

Health Systems Strengthening

Monitoring, Evaluation, and Learning Guide

Anwer Aqil
Eva Silvestre
David Hotchkiss
Lisa Maniscalco



Health Systems Strengthening

Monitoring, Evaluation, and Learning Guide

Anwer Aqil, MD, MCPS, MPH, DrPH, Office of Health Systems, United States Agency for International Development

Eva Silvestre, PhD, MEASURE Evaluation

David Hotchkiss, PhD, MEASURE Evaluation

Lisa Maniscalco, MPH, Office of Health Systems, United States Agency for International Development

September 2017

MEASURE Evaluation

University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517 USA

Phone: +1 919-445-9350

measure@unc.edu

www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement AID-OAA-L-14-00004 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. TR-17-167c

ISBN: 978-1-9433-6455-8



ACKNOWLEDGMENTS

Karen Cavanaugh, former Director, Office of Health Systems, Bureau for Global Health (GH/OHS) at the United States Agency for International Development (USAID), provided the overall leadership, direction and inspiration for the preparation of this guide. Kathryn Panther, former Deputy Director of the GH/OHS, supported the team to keep moving and complete the task at hand.

Special thanks to Kelly Saldana, Director, GH/OHS/USAID, for providing guidance on revising chapter four on acquisition and assistance solicitation, summarizing key messages, and simplifying language and format.

We highly appreciate the comments by our reviewers at USAID: Diana Harper (Office of Policy, Programs, and Planning/GH), Amani Selim (Office of Population and Reproductive Health/GH), David Jacobstein (Bureau of the Democracy, Conflict, and Humanitarian Assistance/Democracy, Human Rights and Governance Cross-Sectoral Programs Team), Joseph Naimoli and Robert Emery (GH/OHS), and Danielle Pearl and Kristin Saarlus (Office of Learning, Evaluation and Research, in the Bureau for Policy, Planning and Learning). Their feedback made the guide better.

We thank the MEASURE Evaluation Knowledge Management team at the University of North Carolina at Chapel Hill for their editorial and production assistance.

CONTENTS

ACKNOWLEDGMENTS	3
ABBREVIATIONS	8
EXECUTIVE SUMMARY	10
Background	10
Purposes.....	10
Contents of This Guide.....	10
<i>Chapter 1. Introduction</i>	10
<i>Chapter 2. Background of Health Systems Strengthening</i>	11
<i>Chapter 3: MEL Section of USAID’s Project Appraisal Document</i>	11
<i>Chapter 4: MEL Requirements in Acquisition and Assistance Solicitation Documents</i>	12
<i>Chapter 5: Review of the MEL Implementation Plan</i>	12
<i>Chapter 6: Monitoring Implementation of the MEL Plan to Build Evidence</i>	12
<i>Chapter 7: Building the Evidence Base for Health Systems Strengthening through Evaluation</i>	13
CHAPTER 1. INTRODUCTION	14
Context of the Guide.....	14
Audience	14
Purposes.....	14
<i>Operational Guidance</i>	15
<i>Addressing Complexity in Health Systems</i>	15
<i>Evidence Generation</i>	17
<i>Learning to Inform Decision Making</i>	18
Resources Used to Develop the Guide.....	19
Organization of the Guide.....	19
CHAPTER 2. BACKGROUND OF HEALTH SYSTEMS STRENGTHENING	20
Introduction	20
What Is a Health System?	20
What Is Health System Performance?	21
What Is Health Systems Strengthening?.....	22
Systems Thinking for HSS.....	22
The Implications of HSS for M&E.....	23
Linking HSS MEL with USAID’s Program Cycle.....	25
Key Messages	26
CHAPTER 3. MONITORING, EVALUATION, AND LEARNING SECTION OF USAID’S PROJECT APPRAISAL DOCUMENT	27

