the baseline (e.g., treatment success rates), or downward, where a result is expected to decrease from the baseline (e.g., indicators measuring disease burden) (Figure 7).

Importantly, if there is no baseline in the results framework, the performance can still be measured; however, the evaluation teams must note the lack of baseline data and address this both in the findings and in the recommendations.

FIGURE 7: RATING METHODOLOGY WORKFLOW AND CALCULATIONS



Note: If no baseline exists and if targets are expected to be upward, the result should be divided by the target; if targets are expected to be downward, the target should be divided by the result.

The performance rating reflects the extent to which the funded program achieved the outcome indicators within the results framework. The baseline and target measurements are important since together they indicate the extent to which the program aimed to increase the coverage or reduce the burden of disease. This is important as the larger the expected increase, the larger the budget required to reach the target.

For example, a 10-percent increase in coverage when the gap between the baseline and the target is 20 percent should be treated differently from when the gap between the baseline and the target is 80 percent (see Table 1). In the former, the performance will be 50 percent (10%/20%=50%), while in the second scenario, the performance will be substantially lower: 12.5 percent (10%/80%=12.5%). Even though the increase in coverage for both is the same, the former advanced more in bridging the gap between the baseline and the target to which the HSS program committed.

TABLE 1: EXAMPLE OF PERFORMANCE CALCULATION IN TWO SCENARIOS

	Baseline	Target	Result	Expected Increase	Actual Increase	Performance
Scenario A	10%	90%	20%	90%-10%=80%	20%-10%=10%	10%/80%=12.5%
Scenario B	70%	90%	80%	90%-70%=20%	80%-70%=10%	10%/20%=50%

To provide an overall assessment of the outcome indicators relevant to the HSS program, the evaluators will summarize the number of indicators that fall into each of the following performance categories (largely based on the Global Fund's categories for output indicators):

1. Performance of 90 percent or above: meets or exceeds targets.

2. Performance of 60–89 percent: adequate trend, targets are not met.
3. Performance of 30–59 percent: inadequate achievement, yet potential demonstrated.
4. Performance of 0–29 percent: no or insignificant trend.
5. Negative performance: undesired effect.

4.4 UNDERSTANDING THE REASONS FOR THE OUTCOME INDICATORS' PERFORMANCE

In assessing whether the performance of outcome indicators is the result of the HSS program, one needs to show not only that the interventions were relevant to the outcome indicators, but also that interventions were of sufficient programmatic performance and that they were implemented in a quality manner, as detailed below.

4.4.1 EXAMINING THE PROGRAMMATIC PERFORMANCE OF THE INTERVENTIONS

Programmatic performance of interventions is based on results versus target, as defined by the program. Such performance is frequently readily available (e.g., in progress reports of Global Fund grants). Where performance is absent, yet both results and targets exist, the above performance calculation should be applied (frequently in output indicators baselines are not factored). The higher the performance, the higher the probability that any improvement in the performance of the outcome indicator is linked to the intervention.

4.4.2 EXAMINATION OF THE QUALITY OF THE INTERVENTION

The quality of the intervention is done by examining the following three dimensions[5]:

1. **Planning:** Denotes the manner in which the activity was planned. This includes whether a written work plan, budget, and/or protocol was prepared; if and how key stakeholders were involved; how the intervention was tailored to the program capacity and local context; and how national and international guidelines, as well as best practices, were factored.
2. **Structure:** Denotes the attributes of the settings in which the intervention occurs. This includes the attributes of material resources (such as facilities, equipment, and money), human resources (such as the number and qualifications of personnel), and organizational structure (such as methods of peer review and methods of reimbursement).
3. **Process:** Denotes what is actually done when implementing the intervention. This includes questions such as how a training was conducted, the extent to which key stakeholders were actually involved, and the extent to which treatment followed best practices and guidelines.

The aim of the above classification is to assist the evaluators in the dimensions of quality to review. The findings and recommendations are not tied to this classification. For examples on questions to explore when assessing quality using these dimensions, please refer to Annex B.

[5] Partially based on Donabedian, A. (1988). "The quality of care: How can it be assessed? ." Journal of the American Medical Association **260**(1743–1748).

Much of the data collection for delivery sites will focus on assessing the quality of the interventions. For other services such as improving the central HIS or the construction of a central warehouse for the medicines, the evaluation team will require a smaller number of site visits.

4.4.3 REVIEWING POTENTIAL CONFOUNDING FACTORS

In addition to examining the quality of the HSS program, the evaluation team will need to identify potential factors beyond the HSS program that might have contributed to the outcome indicators' performance by exploring the following questions per outcome indicator:

1. Were other major health or economic interventions occurring at the same time, among the same population, that could have had a mediating effect, such as another large health program or a national campaign?
2. Were there external circumstances that could have affected the intervention, such as natural disasters, political changes (e.g., floods, civil unrest, elections), or problems with the health system (e.g., delays in paying health workers, organizational changes)?

4.5 EXAMINING DATA QUALITY OF THE OUTCOME INDICATORS

In reviewing the quality of the baseline and the results of outcome indicators, the timeliness, the relevance, and the robustness of the sources need to be examined, through the following steps:

1. **Timeliness:**
    a. **Baseline**: Is the baseline within a reasonable timeframe (usually no more than one year before and no later than three months after the HSS program start date [first disbursement])?
    b. **Result**: Is the result being compared to the nearest target available within the closest timeframe? For example, a result measured two years and eight months after the first disbursement should be compared to the year 3 target, not year 2 target.

Importantly, surveys collecting data on morbidity and mortality are unlikely to be conducted frequently, and the results are released several months, even a year, after being conducted (see Annex C on timeliness of various key data sources). This might limit the ability to locate timely data for either the baseline or the latest target, or both.

2. **Relevance:**
    a. Do the baselines and results correspond precisely to the target indicator as per its definition? Examples: HIV prevalence among blood donors cannot be used as a baseline or a result for the target HIV prevalence among pregnant women; Under-five mortality in urban areas cannot be used as a baseline for national under-five mortality.
    b. Are the targets of the outcome indicators under-ambitious (targets are but a slight improvement from the baseline) or overambitious (e.g., targets of universal coverage where coverage is very low)?
3. **Robustness of data source:** Were the actual results measured (e.g., from facility data or population-based survey), or are assumptions used (as is the case in estimates derived from modeling or expert assessment)?
    a. Are the indicators of disease burden based on robust population-based surveys such as a census and the Demographic Health Surveys (DHS):

i) Are they combined with proxies for facility-based data, and if so, explain why. Examples:

(a) If HIV prevalence from DHS and facility-based antenatal care data are combined, explain why. DHS or other population survey data are usually superior for measuring HIV prevalence among the general population.

(b) In measuring the mortality rate, if both the case fatality rate (e.g., from facility-based data) and mortality rate (e.g., from vital registry) are combined, explain why. A vital registry, if it is comprehensive, is superior for measuring mortality among the general population.

ii) Are there any issues with the quality of the data: Is the sampling representative of the target population? Is the facility-based data complete, accurate, and precise?

b. If baselines or results are based on an assessment of experts or a modeling methodology, assess the quality of the data on which the estimates are based and the assumptions used.

4.6 SUMMARIZING THE ASSESSMENT OF HSS PROGRAM'S EFFECT ON HEALTH OUTCOMES

The evaluators will categorize and summarize the findings and recommendations from this assessment as follows:

1. **Plausible effects of HSS program on the systems and health outcomes:** Review the number and nature (e.g., systems or health, outcome or impact) of outcome indicators in each of the five performance categories (meeting or exceeding targets; adequate trend, targets are not met; inadequate trend, yet potential demonstrated; no or insignificant trend; or undesired effect – see section 4.3). The performance assessment will be accompanied with a summary of the overall trends, including quality of interventions and confounding factors that might have affected those trends. The evaluation team will explain why the examined outcome indicator is plausibly affected by the HSS program, and if there is a need for additional evaluations to establish a more conclusive link.
2. **Assessment of the HSS program design:** Per outcome indicator, consider whether there are any relevant interventions in the HSS program, and, if not, should the outcome indicator be modified or should the HSS program be reprogrammed.
3. **Assessment of the measurements:** Provide an assessment of the ability of the results framework to measure the outcome. Per outcome indicator, the timeliness, relevance, and robustness of the data source will be reviewed, as detailed above.