Introduction	27
Articulate and Review Theory of Change (TOC)	27
Prepare and Review a Results Framework	29
Select and Review Illustrative Performance Indicators	32
Select and Review Systemwide Effects and Indicators	33
Select and Review Context Information.....	34
Guide and Review the Selection of Sources of Data Collection	35
Preparation and Review of the Logical Framework	36
Prepare and Review Preliminary Evaluation Questions and Their Implications	38
Learning Plan	39
Sustainability of Results and Local Ownership	39
Key Messages	39
CHAPTER 4. MONITORING, EVALUATION, AND LEARNING REQUIREMENTS IN ACQUISITION AND ASSISTANCE SOLICITATION DOCUMENTS	40
Introduction	40
MEL in the A&A Documents.....	40
<i>The Importance of MEL in the A&A Solicitation Document</i>	<i>40</i>
<i>MEL Issues to Address in the A&A Solicitation Document</i>	<i>40</i>
Reviewing MEL Sections of Proposals.....	43
Key Messages	43
CHAPTER 5. REVIEW OF THE MONITORING, EVALUATION, AND LEARNING IMPLEMENTATION PLAN.....	44
Introduction	44
Reviewing the MEL Plan’s Objectives.....	44
Reviewing the TOC	45
Reviewing the RF	46
Reviewing Monitoring Information as per the A&A Agreement.....	46
<i>Performance Indicators</i>	<i>46</i>
Reviewing Sources of Information: from What to How and Why	50
Data Collection Methods	52
Data Collection Frequency	52
Outcome Monitoring Techniques	53
Setting up Data Collection and Management	57
<i>Data Collection Process, Entry, and Cleaning.....</i>	<i>57</i>
<i>Data Quality Assessment.....</i>	<i>57</i>
<i>Maintaining Privacy and Confidentiality of Data</i>	<i>58</i>

<i>Analysis Plan</i>	58
<i>Open Data Policy</i>	58
<i>Roles and Responsibilities</i>	58
Establishing Baseline and Targets	59
Reviewing the Approach to Promote the Use of Collected Data at Different Levels.....	61
Reviewing the Evaluation Plan and Additional Studies	62
Key Messages	62
CHAPTER 6. MEL IMPLEMENTATION: BUILDING EVIDENCE THROUGH MONITORING	63
Introduction	63
Monitoring Report	63
<i>Performance Tracking</i>	63
<i>Reporting Systemwide Quantitative Indicators</i>	68
<i>Reporting Contextual Factors</i>	68
<i>Reporting Systemwide Changes/Effects Using Outcome Monitoring Techniques</i>	69
<i>Monitoring Unintended Results</i>	69
Reporting the Use of Information.....	71
Synthesizing Data and Creating Evidence.....	71
Data Quality Assessments.....	73
Learning and Adapting	74
Key Messages	75
CHAPTER 7. BUILDING THE EVIDENCE BASE FOR HEALTH SYSTEMS STRENGTHENING THROUGH EVALUATION	76
Introduction	76
Importance of HSS Evaluations	76
What Is Unique about HSS Evaluation?	76
Types of Evaluations	78
<i>Impact Evaluations</i>	78
<i>Performance Evaluations</i>	78
Planning for Evaluations.....	78
Evaluation SOW.....	79
Evaluation Questions	79
Conceptual Framework for Evaluation	81
Choosing an Evaluation Design	83
Data Collection Methods	85
Data Analysis Plan.....	85

Deliverables	85
Team Composition	85
Bidding for Evaluations	86
Evaluation Design and Work Plan	86
Implementing the Evaluation.....	87
Managing the Evaluation	87
Analysis, Report Preparation, Utilization, and Dissemination.....	87
<i>Analysis</i>	87
<i>Report Preparation</i>	88
Utilization	88
<i>Dissemination</i>	88
Key Messages	89
Summaries of Illustrative Impact Evaluations.....	90
<i>Example 1. Paying Primary Healthcare Centers for Performance in Rwanda</i>	90
<i>Example 2. Contracting Primary Healthcare Services to Improve and Increase Coverage in Cambodia</i>	91
<i>Example 3. Using Public Policy to Improve the Use of Health Services by the Poor</i>	92
REFERENCES	93

ABBREVIATIONS

A&A	Acquisition and Assistance
ADS	Automated Directive Systems
ANC	antenatal care
CDCS	country development cooperation strategy
CHIS	community health information system
DHIS	district health information system
DHS	Demographic and Health Survey
DQA	data quality assessment
eHMIS	electronic health management information system
GIS	geographic information system
HEW	health extension worker
HFG	Health Finance and Governance
HIS	health information system(s)
HMIS	health management information system(s)
HMN	Health Metrics Network
HRH	human resources for health
HRIS	human resources information system(s)
HS	health system(s)
HSS	health systems strengthening
IMCI	integrated management of childhood illnesses
IP	implementing partner
IR	intermediate result
IRS	indicator reference sheet(s)
ITN	insecticide-treated bednet
LF	logical framework
LMIS	logistics management and information system
M&E	monitoring and evaluation
MEL	monitoring, evaluation, and learning
MICS	Multiple Indicators Cluster Survey
MSC	most significant change
NHA	National Health Account
OHS	Office of Health Systems

PAD	Project Appraisal Document
PEPFAR	United States President's Emergency Plan for AIDS Relief
PNC	postnatal care
PRISM	Performance of Routine Information System Management
RF	results framework
RFA	request for applications
RFP	request for proposals
RHIS	routine health information system(s)
ROA	rapid outcome assessment
SDG	Sustainable Development Goal
SDH	socioeconomic determinants of health
SMART	specific, measurable, achievable, realistic, and time-bound
SNNPR	Southern Nations, Nationalities, and Peoples Region
SOW	scope of work
TOC	theory of change
TT2	two tetanus toxoid injections
UHC	universal health coverage
USAID	United States Agency for International Development
USG	United States Government
WHO	World Health Organization

EXECUTIVE SUMMARY

Background

As the number of health systems strengthening (HSS) projects funded by the United States Agency for International Development (USAID) increases, so, too, does the need to build the capacities of staff at the missions and at headquarters to plan, manage, and conduct monitoring and evaluation (M&E) of HSS projects. Evidence on how HSS interventions¹ strengthen the performance of health systems (HS) and contribute to sustainable improvements in health status are scarce and scattered, with limited dissemination. Health systems are multifaceted and multilayered. The interactions among people, institutions, processes, and outputs makes these systems complex. The complexity arises not only from whether the interventions are simple or complex, practical, and economically feasible, but also from a lack of clarity on causal linkages between HS interventions and the system's performance. HSS monitoring, evaluation, and learning (MEL) is different from the MEL of other types of projects because of its systemic nature, interactions among HS components, and the need to generate evidence on causal pathways and linkages. The guide is complemented by a comprehensive list of HS indicators ([Health Systems Strengthening: A Compendium of Indicators](#)) and an HSS MEL needs assessment ([Health Systems Strengthening: A Literature Review](#)).

Purposes

This guide fulfills USAID's staff capacity-building HSS MEL needs, by providing step-by-step instructions on planning, implementing, and evaluating an HSS project. It has the following purposes:

- Provide operational guidance on planning, implementing, and evaluating HS MEL activities
- Address the complexity of HS in MEL activities
- Describe methods/techniques to generate evidence on the effectiveness of HSS interventions in improving HS
- Apply evidence from HSS MEL to learning, adaptive management, and designing HS projects

Contents of This Guide

This guide's overarching intent is to encourage the design and implementation of the project MEL component using existing resources such that the project's capacity to generate evidence around achievements, systemwide changes, and learning is enhanced. Summaries of each chapter follow.

Chapter 1. Introduction

This chapter introduces the context, audiences, and purposes of the HSS MEL guide. It provides information on what resources were used to develop the guide and instructions to facilitate readers' selection of chapters according to their needs.

Those familiar with background literature in HSS can start from Chapter 3.

¹ USAID (2009) defines interventions as "an action or entity that is introduced into a system to achieve some result. In the program evaluation context, an intervention refers to an activity, project or program that is introduced or changed (amended, expanded, etc.)."

Chapter 2. Background of Health Systems Strengthening

This chapter defines the concepts of HS, HS performance, HSS, and systems thinking. It outlines why HSS MEL is uniquely different from the MEL of other projects. Key messages are:

- Defining the concepts of HS, HS performance, HSS, and systems thinking is important for creating boundaries for HSS MEL.
- HSS is inherently systemic. Therefore, HSS MEL has the following requirements:
 - HSS MEL needs to identify and measure direct and indirect causal linkages among HSS interventions, performance, and impact.
 - HSS MEL needs to detect the systemic changes in other HS functions, as well as changes in organization and in people's relationships, roles, rules, and resources.
 - To understand systemwide effects, HSS MEL must account for interaction effects among HS interventions in one function and other HS functions, including unintended outcomes of interactions.
 - Contextual monitoring needs to be incorporated to exclude alternative explanations for change.
 - Outcome monitoring must be used to explore causal pathways and linkages.
 - HSS MEL needs to be flexible to adjust to a project's emerging needs.
- HSS MEL is based on USAID's program cycle, which emphasizes what is needed for HSS MEL under each phase (plan, implement, monitor, evaluate, and learn across these phases).