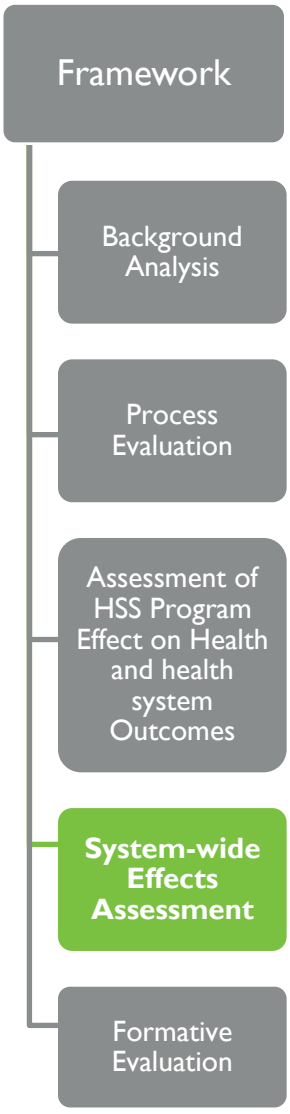
5. ASSESSMENT OF THE HSS PROGRAM'S SYSTEM-WIDE EFFECTS

Relevant components of the health system, supported by the HSS program, will be assessed. For the purpose of the evaluation, the health system components identified are loosely based on the World Health Organization (WHO) framework and include the following: facility management and organization, infrastructure, health information systems, procurement and supply chain management, health workforce, health financing, and stewardship and governance. For each of the seven components, the evaluator will qualitatively review the system-wide effects using the following five criteria: equity, quality, efficiency, accessibility, and sustainability.

A checklist of topics to explore is provided in Annex D. The methodology here is similar to the one in the process evaluation: the questions in Annex D aim to direct the evaluation team toward key strengths and challenges. Annex D should not be viewed as a questionnaire, but rather a menu of relevant questions from which an interviewer can pick, depending on the interviewee and the context.

The goal of this assessment is to provide the management of the HSS program and its stakeholders with feasible, concrete recommendations to increase the positive effects of the program and to reduce or eliminate any negative ones.

As previously noted, only SDAs relevant to the HSS program will be examined. The evaluation team will have latitude to select relevant health system components for examination, to prioritize some performance dimensions over others, and to examine additional relevant questions, depending on the HSS program. Each question guides the evaluation team to examine the potential positive and negative effects of the HSS program on a given area of the health system. As with the process evaluation, answers are obtained through program documents and routine data, and through interviews with key stakeholders and technical partners. Through this process, the evaluation team will conduct the following:



a. Solicit responses on positive and negative effects.

b. Triangulate the findings by soliciting feedback on evidence and opinions previously documented and provided by other interviewees. The evaluation team should not provide the information source if the source is another interviewee.

c. Supplement findings, where possible, with supporting documentation from interviewees.

d. Test, modify, or refine recommendations developed by the evaluation team based on the assessment (see Phase 3 in chapter7).

e. Work with stakeholders to distill recommendations into action items, each with a timeframe, responsibilities, and, if possible, an estimated cost.

The evaluation team will need to interview a wide range of key informants from various organizations, including the MoH, service providers, donors, and implementing partners, on identical topics. This **triangulation,** supplemented by relevant documentation, will reduce the subjectivity of the collected information.

Table 2 provides examples of potential system-wide effects of selected interventions. Additional examples can be found in a technical report from the Partners for Health Reform*plus* project[6].

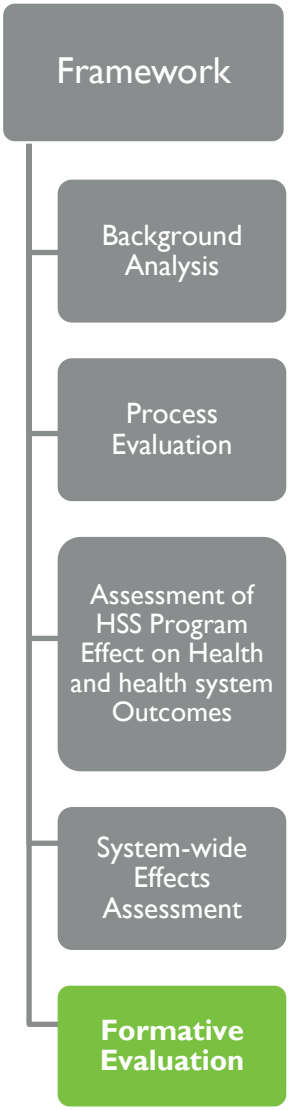
TABLE 2: EXAMPLES OF POTENTIAL SYSTEM-WIDE EFFECTS FOR SELECTED INTERVENTIONS

Intervention	Strengths	Challenges	Recommendations
Building and operating new primary health facilities	Improves accessibility to health services	None	
Building ART and DOTS clinics in all health facilities	Improves accessibility to HIV and TB services	Substantial investments in areas where burden of HIV and/or TB is minimal Limited capacity of the program to implement such an ambitious plan	Prioritize implementation of this activity in areas with higher disease burden
Recruitment of new health workers with donor funding	Reduces the shortage in health workers	Nonsustainable, as funded by donors	Formulate a transition plan, where financial responsibilities of employing the additional health workers are gradually moved to the MoH
Establish a 5-year M&E plan for the HIV and TB programs in low HIV and TB prevalence countries	Strengthens the M&E capacity of both programs	Diversion of M&E-related resources from the wider health services	Expand the M&E plan to the M&E of the entire health system; apply essential, interim solutions for the M&E of HIV and TB programs

[6] Bennett, S. and A. Fairbank (2003). The system-wide effects of the Global Fund to Fight AIDS, Tuberculosis and Malaria: A conceptual framework. Technical report no. 031. Bethesda, MD, The Partners for Health Reformplus Project, Abt Associates.

6. FORMATIVE EVALUATION

The formative evaluation reviews the HSS program in the context of external dynamics that have occurred since the HSS program was designed (see Figure 8). The formative evaluation is aimed at identifying those areas of the HSS program that may need to be redesigned due to the following factors:



1. **Country has changed NSP:** This will consist of a review of the HSS program against the NSP, to assess the extent it complies with the program and whether any changes are required. For example, if the HSS program emphasizes strengthening the central level while the NSP promotes decentralization, there will be a need to consider recommending changes to the HSS program.
2. **New normative guidelines/policy documents have become available:** This will consist of a review of recent guidelines and policies to see if the HSS program needs to be aligned with the guidelines/policies; for example, changes in guidelines that reduce dependency on lab services, such as an increase in syndrome management for sexually transmitted infections and/or an increase in rapid HIV testing. This will require revisiting the HSS program that focuses heavily on strengthening lab services.
3. **Changes in financial resources have occurred:** This will involve changes such as whether a new donor is providing support for areas that overlap with the HSS program, or whether a donor pulls out of supporting the program. This review will also factor potential changes in country eligibility for future funding. For example, a country that expects its income level to increase and, as a result, is less likely to get donor funding will need to reduce donor dependency of concurrent costs of the HSS program.
4. **Changes in demographic and disease burden**: This will consist of new information on the demographic and disease burden, including more robust estimates on the disease burden that allow for better resource allocation. For example, lower estimates of HIV prevalence will require the estimated number of people in need of antiretroviral therapy and prevention of mother-to-child transmission services to be lowered, and the amount of resources allocated for these services to be reduced.

The methodology used here is similar to the one applied in the process evaluation and the system-wide effects assessment: triangulation of information obtained from HSS program documentation and interviews with key informants

FIGURE 8: SCREEN SHOT OF THE FORMATIVE EVALUATION

SDA of intervention	Detailed description of intervention	Relevant country changed NSP	Relevant new normative guidelines/policy documents	Relevant changes in financial resources	Relevant changes in demographic and/or disease burden	Proposed follow-up action	Timeframe of implementation	Estimated budget (if known)
Example: Health Workforce								
Example: Stewardship and Governance								
From Background Analysis worksheet								
From Background Analysis worksheet								
From Background Analysis worksheet								
From Background Analysis worksheet								
From Background Analysis worksheet								
From Background Analysis worksheet								
From Background Analysis worksheet								

PART B: THE GUIDELINE

Phase 1	Phase 2	Phase 3	Phase 4
Preparation and initiation	**Data collection and analysis**	**Formulation of recommendations**	**Evaluation completion**
1. Assemble evaluation team and data collectors	5. Collect data	7. Present preliminary findings to stakeholders and formulate recommendations	8. Draft evaluation report
2. Notify program and obtain national authorizations	6. Analyze the findings		9. Review and collect feedback from country and commissioning organization
3. Request documentation			10. Finalize evaluation report
4. Formulate data collection plan			11. Initiate follow-up of recommended actions

7. GUIDELINE

The evaluation consists of four phases, detailed in figure in the previous page and Box 3. Phase 1 includes preparation and initiation. Phases 2 and 3 are conducted in-country and consist of data collection, analysis, and formulation of recommendations with the stakeholders. Phase 4 consists of writing reports and incorporating additional comments to finalize the evaluation report and recommendations. Box 3 provides a detailed explanation of each of these steps.