Chapter 3: MEL Section of USAID's Project Appraisal Document

USAID's Project Appraisal Document (PAD) is the foundation of the USAID program cycle. The MEL section of this document lays out how performance, context, systemwide effects, and outcomes will be monitored. This chapter's key messages are:

- Prepare or review the HSS PAD by studying the theory of change (TOC), which provides the rationale for the MEL activities.
- Ensure that the HS intervention's interactions with other HS functions are depicted in the TOC and in associated results and logical frameworks.
- Ensure that the HSS MEL section discusses collection of systemwide changes and contextual information in addition to the required outcome/impact and performance indicators and sources of information.
- Verify that the prospective contractor will be required to identify at least one outcome monitoring technique as part of the MEL plan to track systemwide project changes.
- Describe whether an impact evaluation is needed to assess an innovative intervention, untested ideas, or tested ideas in a new context.
- Explain learning plan, sustainability, and local ownership in the MEL section.

Chapter 4: MEL Requirements in Acquisition and Assistance Solicitation Documents

USAID's Acquisition and Assistance (A&A) solicitation documents include MEL considerations delineated in the PAD. The A&A solicitation documents will lay out the MEL expectations for potential implementers, such as the following:

- Collect performance indicators and systemwide and context information.
- Conduct outcome monitoring to explore intermediate outcomes, systemwide changes, and unintended consequences.
- Explore causal pathways between the HSS intervention and its outcomes or impact.
- Assess the intervention's interaction with relevant HS functions.
- Explain the learning plan, sustainability, and local ownership.
- Describe the roles and relationship of the project implementing partner (IP) with the IP conducting the impact evaluation, if there is one.

Chapter 5: Review of the MEL Implementation Plan

Once an activity is awarded, the IP develops a MEL plan for it. This chapter focuses on reviewing the MEL plan to make sure it aligns with the HSS MEL principles described in previous chapters. Considerations for this review are the following:

- Ensure that the activity's TOC and results framework (RF) are aligned with each other. Encourage the development of a separate TOC for each subactivity to facilitate better implementation and MEL.
- Confirm that the performance, systemwide, and context indicators are described in alignment with results and that they have appropriate sources of information.
- Verify that systemwide effects, unintended results, and causal linkages are captured using at least one outcome monitoring approach.
- Confirm that the collection, management, quality assessment, triangulation, and validation of data are described in the MEL plan.
- Ensure that an analysis plan is available to generate evidence in a way that excludes alternative explanations from various sources.
- Ensure that the use of information is part of the MEL plan.
- Determine that baseline and target setting is part of the plan for tracking progress.
- Ensure that the plan includes the rationales for types of evaluations and special studies.

Chapter 6: Monitoring Implementation of the MEL Plan to Build Evidence

This chapter outlines the steps in ensuring that the MEL plan is implemented as designed. It also describes how MEL results can be logged in the activity monitoring report. Here are the key steps:

- Report performance tracking results, including selected systemwide quantitative indicators.
- Report contextual monitoring results.
- Report outcome monitoring: systemwide effects, unintended consequences.
- Document that information is used at different levels for policy and management decisions.
- Synthesize data for evidence generation.
- Prepare a data quality assessment report, when needed.
- Identify and synthesize what has been learned from monitoring.

Chapter 7: Building the Evidence Base for Health Systems Strengthening through Evaluation

The final chapter deals with evaluations of HSS interventions. The key messages from this chapter are as follows:

- Start planning an evaluation—especially an impact evaluation—in the project design phase.
- Ensure that conditions are conducive to an impact evaluation of an HSS intervention.
- Prepare a conceptual framework for evaluation based on the TOC and tied to evaluation questions.
- Confirm that documenting systemwide effects and identifying causal pathways are part of the evaluation.
- Select the evaluation type and design, by considering an attribution–contribution continuum.
- Request and review an impact evaluation protocol for approval before implementing the project.
- Make sure that a communication plan for the evaluation findings has been developed.
- Assure that the evaluation’s scope of work (SOW) states how evaluation findings will be used.
- Involve MEL advisors from USAID’s mission and headquarters levels in the design phase of the evaluation.