Box 3: Checklist for HSS Evaluation Implementation

Phase 1: Preparation and initiation

Step 1. Assemble evaluation team and data collectors

Step 2. Notify program and obtain national authorizations

- Notify program
- Notify other key stakeholders (as appropriate)
- Obtain national authorization, if required

Step 3. Request documentation and list of key informants

- Ask the HSS program to assign three focal points: one for the process evaluation, another for the assessment of the outcome indicators, and a third for the assessment of the system-wide effects
- Obtain HSS program documents
- Obtain national health systems analysis report, national strategies and plans
- Obtain list of key informants

Step 4. Formulate data collection plan

- Conduct the background analysis
- Identify key informants to interview for the process evaluation and the system-wide effect assessment
- Select the outcome indicators from the results framework of the HSS program to be assessed in this evaluation
- Formulate questions per outcome indicator
- Assess what data are required from service delivery sites and formulate a questionnaire and design the sample of sites to be visited

Cont next page

Box 4: Checklist for HSS Evaluation Implementation. Cont…

Phase 2: Data collection and analysis

Step 5. Collect data

- Brief the HSS program management and office of the commissioning organization
- Conduct data collectors' training and piloting of the questionnaire
- Collect data and interview key informants

Step 6. Analyze the findings

- Conduct process evaluation:
  - Identify a list of strengths and challenges in management of the HSS program
  - Formulate recommendations for discussion with stakeholders
- Assess plausible outcome:
  - Assess relevance of selected outcome indicators to HSS program
  - Calculate the performance of the attributable outcome indicator
  - Review the performance and the quality of the interventions
  - Review potential confounding factors
  - Review quality of the outcome indicators in term of timeliness and relevant and robustness of data source
  - Summarize outcome indicator assessment:
    - Plausible outcome of the HSS program
    - Assessment of the HSS program design
    - Assessment of the data quality of the outcome indicators
- Assess the HSS program's system-wide effects
  - Identify a list of positive and negative effects of the HSS program
  - Formulate recommendations for discussion with stakeholders
- Conduct formative evaluation
  - Identify changes in NSP, guidelines, policies, the funding environment, and demographics and disease burden that are relevant to the HSS program
  - Formulate recommendations for discussion with stakeholders

Phase 3: Formulation of recommendations

Step 7. Present preliminary findings to stakeholders and formulate recommendations

- Discuss the findings and recommendations in one-to-one interviews and small group discussions:
  - Accept, modify, or reject recommendations
  - Prioritize up to 15 recommendations for discussion
  - Develop initial implementation plan of prioritized recommendations
  - Organize and conduct a stakeholders' workshop
    - Accept, modify, or reject prioritized recommendations
    - Develop final implementation plans of prioritized recommendations

Phase 4: Evaluation completion

Step 8. Draft evaluation report

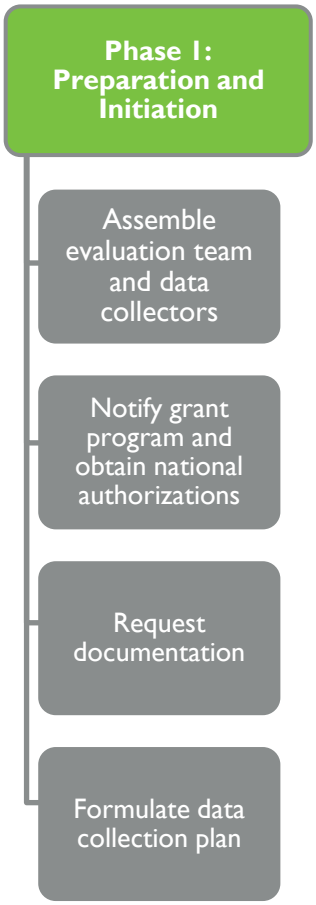
Step 9. Review and collect feedback from country and commissioning organization

Step 10. Finalize evaluation report

Step 11. Initiate follow-up of recommended actions

PHASE 1: PREPARATION AND INITIATION

The first phase of the evaluation occurs prior to the evaluation team being on-site at the location of the program/project. Responsibility for Phase 1 rests partly with the organization commissioning the evaluation and partly with the evaluation agency. It is important to note that it is assumed that the HSS program is being selected per internal criteria of the organization commissioning the evaluation.



STEP 1. ASSEMBLE EVALUATION TEAM AND DATA COLLECTORS

Each evaluation is estimated to be conducted over a period of 8 to 12 weeks by a team of three evaluators, with a two-week in-country data collection. The core evaluation team typically includes an evaluation lead, an evaluation specialist, and an evaluation analyst, the latter being the junior team member who handles logistical coordination and assists with team coordination, data analysis, and writing. The required skill matrix is detailed in Table 3, and will differ slightly from one evaluation to another. Most of the skills in Table 3 are likely to be with the lead and/or specialist and, to a lesser extent, with the analyst.

During fieldwork, the evaluation team expands. It typically includes one or two local, senior data collectors that are hired to supervise and coordinate data collection and an additional two to six junior data collectors, depending on the scope of the evaluation and the study sample.

TABLE 3: LIST OF REQUIRED SKILLS OF THE EVALUATION TEAM

Required Skills
1. Expertise in the main building blocks covered by the HSS program
2. Knowledge of the prime diseases to which the HSS program is responding
3. Knowledge of the mechanism through which the funding is provided
4. Development of results frameworks and/or M&E plans, including formulation of indicators and target setting
5. Experience in financial management/audit/budgeting
6. Experience in, or at least knowledge of, program management unit
7. Experience in, or at least knowledge of, sub-contractor selection and management
8. Knowledge of outcome and impact evaluation
9. Knowledge of data quality
10. Experience in developing questionnaire for data collection from delivery points
11. Expertise in stakeholder engagement
12. Knowledge of the official language in the country of the examined program, and including good writing skills for preparation of the report

The data collectors should not be employees of the evaluated HSS program. The number of days expected per team member is detailed in Table 4. The evaluation team should be finalized, and roles and responsibilities should be clearly defined prior to the fieldwork.

TABLE 4: ESTIMATED DAYS PER TEAM MEMBER

Stage (duration)	Evaluation Lead	Evaluation Specialist	Evaluation Analyst	Two Senior Data Collectors	Four Junior Data Collectors
Planning (5 weeks)	5 days	5 days	10 days	0	0
Fieldwork (2 weeks)	11 days	11 days	11 days	10 days each	10 days each
Data analysis & writing (10 weeks)	25 days	25 days	40 days	0	0
Total estimated days	41	41	61	20	40

Notes:

1. The estimated number of days might vary from one program to another, depending on the complexity of the HSS program. The above assumes the evaluation team conducts a comprehensive review of relevant documents, interviews approximately 40 key informants, visits 12 sites, and conducts one stakeholder meeting.
2. The duration of each stage (shown in first column, in weeks) is the time the team members should allow for the evaluation. For example, while only 5 days are required from the lead during the planning stage, s/he should expect to spread it over 5 weeks, during which approvals are obtained and arrangements are taking place.

STEP 2. NOTIFY PROGRAM AND OBTAIN NATIONAL AUTHORIZATIONS

A. Notify Program

The organization commissioning the evaluation should notify the country/program about the evaluation as soon as possible and obtain national and other relevant authorizations. The organization should also notify other organizations, as appropriate, about the evaluation and request cooperation. The evaluation team is expected to comply with national regulations regarding data confidentiality and ethics. It is the evaluation team's responsibility to identify such national regulations and adhere to them.

It is important that the organization commissioning the evaluation stress the need for the relevant M&E unit staff member(s) to accompany the evaluation team on its site visits. A notification letter/email should be accompanied by the initial documentation request from the M&E unit (list of documents can be found in Table 5).

After a notification letter/email has been sent, the organization commissioning the evaluation should send a copy of the notification letter/email to all relevant stakeholders, including the following:

- Host country officials related to the program/project being evaluated
- National M&E agency, as appropriate
- Donors, development partners, international implementing partner organizations, and relevant M&E working group representatives.

The evaluation team should follow up with the selected program/project about the pending evaluation, timeframes, contact points, and the need to supply certain information and documentation in advance (the latter is detailed in Step 3).

B. Obtain National Authorization

In certain cases, special authorization for conducting the evaluation may be required from another national body, such as an ethics committee. Such a request should be sent by the organization commissioning the evaluation. The recipient(s) of the authorization letter will vary according to what

program or project is being evaluated. The national authorization and any other relevant permission to conduct the evaluation from donors supporting evaluation sites or program/project officials should be included in the final evaluation report as an attachment.

TABLE 5: INFORMATION TYPE AND DOCUMENTATION TO BE OBTAINED

Information Type		Potential Documentation	Source for Global Fund Documents
HSS Program-Related Documents	**Funding request**	Proposal of reviewed grant	Country page on Global Fund website
	Grant agreement	Grant agreement	Country page on Global Fund website
	Revisions in grant agreement	Condition precedents (CP) and management letters	Country page on Global Fund website (CPs can be found in GPR – see below)
	Results framework	Performance framework	Country page on Global Fund website (within grant agreement)
	Budget and work plan	Grant budget and work plan	From HSS program or the Global Fund
	M&E plan	M&E plan	From HSS program or the Global Fund
	Assessment of the M&E system	Monitoring and Evaluation Systems Strengthening Tool (MESST)	From HSS program or the Global Fund
	Reports on grant performance and funding decisions	Progress update/disbursement decision (PU/DR), disbursement decision forms. Grant Performance Reports (GPR), and Grant Score Cards (GSC)	From HSS program or the Global Fund
	Disbursement decision forms	Disbursement decision forms	The Global Fund
	Reports on expenditures	Enhanced Financial Reporting (EFR)	From HSS program
	Assessment of the data quality	On-site data verification (OSDV and Data Quality Audit (DQA)	From HSS program – results may be reported in GPR and GSC
	Assessment of the quality of services	Rapid Service Quality Assessment (RSQA)	From HSS program – results may be reported in GPR and GSC (see above)
Other	**Progress report of the country**	Varies by the scope of the HSS program. For example, UNGASS report of the country will be relevant where the HSS programs is linked to an HIV response	From HSS program and from the website
	Relevant country policy documents	Strategies of the targeted diseases, thematic strategies of areas targeted by the investment., (e.g., HRH strategy)	From HSS program and from the website
	International guidelines and best practices	For example, treatment guidelines of the targeted diseases	From HSS program and from the website
	Evaluations and assessments	For health system: World Bank's joint assessment of national strategy (JANS), Health Systems 20/20's Health System Assessments (HSA) For HIV: AIDS Indicator Survey (AIS), behavioral surveillance surveys, Priorities for Local AIDS Control Efforts (PLACE) For TB: TB prevalence survey For Malaria: Malaria Indicator Survey (MIS) Cross-cutting: Demographic Health Surveys (DHS), UNICEF's Multiple Indicator Cluster Surveys (MICS)	From HSS program and from the website

STEP 3. REQUEST DOCUMENTATION

The evaluation team will need three types of documentation at least two weeks in advance of the country mission:

1. HSS program-related documents.
2. National health systems analysis reports, national strategies, service provision guidelines, and policy plans.
3. List of implementers and stakeholders recommended to be involved in the evaluation process and to be informed by the evaluation results. This list would be the basis for planning the in-country data collection.

It is important to ask the evaluated HSS program to assign three focal points: one for the process evaluation, another for the assessment of the outcome indicators, and a third for the assessment of the system-wide effects. This will reduce the dependency on one person and will enable the evaluation team to spread the requests, including the various required documentation, among the three focal points.

The organization commissioning the evaluation should reinforce the importance of these focal points' involvement in the evaluation to ensure that the data used and ideas reflected in the evaluation best represent the reality of the program. The recommended information sources and documentation to be obtained and a list of potential people to interview are detailed in Table 5 and Table 6, respectively.

TABLE 6: LIST OF POTENTIAL INTERVIEWEES

Potential Interviewees
Oversight of the HSS program: local stakeholders (MoH, other public, private, and civil society actors), international donors, and other partners
Funding recipients: Program management entity and subcontractors
Technical coordinators of the reviewed program
M&E officers at program national/regional level
M&E officers of central MOH M&E unit
MoH Department of Planning, regional and district officials
Service delivery personnel
Persons infected and affected: e.g., network of people living with HIV
Other technical partners and donors

STEP 4. FORMULATE DATA COLLECTION PLAN

During this step, the data to be collected and from whom are defined.

A. **Identify key stakeholders and informants**: Formulate a list of stakeholders to be invited to a workshop on the findings and recommendations of the evaluation, in coordination with the commissioning organization and the HSS program management. This list will be the basis for identifying interviewees for the next steps. The evaluation team should identify additional people to interview (e.g., by referral from other interviewees).

B. **Conduct background analysis**: Populate the "Background Analysis" worksheet in the template based on desk review of the documentation received and through documents on the website.

C. **Identify key informants for process evaluation**: Identify relevant interviewees and set up interviews with them.

D. **Assess the potential effects of the HSS program on the outcome indicators in the results framework**: Formulate questions per outcome indicator, as presented in Table 7. *This is*

the most time-consuming part of the data planning and data collection. It will require developing a simple and short questionnaire for the site visits that senior and junior data collectors can administer.

E. **Assess the HSS program's system-wide effects**: Review the SDAs that the HSS program are likely to affect and select interviewees accordingly.

Note that several key informants will be interviewed on more than one component of the evaluation.

TABLE 7: PROCESS TO FORMULATE QUESTIONS PER OUTCOME INDICATOR

Information Required	Information Source at the Planning Phase	Questions to Raise during Fieldwork
Is the examined outcome indicator logically linked/achievable with the HSS program?	Funding proposal, results framework, literature review of studies on linkages between the examined interventions and outcome indicators	None, if the documentation reviewed in the planning phase is sufficient.
Is the programmatic scale of the HSS interventions sufficient enough to have an effect on the examined outcome indicator?	Funding proposal, results framework, M&E work plan, expenditure reports, performance reports	If the HSS interventions are at the same scale and geographical area of the outcome indicator, and are tied to the HSS program, no further action is required. Otherwise, they will need further documentation of the expenditures of other stakeholders, usually obtained in the country visit through interviews.
Is the time lag between the intervention implementation and the measured outcome indicator sufficient to expect a linkage between the two? (see Annex C)		None, if the documentation reviewed in the planning phase is sufficient. Requires the dates of the implementation of the activities, obtained from the performance assessments, and the period to which the outcome indicators relate.
Is the intervention of sufficient quality to achieve the desired effect? Three dimensions are examined: Planning, Structure, and Process.	Performance reports, assessment of the quality of services	See Box 5
Confounding factors: a. Were there any other major health or economic interventions occurring at the same time among the same population that could have had a mediating effect, such as another large health program or a national campaign? b. Were there external circumstances that could have affected the intervention, such as natural disaster, political changes, or other problems with the health system?	Performance reports and annual assessments	Site visits and interviews with stakeholder will be necessary.

Information Required	Information Source at the Planning Phase	Questions to Raise during Fieldwork
Quality of the outcome indicators – **Timeliness:** Baseline: Is the baseline within a reasonable timeframe (usually no more than a year before and no later than three months after the HSS program start date [first disbursement])? Result: Compared to the nearest target, is the result time-wise?	Results framework, M&E work plan, performance reports	None, if the documentation reviewed in the planning phase is sufficient.
Quality of the outcome indicators - **Relevance:** Do the baselines and results correspond precisely to the target? Are the targets of the outcome indicator underambitious or overambitious?	Funding proposal, results framework, M&E work plan, performance reports	None, if the documentation reviewed in the planning phase is sufficient.
Quality of the outcome indicators – **Robustness of data source:** a. Were the results actual measures, or are any estimations used? b. Are indicators of disease burden based on robust population-based surveys? If they are combined with proxies for facility-based data, explain why. c. Are there any issues with the quality of data: is the sampling representative of the target population? Is the facility-based data complete, accurate, and precise?	Funding proposal, results framework, M&E work plan, performance reports	Question C will require understanding the source data, through review of data source and its methodologies, as well as through interviews with relevant stakeholders.