CHAPTER 1. INTRODUCTION

We will not be successful in our efforts to end deaths from AIDS, malaria, and tuberculosis unless we do more to improve health systems around the world. —United States President Barack Obama, 2009²

Context of the Guide

USAID's *Vision for Health Systems Strengthening* (USAID, 2015a) presents the agency's strategy for an integrated, comprehensive, and holistic approach to improve HS at the country level. An important premise of this document is that weak HS limit the effectiveness of governments and international partners in scaling up the availability and use of health services that are essential to two USAID global health initiatives: ending child and maternal deaths and controlling the HIV/AIDS epidemic. By helping countries strengthen such core HS functions as health financing, human resources for health, and health information, cost-effective interventions for combating disease and other health problems will be more effectively delivered and sustained. This would subsequently contribute to improved health status.

Evidence on how HSS interventions strengthen HS performance and contribute to sustainable improvements in health status is scarce and scattered and its dissemination is limited. To build evidence demonstrating the effectiveness of HSS interventions for preventing child and maternal deaths and controlling the HIV/AIDS epidemic, USAID's Office of Health Systems (OHS) has launched the *Marshaling the Evidence* initiative (USAID, 2016). The USAID *Vision* document calls for greater investments in monitoring and evaluation (M&E) and implementation science research to generate evidence on how to tailor HSS interventions to diverse country contexts.

The HSS MEL Guide, grounded in *USAID's Vision for Health Systems Strengthening* (USAID, 2015a), provides state-of-the-art MEL methods for planning, implementing, evaluating, and generating evidence on HSS programs, projects, and activities and accounting for complexity within and outside the HS. The HSS MEL guide was originally intended to have two appendices—a comprehensive list of HS indicators and a literature review of existing MEL guidance for HSS. However, to avoid an overly lengthy guide, we decided to publish them separately. *Health Systems Strengthening: A Compendium of HS Indicators* (Diana, Yeager, & Hotchkiss, 2017a) is a wide-ranging list of indicators from multiple, internationally recognized sources around all six functions of the health system. It is informative reference material to review before selecting the HS indicators relevant to the project needs. *Health Systems Strengthening: A Literature Review* (Diana, Yeager, & Hotchkiss, 2017b) was conducted to assess the availability of guidance on monitoring and evaluating HSS, and to list and summarize these resources for others in this field.

Audience

The primary audiences for the HSS MEL Guide are the USAID mission and headquarters staff. However, the use of a program cycle for providing MEL guidance is generic, making the guide practical and useful for other HS practitioners, planners, managers, and MEL specialists.

Purposes

The HSS MEL Guide has the following purposes:

- Provide operational guidance on planning, implementing, and evaluating HSS MEL activities.
- Address the complexity of HS in MEL activities.

² "Statement by the President on Global Health Initiative." The White House Press Office. <https://obamawhitehouse.archives.gov/the-press-office/statement-president-global-health-initiative>

- Generate evidence on the effectiveness of HSS interventions in improving HS processes, outcomes, and systemwide changes.
- Apply evidence from HSS MEL to learning, adaptive management, and designing HSS projects.

Operational Guidance

Many HSS projects³ struggle during planning and implementation to link HSS intervention(s) to other HS functions and outcomes. This struggle is compounded by having multilevel (macro/micro) interventions that are intended to have direct (proximate) and indirect (distal) connections to HS outcomes. The gaps in establishing correct linkages (or pathways) among interventions, mediators, and HS outcomes during the design phase of a project make it difficult to identify, track, and measure changes at different levels associated with an HSS project. In addition, the MEL model that depicts linear pathways among inputs, processes, outputs, and outcomes is often used to explore changes in only the primary outcome of an intervention. As a result, the effects of the intervention's interactions with other HS functions and the unintended consequences of the intervention are often overlooked, as are other interventions within and outside of the HS (context) that can affect the targeted outcomes. Because of these issues, teasing out the effects of HSS interventions from those of other factors is challenging. We must find ways to address the complexity in which HS exist. Understanding HS complexity requires M&E systems that are flexible to adapt and use different sources of information to collect performance and contextual data; that employ other outcome MEL techniques to capture systemic and unintended changes; and that make sense from the cloud of complexity.