Box 5: Developing a Questionnaire and Sampling Sites of Data Collection

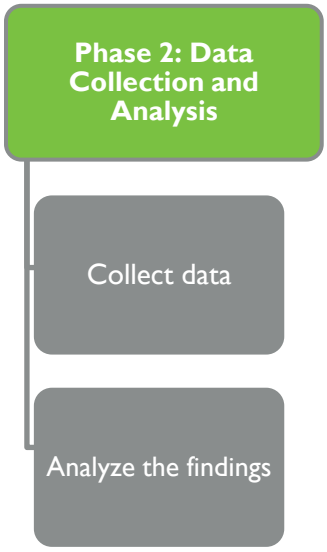
For selected interventions that are likely to be of major relevance to an adequate or well-performing outcome indicator, the quality of the intervention's implementation will be assessed, since poor quality will question the possibility that the HSS intervention had an effect on performance. The evaluation team can usually assess quality by reviewing documentation, conducting interviews, and visiting one or two sites. However, for some interventions, more comprehensive site visits will be required. To determine how indicators will be assessed for their quality, the steps below will be followed:

1. **Determine quality measures:** Per examined indicator, the evaluation team will determine what will be considered as a measure of quality of the examined intervention. Some examples include documented ability to achieve objectives, feasibility, and appropriateness to environment.
2. **Conduct desk review:** The evaluation team will review documentation for information on the quality of the examined intervention.
3. **Prioritize:** Where measuring the quality of an intervention requires visiting a number of service delivery points or interviewing a number of key informants on the same intervention, the evaluation team will consider developing a structured questionnaire that can be administered by data collectors. Given the limited time and resources, data collectors will be limited to 12–15 delivery sites if the questionnaire is long, or 24–30 delivery sites if the questionnaire is short. As such, the quality of one to two interventions can be examined, and there might be a need to prioritize. This will be done based on the following criteria:
    a. The share of the HSS program allocated to each intervention
    b. The number of outcome indicators attributed to each intervention
    c. Whether a set of service delivery points can be used to examine the quality of more than one intervention.

    A good source of questions on quality of service can be taken from the Rapid Service Quality Assessment (RSQA) tool of the Global Fund (http://www.theglobalfund.org/en/me/documents/MEQualityServices/)
4. **Sampling:** Sampling of service delivery points will typically follow the on-site data verification (OSDV) sampling guidelines, where ideally, random selection will be applied. The random selection can be clustered and stratified, i.e., focusing on more important regions and/or districts from the perspective of the HSS program(s) (both programmatically and financially), and then randomly selecting several delivery points of each type (e.g., from a sample of hospitals, and, separately, from a sample of primary health facilities). Still, a site may be purposefully selected, if there are reasons to do so. Those sites should be reported in the final evaluation report. Per the guideline of the OSDV and/or RSQA, the data collection form service delivery points will be performed in at least two regions, including two districts per region, and with two sites visited in each district (i.e., at least eight sites overall). The estimated level of efforts is 6–12 days.

PHASE 2: DATA COLLECTION AND ANALYSIS

STEP 5. DATA COLLECTION



A. Brief the HSS program management and office of the commissioning organization

B. Upon arrival, the evaluation team will brief the HSS program management and, if relevant, the local office of the commissioning organization. In addition to explaining the scope and timeline of the evaluation, this briefing will be used for arranging any remaining logistical issues.

C. Conduct data collectors' training and piloting of the questionnaire. The next day will be used for training the data collectors on data collection of a structured questionnaire from service delivery points. As detailed in Box 5, the evaluation team will develop a questionnaire to assess the quality of key interventions with linkages to the selected outcome indicators. The questionnaire will be piloted and further refined the following day, after the first day of data collection.

D. Collect data and interview key informants. Most of the fieldwork will focus on data collection from service delivery points through interviews with key informants. An interim analysis of findings and recommendations will be conducted at the end of the first week and in the second week. This is done prior to the stakeholder meeting at the end of fieldwork, at which time findings are validated, recommendations are discussed, and action plans are developed for selected recommendations.

Table 8 provides the recommended schedule of the evaluation team during the fieldwork, as detailed below.

TABLE 8: RECOMMENDED SCHEDULE OF THE EVALUATION TEAM DURING THE FIELDWORK

Week	Day	Schedule
First week	Monday	Briefing of the HSS program management and, if relevant, the local office of the commissioning organization
	Tuesday	Training + piloting, interviews
	Wednesday	Data collection (from service delivery points), final feedback for data collectors, and interviews
	Thursday	Data collection, interviews
	Friday	Data collection, interviews
	Saturday	Interim analysis of findings and recommendations
Second week	Sunday	Day off
	Monday	Data collection, interviews
	Tuesday	Data collection, interviews
	Wednesday	Data collection, interviews, preparation of preliminary findings, and recommendations to discuss with stakeholders
	Thursday	Stakeholder meeting, data collection
	Friday	Interviews and debriefing

STEP 6. ANALYZE THE FINDINGS

A. Conduct process evaluation

The process evaluation consists of two layers:

1. Providing a short description of the four elements related to the performance of the HSS program:
    a. Summary of programmatic performance: The degree to which the HSS program is achieving its programmatic targets (if any) and the challenges it faced.
    b. Summary of financial management: Summary of how the funding is being managed, including the burn rate (money expedited versus the planned budget).
    c. Summary of data quality: Summary of data quality as reported by assessment reports.
    d. Review of overall HSS program management: Based on the above, an overall assessment of the HSS program performance. The better the execution of the HSS program, the higher the likelihood to observe systems and health outcomes.

The above performance review should be based on existing documents, such as progress reports and annual reviews.

2. Identifying the key strengths, challenges, and recommendations for each of the five elements (structure and capacity, program management and processes, M&E, coordination with stakeholders, and alignment and harmonization) from the data collected by the evaluation team and data collectors.

    The evaluation team might decide to select a subset of questions from Annex A to guide the interviews.

B. Assess plausible effects of the HSS program on the outcome indicators

The assessment of the plausible effects of the HSS program on the outcome indicators is detailed in section 4, and includes three areas:

1. Plausible outcome of the HSS program
2. Assessment of the HSS program design
3. Assessment of the data quality of the outcome indicators

The Excel template is designed to facilitate this assessment.

C. Assess HSS program's system-wide effects

This will identify positive effects, negative effects, and recommendations resulting from each SDA that emerged. The evaluation team might decide to select a subset of questions from Annex D to guide the interviews.

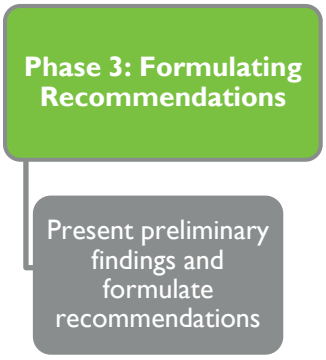
D. Conduct formative evaluation

An analysis of changes in NSP, guidelines, policies, the funding environment, and demographic and disease burden will identify adjustments to the HSS program that might be required and will result in formulating recommendations accordingly.

The Excel template is designed to facilitate this assessment.

PHASE 3: FORMULATION OF RECOMMENDATIONS

STEP 7. PRESENT PRELIMINARY FINDINGS TO STAKEHOLDERS AND FORMULATE RECOMMENDATIONS



Recommendations should be developed from both discussions with key informants and brainstorming with the evaluation team:

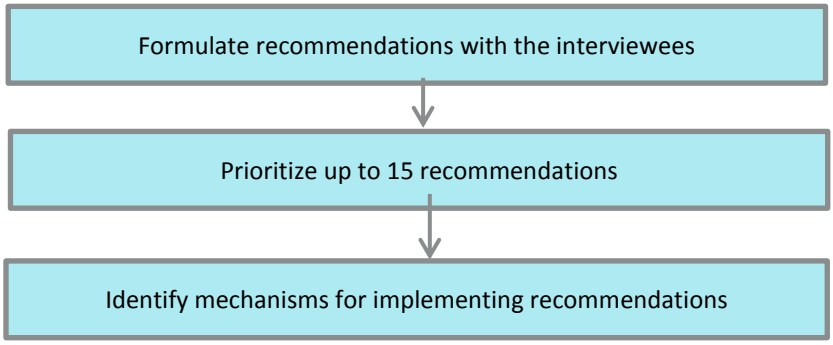
A. Discuss the findings, conclusions, and recommendations in one-on-one interviews and small group discussions

Recommendations will be formulated and tested during the stakeholders' interviews. Once the evaluation team has a preliminary list of findings and recommendations, usually within five days into the country visit, the team should discuss this with the organization commissioning the evaluation, the HSS program management, and those stakeholders to which the recommendations are relevant. The discussion will usually be conducted in small groups in which recommendations will be accepted or rejected, and if accepted, refined to prepare a preliminary work plan per recommendation.

B. Hold daily brainstorming sessions on recommendations discussed with key informants

While a report can include a range of recommendations, turning those recommendations into reality requires the agreement of those who can implement the changes or can implement conditional requirements for future investments in the program. Not all recommendations can be implemented, because of technical or political considerations. It is necessary to assess what issues can be addressed, how to prioritize these issues, and then find the mechanisms through which the recommendations for the prioritized issues can be addressed. This section focuses on formulating the recommendations and prioritizing them, and the mechanisms through which they can be addressed. The workflow and criteria are shown in Figure 9.

FIGURE 9: FORMATIVE EVALUATION WORKFLOW AND CRITERIA



The worksheets of the process evaluation and the assessment of HSS progress and system-wide effects contain columns to list recommendations. Recommendations should follow several criteria:

1. **Legal framework:** Recommendations should be in line with the law of the country. For example, recommendations of a salary increase of health workers to reduce their attrition is unlikely to be feasible if this requires change in the law.

2. **Local context:** It is important to consider the local context when making recommendations. For example, although supportive supervision using mobile devices can be a best practice to improve quality of services and data, this recommendation might need to be adjusted in areas with limited mobile network.
3. **Resources:** Recommendations should be feasible in terms of the financial and human resources available. For example, recommendations of task-shifting from a medical doctor to nurses would be counterproductive in places where a shortage in nurses is more acute than of medical doctors.
4. **Mechanism for implementation:** A recommendation should have a feasible process for it to turn into a reality. This can occur either through a buy-in of relevant stakeholders, who then commit to carry it forward, or through conditioning continuation of some or all of the HSS program in the implementation of a recommendation.

It is important to note that the evaluation team might be notified that some recommendations are already being implemented. The evaluation team must confirm this through review of relevant documentation and, if necessary, through site visits. If a recommendation is indeed implemented, yet is not well communicated to other stakeholders, the evaluation team should use its communication channels to update the stakeholders accordingly.

C. Prioritize recommendations

The evaluation should be solution oriented, in that recommendations are sought for each area that needs to be improved. However, the evaluation team is strongly encouraged to focus on up to 15 recommendations that they wish to discuss in detail with the stakeholders. Prioritization of the recommendations will be based on the recommendation's relevance, on the stakeholders' input during interviews, and on guidance from the organization commissioning the evaluation.

D. Identify mechanisms to implement recommendations

Prioritized recommendations will need to have concrete mechanisms for implementation. There are three major forums for consultation with stakeholders:

1. One-on-one interviews
2. Small discussion groups
3. Stakeholders' meeting toward the end of the fieldwork, where stakeholders are divided into groups, each reviewing three to five recommendations, for which they formulate a work plan with steps and a timeline, and identify individuals to supervise the execution of each step.

Through this stakeholder engagement, work plans are developed, yet they are not binding: it is up to the policymakers of the HSS program as to whether the work plans will be implemented. The organization commissioning the evaluation (if different from the implementer of the HSS program) may decide whether to condition the implementation of some or all of the work plans on the continuation of funding to the HSS program. If requested by the stakeholders, the evaluation team may carry out limited follow-up analyses.

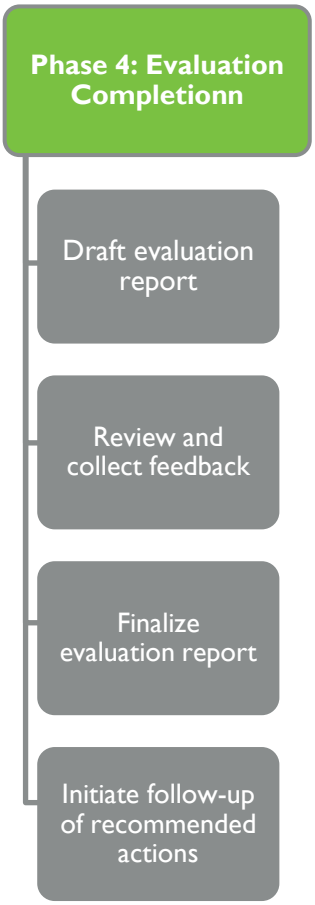
Table 9 shows the potential schedule of such a workshop.

TABLE 9: POTENTIAL SCHEDULE OF DISSEMINATION WORKSHOP

09:00–09:15	Opening remarks
09:15–10:15	A presentation of the major findings and recommendations of the evaluation
10:15–10:30	Coffee break
10:30–12:00	Discussions within small groups, each assigned with a specific set of findings and recommendations. Each group will be presented with the findings and recommendations and then will be asked to review the findings, conclusions, and recommendations and prepare a work plan per recommendation
12:00–12:45	Lunch
12:45–13:45	Reports from groups on identified issues
13:45–14:00	Closing remarks

PHASE 4: EVALUATION COMPLETION

STEP 8. DRAFT EVALUATION REPORT



Within one to two weeks, the evaluation team should complete its review of all of the evaluation documentation and interviews produced during the mission and complete a draft evaluation report with all findings, recommendations, and stakeholders' work plans for the implementation of the recommendations. Any major changes in the evaluation findings made after the closeout meeting in-country should be clearly communicated to the organization commissioning the evaluation and to the implementer of the HSS program. The draft of the evaluation report will be sent to the organization commissioning the evaluation and to the implementer of the HSS program. Table 10 shows the suggested outline for the evaluation report.

TABLE 10: SUGGESTED OUTLINE FOR THE FINAL EVALUATION REPORT

Executive summary

Acknowledgments

1. Introduction

- 1.1 Purpose of the evaluation report
- 1.2 Background on the country
- 1.3 Background on the HSS program

2. Methodology

- 2.1 Methodology
- 2.2 Data collection
- 2.3 Key informant interview and stakeholder engagement

3. Findings, recommendations, and work plans

- 3.1 Assessment of the implementation of the HSS program
- 3.2 Plausible outcome of the HSS program
- 3.3 Assessment of the HSS program design
- 3.4 Assessment of the data quality of the outcome indicators
- 3.5 Assessment of the system-wide effects
- 3.6 Assessment of the relevance of the HSS program

4. Summary of findings, conclusions, and recommendations

References

STEP 9. REVIEW AND COLLECT FEEDBACK FROM COUNTRY AND COMMISSIONING ORGANIZATION

To build consensus and facilitate data quality improvements, the evaluation team needs to share the draft evaluation report with the organization commissioning the evaluation and with the implementer. The latter will be given an opportunity to provide feedback on the evaluation findings. The review period should be limited to two weeks unless otherwise agreed upon. This response will need to be included in the final evaluation report.

STEP 10. FINALIZE EVALUATION REPORT

Once the organization commissioning the evaluation and the implementer have reviewed the draft evaluation report and provided feedback, the evaluation team will complete the final evaluation report. While the evaluation team should elicit feedback, it is important to note that the content of the final evaluation report is determined by the evaluation team exclusively. Once the report is finalized, it should be disseminated.

STEP 11. INITIATE FOLLOW-UP OF RECOMMENDED ACTIONS

The HSS program management will be expected to send follow-up correspondence once the agreed-upon changes/improvements have been made. If the organization commissioning the evaluation wants the evaluation team to be involved in the follow-up of identified strengthening measures, an appropriate agreement may be reached. The organization commissioning the evaluation and/or the evaluation team should maintain a "reminder" file to alert itself as to when these notifications are due.