This HSS MEL guide assumes that MEL is essential at every stage of the USAID project cycle, from planning, to implementation, to evaluation of the project or activity. Therefore, operational guidance is provided for the MEL actions needed at each of these stages.

Addressing Complexity in Health Systems

Health system organization is multifaceted, and the interactions among people, institutions, and HS processes and functions at different levels make the HS complex. This complexity must be addressed for better M&E. USAID Policy, Planning and Learning (PPL) has developed guidance for complexity-aware monitoring (USAID, 2013). This guidance is based on frameworks from Patton (2011) and Snowden and Boone (2007).

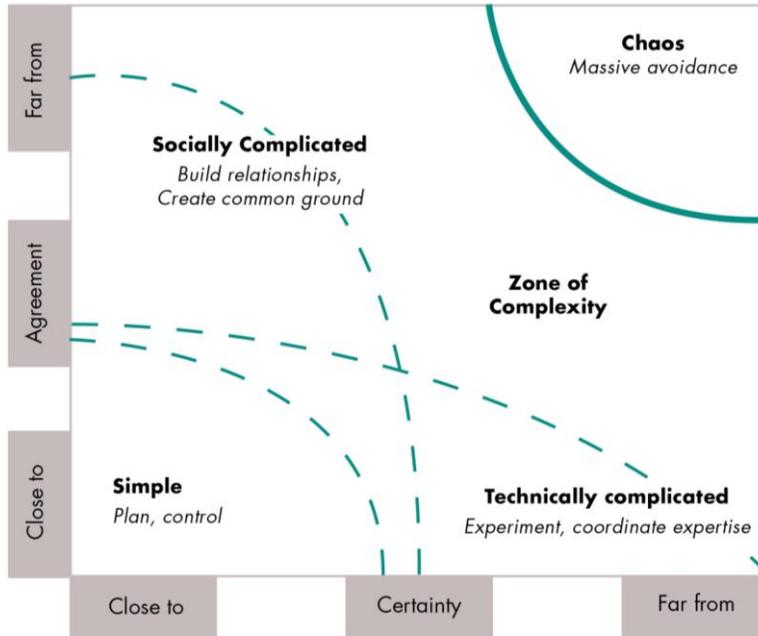
The matrix in Figure 1 (Patton, 2011) states that complexity arises on the one hand from the stakeholders' degree of certainty about the solution to solve a problem, and on the other hand from the extent to which stakeholders agree that the solution can be implemented. Thus, the more certain stakeholders are that a solution is the right one to solve the problem and that the solution is implementable, the less complex the intervention will be, and vice versa. An intervention's level of complexity determines which M&E strategy should be employed to monitor, evaluate, and learn from it.

The framework shown in Figure 2 (Snowden & Boone, 2007) determines the level of complexity of an intervention by appraising whether links between cause (intervention) and effects (outcomes) are easy or difficult to determine. The framework categorizes the conditions for assessing the clarity of linkages between causes and effects as *obvious* (causal linkage is straightforward), *complicated* (multiple causal linkages but possible

³ A USAID project is a set of executed interventions, with an established timeline and budget, intended to achieve a discrete development result (i.e., the project's purpose) by resolving an associated problem. It is explicitly linked to the CDCS Results Framework. An activity is a component of a project that contributes to the project's purpose. An activity refers to an award (such as a contract or cooperative agreement) or a component of a project (such as policy dialogue) that may be undertaken directly by mission staff.