ANNEX A: PROCESS EVALUATION CATEGORIES AND QUESTIONS

Annex A contains a list of questions, per element and sub-element for five areas covered by the process evaluation: structure and capacity, program and financial management and processes, monitoring and evaluation, coordination with stakeholders, and alignment and harmonization.

These are guiding questions, which aim to direct the evaluation team toward key strengths and challenges. There is no need to answer every question; rather, the questions are intended to guide the process evaluation toward its main goal: to provide the management of the HSS program and its stakeholders feasible, concrete recommendations to enhance the program. By engaging with stakeholders, a subset of these recommendations will be broken down into action items, each with a timeframe, responsibilities, and, if possible, an estimated cost.

STRUCTURE AND CAPACITY

1. **Organizational structure and accountability:**
    a. Does the program management have clear organizational leadership with clear accountability required to implement the proposed program?
    b. Does the program management have a governing board or governing body outside of its employees? Describe their function and evaluate how they help or hinder the function of the program management.
    c. Is there an organizational chart? Are the responsibilities in the chart clear?
    d. Examine two or three poor performing areas or failures: Describe them briefly. What were the factors that led to the failure?
    e. Examine two or three strengths/successes: Describe them briefly. What were the factors that led to the success?
2. **Program management's ability, systems, and resources:**
    a. Does the program management have personnel with adequate expertise for managing outsourced activities? Are there adequate human resources, financial, and communication systems and plans for managing outsourced activities?
    b. Review the program management's general criteria for evaluating subcontractors: Is there transparent competition; what steps are taken to ensure that the selected subcontractors have a good track record?
    c. Does the program management have the ability to arrange technical assistance to subcontractors where needed? Provide concrete examples.

PROGRAM AND FINANCIAL MANAGEMENT AND PROCESSES

3. **Financial management system:**
    a. Does the program management have a financial management system that can correctly record all transactions and balances?
    b. Do the expenditures in the financial management system correspond to those reported to the funding sources?
4. **Safeguards:** What safeguards exist to prevent loss, waste, fraud, and abuse of funding and of HSS programming assets? Examples include existence of robust financial management and audits at both the program management and subcontractor levels; monitoring of pricing of items purchased by program management and subcontractors; contracts with subcontractors, which are performance/deliverable oriented rather than related to level of effort; proper documentation of expenses; and sign-off of medium and large expenses by at least two persons.
5. **Operations of program management:**
    a. Does the program management base its operations on an annual plan (e.g., work plans, M&E, etc.) as part of routine management actions and benchmarking (e.g., key performance indicators, annual plan budgets)?
    b. Does the nominated program management share these documents internally and with key partners?
6. **Operations and supervision of subcontractors:**
    a. Does the program management have operational procedures and plans in place for managing subcontractor s, including plans for monitoring the program implementation at subcontractor level and reviewing subcontractors' financial and program reports for completeness and technical soundness?
    b. Does the program management review documentation of the subcontractor's performance assessment, and does the review include clear recommendations for improvements, and whether the subcontractor's performance improved over time?

MONITORING AND EVALUATION

7. **M&E plan of the program:**
    a. Has the program management compared the M&E plan with the HSS program's results framework and assessed the extent to which the indicators in the results framework are captured in the M&E plan, including definition and measurement methods?
    b. Is the HSS program aligned with the national M&E plan? Has the program management compared the activities in the HSS program's M&E plan with those in the national M&E plan?
8. **Alignment with the national M&E unit:** Are results shared between the program M&E unit and the national M&E system? Is there any duplication in data collection? Are there discrepancies in the results reported?

9. **Capacity of the M&E unit and its information system:** Does the current M&E unit and its information system have the capacity and dedicated personnel to collect and provide programmatic reports for the proposed program, including reports from subcontractors in a regular, reliable, and quality manner?

COORDINATION WITH STAKEHOLDERS

10. **Key stakeholders and beneficiaries:** Ask interviewees to list the key stakeholders and beneficiaries of the program.
11. **What is the roles of each key stakeholder:**
    a. Are the stakeholders engaged in policy formulation, and if so how? Please provide two examples of successes or failures in the stakeholder engagement.
    b. What are the troubleshooting solutions for "difficult" stakeholders?
12. **Communication mechanisms:**
    a. What are the available communication mechanisms in place for facilitating communication and coordination between the program and its various stakeholders (monthly meeting, regular email updated – as for documentation)?
    b. To what extent are these mechanisms helpful for implementation?
13. **Health of the partnership:**
    a. What is the "health of the partnership"?
    b. Provide one or two examples where the partnership solved an issue and/or created a problem.

ALIGNMENT AND HARMONIZATION

14. **Alignment and harmonization:** Based on the answer to the question on key stakeholders and beneficiaries, review two or three potential overlaps between the various donors and international organizations (e.g., UNAIDS, WHO, World Bank, USAID, United Kingdom Department of International Development) and assess if there is an actual overlap, its reasons, and how this is being handled.

ANNEX B: EXAMPLES OF THE ASSESSMENT OF THE QUALITY OF INTERVENTION

Example 1 – Outcome indicator: percentage of health facilities that submit reports to the national level on time.

In this example, three interventions were found to be logically linked to the above outcome indicator:

a. Conduct an assessment of the information needs of laboratories and policymakers:

- **Planning: Was a written scope for the assessment/terms of reference (ToR) developed prior to the assessment? Was the selection process of the assessors robust?** Was there a buy-in for the assessment and recommendations?
- **Structure: Did qualified assessors conduct the assessment? Were sufficient resources available?**
- **Process:** Was an assessment report produced? Does it include a clear and robust methodology? Did it include both strengths and challenges of information needs of laboratories? Were the recommendations focusing on comprehensive, yet relevant information needs of laboratories? Were the findings and recommendations communicated effectively?

b. Design a modular integrated lab information database situated at the reference lab and revise and modify the database after piloting:

- **Planning:** Was a quality ToR prepared for the scope of the design?
- **Structure: Was the design done by qualified personnel?** Did the design adapt to the existing information technology capacity of the program? Was the design tailored to the existing information management of the program?
- **Process:** Did the design take into consideration the assessment of the information needs? Was it done in consultation with all key stakeholders?

c. Prepare guidelines and procedures for using the web-based lab information database:

- **Planning:** Was a quality ToR prepared for the scope of the development of the guidelines? Did the guidelines and procedures take into consideration the assessment of the information needs? Was planning done in consultation with all key stakeholders and was their buy-in obtained?
- **Structure:** Were the guidelines written by qualified personnel?
- **Process:** Are the guidelines clear and tailored to the target population? Do the guidelines include best practices in reporting of lab information? Were the guidelines and procedures communicated effectively? Did all the labs that the evaluation team visited have a copy of the guidelines and were they able to refer to those guidelines?

Example 2 – Outcome indicator: percentage of births attended by trained health personnel .

In this example, one indicator – number of referrals obtained by community health workers of pregnant women to receive transportation costs to access health centers – was found to be logically linked to the above outcome indicator:

- **Planning:** Was a protocol written for implementing this intervention, including details on how much to pay and who is entitled to these benefits? Was the intervention piloted?
- **Structure:** Were community health workers trained on delivery? Is this intervention properly funded? Is it sustainable? Is a supportive supervision mechanism in place?
- **Process:** Did the pregnant women receive sufficient, yet not excessive, reimbursement for their transportation? Was the reimbursement for the transportation cost also provided for those delivering at the clinic? Was the transportation reimbursement given sufficiently in advance of the need, or at least promised, in a manner that the pregnant women knew that the transportation cost would not be a barrier for them to deliver in a clinic? Was the transportation reimbursement given mainly to those that otherwise would not be able to afford to deliver at a distant health facility?

ANNEX C: TIMELINE OF REPORTING ON RESULTS FROM VARIOUS SOURCES

Category	Source of Results	Duration between Target and Report Due Date	Comment on the Delay
Cross disease	Census	13.5 months	12 months to compile results + 45 days for verification/sign-off
Cross disease	Community health workers records	45 days	45 days for verification/sign-off
Cross disease	Demographic Health Survey (DHS)	13.5 months	12 months to compile results + 45 days for verification/sign-off
Cross disease	Facility records	45 days	45 days for verification/sign-off
Cross disease	Health information systems (HIS)	45 days	45 days for verification/sign-off
Cross disease	Modeling	45 days	45 days for verification/sign-off
Cross disease	Multiple Indicator Cluster Survey (MICS)	13.5 months	12 months to compile results + 45 days for verification/sign-off
Cross disease	Population-based survey	13.5 months	12 months to compile results + 45 days for verification/sign-off
Cross disease	Population-based survey with verbal autopsy	19.5 months	12 months to compile results + 6 months for verbal autopsy + + 45 days for verification/sign-off
Cross disease	Program monitoring	45 days	45 days for verification/sign-off
Cross disease	Second generation surveillance	45 days	45 days for verification/sign-off
Cross disease	Surveillance	45 days	45 days for verification/sign-off
Cross disease	Vital registration	3 months	45 days to compile results + 45 days for verification/sign-off
HIV/AIDS	AIDS Indicator Survey (AIDS)	6.5 months	5 months for preliminary report in less optimal scenario + 45 days for verification/sign-off
HIV/AIDS	Behavioral survey (BSS, FHI)	4.5 months	3 months to compile results + 45 days for verification/sign-off; Reports are available approximately 2–3 months after field work is completed
HIV/AIDS	HIV sentinel surveillance	45 days	45 days for verification/sign-off
HIV/AIDS	Priorities for Local AIDS Control Efforts (PLACE)	7.5 months	6 months to compile results + 45 days for verification/sign-off; Reports are available typically within 2–3 months
Malaria	Coverage of IRS program data	13.5 months	12 months for estimating population/structures in need + 45 days for verification/sign-off

Category	Source of Results	Duration between Target and Report Due Date	Comment on the Delay
Malaria	Malaria Indicator Survey (MIS)	6.5 months	6 months to compile results + 45 days for verification/sign-off
Malaria	Reach of IRS program data	45 days	45 days for verification/sign-off
TB	TB prevalence survey	13.5 months	12 months to compile results + 45 days for verification/sign-off
TB	Case detection rate	7.5 months	6 months to compile results + 445 days for verification/sign-off; align to TB global report
TB	Notification rate from surveillance system	7.5 months	6 months to compile results + 45 days for verification/sign-off
TB	Treatment success rate	19.5 months	12 months until the last patients of the cohort report successful treatment + 6 months to compile results + 45 days for verification/sign-off; align to TB global report

ANNEX D: GUIDING QUESTIONS TO ASSESS THE SYSTEM-WIDE EFFECTS OF THE HEALTH SYSTEMS STRENGTHENING PROGRAM

Annex D contains a list of questions, per SDA. These are guiding questions, which aim to direct the evaluation team toward key strengths and challenges. There is no need to answer every question; rather, the questions are intended to guide the assessment toward its main goal: to provide feasible, concrete recommendations to enhance the program's positive effects on the wider health system and mitigate the negative ones. By engaging with stakeholders, a subset of these recommendations will be broken down into action items, each with a timeframe, responsibilities, and, if possible, an estimated cost.

Each of the questions below should be answered in relation to how the HSS program strengthened or weakened the examined HSS element in the question.

HEALTH INFORMATION SYSTEMS (HIS) INTERVENTIONS

If and how the HSS program strengthened or weakened the following:

Quality

1. Financial and physical resources to support HIS in central and district budgets?
2. HIS-related strategies, plans, guidelines, and procedures?
3. Availability and accessibility of data sources?
4. Timeliness, completeness, reliability, and accuracy of HIS data from the public sector?
5. Timeliness, completeness, reliability, and accuracy of HIS data from the private sector?
6. Qualified human resources to operate, compile, and analyze health information (recruitment and training)?
7. Transparency of health information?
8. Feedback of results of analyses to data providers, allowing them to improve their performance and data quality?

Efficiency and sustainability

9. Use of health information for planning, budgeting, or fundraising interventions (e.g., a change in budget levels in response to new information that the data reveals, fund allocation/budgeting proposals utilizing HIS data for advocacy)?
10. Mechanism to review the utility of current HIS indicators for the planning, management, and evaluation process, and to adapt and modify accordingly?

11. Integration of vertical HIS with the HIS of the wider health systems?

FACILITY MANAGEMENT AND ORGANIZATION

If and how the HSS program strengthened or weakened the following:

Accessibility and equity

12. Access of marginalized groups to health services?
13. Quality of patient visits (wait time and frequency)?
14. The referral system?
15. Operating hours for public and private health service providers (including establishment or modifications in the appointment system)?

Efficiency and sustainability

16. Duplications in roles and responsibilities?
17. Organizational structure for facility management (e.g., increase virtualization or promote integration)?

Quality

18. Quality of services provided in the facility?
19. Implementation of clinical standards?
20. Supportive supervision of health services and other quality measures (e.g., mentoring)?

Infrastructure

If and how the HSS program strengthened or weakened the following:

Accessibility and equity

21. The physical accessibility of the population to health services (including accessibility to disadvantaged populations)?

Sustainability and quality

22. Infrastructure of health facilities?
23. Equipment at health facilities?
24. Maintenance and upkeep of facilities?

PROCUREMENT AND SUPPLY CHAIN MANAGEMENT

If and how did the HSS program strengthened or weakened the following:

Quality

25. The system/s for pharmaceuticals registration, post-marketing surveillance, pharmacovigilance, and control?
26. The selection of national essential medicines and maintaining of the national essential medicines list?
27. The National Essential Medicines Policy (NMP) or other government documents that sets objectives and strategies for the pharmaceuticals sector?

28. Any laws, regulations, and standard operational procedures in regards to pharmaceuticals?
29. The National Drug Regulatory Authority responsible for the promulgation and enforcement of regulations?
30. Mechanisms for prescribing and dispensing pharmaceuticals?
31. Stock-outs?
32. Storage and distribution?

Efficiency

33. Efficiency measures, e.g., use of generic drugs, planned procurement (versus emergency procurements, which are more costly)?
34. The procurement process, including the use of competitive bids?
35. The inventory mechanism?
36. Inventory loss?

Sustainability

37. Recurrent cost of pharmaceuticals covered through domestic sources (to ensure continuation of services in case of decline in donor support)?

Accessibility

38. Licensing provisions or incentives for private wholesalers and retailers?
39. Out-of-pocket expenditures for medicines?

HEALTH WORKFORCE

If and how the HSS program strengthened or weakened the following:

Accessibility and Equity

40. Ratio of health workers to the population, by cadre?
41. Distribution of health workforce, and its effects on the services provided/available to disadvantaged populations?

Quality

42. The hiring process?
43. Pre-service and in-service training, including the training quality and the extent to which it responds to the needs of the health care system?
44. Compensation, including competiveness in the local and regional labor market?
45. Human resources management?
46. Availability and use of HRH information systems?
47. HRH-related planning, policies, and strategies?

Efficiency

48. Mechanisms used to monitor and improve health workers' performance, productivity, and expectations?

Sustainability

49. Sustainability of HRH costs, including the extent of concurrent HRH costs funded by donors?

50. Integration of health workers in vertical programs to the wider health system?

HEALTH FINANCING

If and how the HSS program strengthened or weakened the following:

Sustainability

51. The ability of the government to sustain health services without donor funding?

Quality

52. The ability to realize planned budgets?

53. The process of budget formulation?

Efficiency

54. Budget allocation between the central and local government?

55. Local-level spending authority and capacity?

56. Budget allocation mechanisms that promote more efficient spending of public resources at the local level?

57. The oversight to maximize expenditure of limited resources?

58. Payment mechanisms to improve providers' efficiency?

Accessibility and equity

59. Contracting mechanisms between the MoH and public or private service providers?

60. Out-of-pocket, including policies for user fee payments in the public sector (most importantly, user fees and waivers for disadvantaged populations), and allocation of user fee revenues?

61. Health insurance, including population coverage (including disadvantaged populations), a service covered, and provider payment mechanism?

STEWARDSHIP AND GOVERNANCE

If and how the HSS program strengthened or weakened the following:

Quality

62. The national/regional/local oversight framework?

63. The program's capacity to promote certain health issues in policies, plans, and budgets for health services?

64. Involvement of stakeholders in overseeing service delivery?

65. Reliance on evidence in policymaking and planning?

66. Does government regulate the private health sector?

Accessibility

67. Involvement of stakeholders (civil society organizations, private sector infected and affected people, donors, implementers, etc.) in the decision making?

68. Responsiveness to stakeholders?

Efficiency

69. Procedures for reporting, investigating, and adjudicating misallocation or misuse of resources?