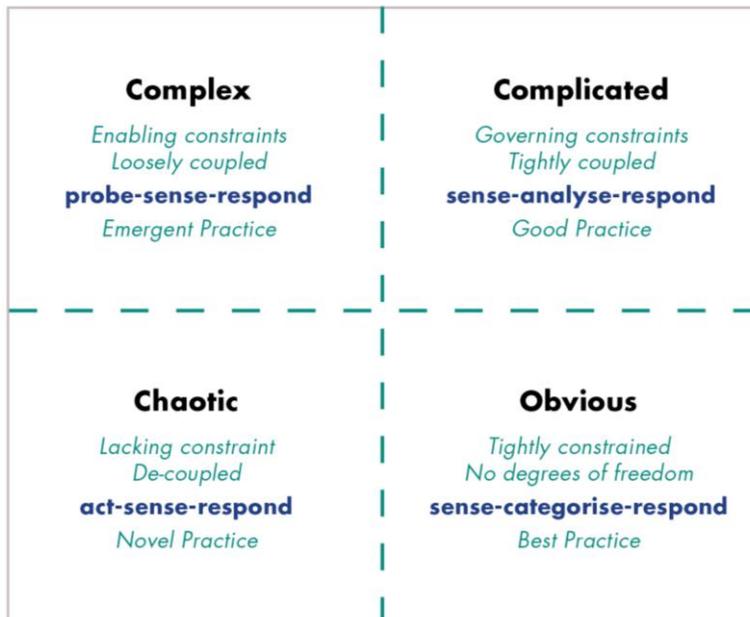
to determine), *complex* (multiple and intertwined linkages that are difficult to determine), and *chaotic* (no pattern of casual linkages discernable). The framework provides strategies to deal with each condition.

Figure 1. Agreement and certainty matrix



Source: Patton, 2011

Figure 2. Cynefin framework



Source: Snowden & Boone, 2007

These frameworks are helpful for understanding HS complexity, as well. For example, some HS interventions such as those related to governance, financing, and health information system functions are at the distal level and affect HS performance indirectly, through better management of service delivery and other HS functions. The causal linkages between these interventions and their effects on health system performance are not clear, leading to challenges in identifying the degree of change in response to an intervention or in attributing a change to a given intervention. In addition, in many low- and middle-income countries, the HS are not fully functional or they have weaknesses in more than one HS function. Thus, when HS interventions are introduced, assuming that a particular intervention is addressing the gap in one function without considering other gaps or how the intervention will affect the interrelationship of health system functions is likely to create uncertainty in identifying causal linkages between intervention and intended outcomes as well as systemwide changes. The contextual factors also augment this complexity. Automated Directive Systems (ADS) 201.3.5.5 promotes the use of complementary monitoring techniques in addition to performance and context monitoring in situations where results are difficult to predict because of dynamic contexts or unclear cause-and-effect relationships (USAID, 2017b). Because HSS interventions are complex, with limited evidence on casual linkages, this guide recommends establishing outcome monitoring⁴ as part of the monitoring for evidence generation.

Evidence Generation

Evidence can be defined in several ways. For this guide's purposes, it is defined as information that substantiates the effects of an HSS intervention on relevant HS functions, outcomes, and impact, and/or affecting HS stakeholders' relationship, roles, rules, and resources within the system.

Evidence of the effectiveness of an intervention is needed to assess causation. Several criteria can determine causation:

- Strength of association between cause and effects (effect size)
- Temporality (cause always comes before the effects)
- Consistency (different locations must show the same effects)
- Theoretical plausibility (a mechanism for the cause and effect to be conceptually plausible)
- Specificity in the causes (the more specific or direct the relationship between intervention and outcome, the clearer the causal relationship)
- Dose response relationship (the more exposure you have to the cause, the more likely you are to see the effect)

Greenland (2005) uses other criteria for casual relationships: a simple structure that you can see through to the data and the phenomenon under study:

- No obvious plausible source of major bias
- Serious efforts to detect plausible biases—efforts that have come to naught
- Insensitivity to small and moderate biases (in other words, to determine causation, alternative explanations must be excluded)

We need to answer the following basic questions to exclude alternative explanations for observed changes (evidence) and conclude that the intervention caused the observed changes:

⁴ There are multiple monitoring approaches, based on qualitative methods, that explore the causal linkages between intervention and impact and create understanding of the effects of the contextual factors. Collectively, we will call use of these approaches "systemwide monitoring" to emphasize what is monitored on a regular basis rather than a particular approach. Additionally, this simplifies consistent use of the term. The specific systemwide approaches are discussed in detail in Chapter 5.

- Did the changes in the target area occur because of socioeconomic changes that affected the population?
- Did the changes occur because of other socioeconomic and demographic interventions in the target area that affected health?
- Did the changes occur because of other health intervention(s) in the target area?
- Did the changes occur because of special characteristics of the target area?
- Did the changes occur because of special characteristics of the participants?
- Was the intervention implemented as planned?
- Were the changes observed due to measurement errors? Or were no changes found due to faulty measurement?

USAID promotes the gold standard—a prospective randomized controlled trial—for an impact evaluation to answer these questions. However, given the inherent complexity of HSS projects because of the multiplicity of relationships and levels within an HS, as well as cost, ethical, political, and other constraints, a randomized controlled impact evaluation design is rarely possible. Other methods—quasi-experimental methods, contribution analysis, and qualitative methods—should be considered. In addition, if determining strength of association is not required, then qualitative methods can be used to determine causation. This guide encourages the use of mixed methods to generate evidence.

The HSS MEL guide considers M&E a continuum for generating evidence. Monitoring identifies gaps in implementation and tracks progress toward a project’s quantitative targets. Monitoring also periodically measures changes in outcomes resulting from the project activities and explores their causal pathways to determine evidence. Monitoring findings are used for feedback, mid-course corrections, follow-up, and learning. Monitoring can also be used to document contextual changes and identify conditions to guide how the project could be adapted to a different context. Evaluation could take the form of a performance or impact evaluation to identify whether the project was implemented as planned and whether the objectives were achieved, or to assess the magnitude of change and causal linkages. This document provides guidance that USAID staff can use to plan MEL activities designed to generate evidence of change throughout the MEL continuum.

Learning to Inform Decision Making

This document provides guidance on using evidence to inform decisions regarding the design and implementation of an HSS project, including making mid-course corrections during the project’s implementation and determining whether to scale up the intervention. The learning can be shared through various USAID knowledge management networks.

This guidance describes the elements of an “ideal” MEL situation: collecting performance, context, and systems-monitoring data; using outcome-monitoring approaches and a mix of quantitative and qualitative methods; and having enough money to carry out the MEL plan as intended for evidence generation and learning. The decision to conduct an impact evaluation is made in the design phase and resources are allocated accordingly. We recommend that “ideal” MEL procedures be followed. However, if that is not possible, then the project should document the rationale for deviating from the ideal MEL plan and the implications for generating evidence and learning.

Resources Used to Develop the Guide

All steps in this guidance are aligned with the following United States Government (USG) and USAID policy and guidance documents:

- *USAID's Vision for Health Systems Strengthening* (USAID, 2015a)
- *Evaluation: Learning from Experience. USAID Evaluation Policy* (USAID, 2011a)
- *Acting on the Call: Ending Preventable Child and Maternal Deaths* (USAID, 2014a)
- PEPFAR Blueprint: Creating an AIDS-free Generation (United States President's Emergency Plan for AIDS Relief [PEPFAR], 2012)
- ADS Chapter 200: Development Policy (USAID, 2015b)
- ADS Chapter 201: Program Cycle Operational Policy (USAID, 2017b)
- ADS Chapter 303: Grants and Cooperative Agreements to Non-Governmental Organizations (USAID, 2017a)
- 2014 Local Systems: A Framework for Supporting Sustained Development (USAID, 2014b)

Both ADS 201 and the local systems framework (LSF) underscore that an intervention is not an isolated activity but that it is implemented in a system of relationships among stakeholders. Systems strengthening cannot be discussed without dealing with the complexity arising from the multiplicity of relationships at different levels within and outside the system. Therefore, learning is given due attention to emphasize documentation on what has changed and will be useful for moving forward.

Organization of the Guide

The guide and chapters are organized around the USAID project cycle—plan (PAD/Acquisition and Assistance [A&A] document), implement, monitor, and evaluate. The second chapter provides an overview of HSS history and concepts and could be skipped if you are not interested in a conceptual discussion. Each chapter, although related to previous chapters, stands alone and could be read without losing any linkage or value, especially the last chapter on evaluation. Each chapter starts with a scenario to encourage readers to think about the steps to take before starting a description of the steps generally needed on that topic, based on state of the art and USAID ADS.

CHAPTER 2. BACKGROUND OF HEALTH SYSTEMS STRENGTHENING

Introduction

Emphasis on HSS has increased in recent years, but clarity on what HSS means remains elusive. This chapter reviews the definitions of the HS, HS performance, HSS, and systems thinking, to create a common understanding of what HSS entails and the implications of HSS MEL.

What Is a Health System?

To monitor and evaluate an HS, we must define its boundaries, content, and context to measure change and exclude contextual factors responsible for the change. An HS consists of all the people, institutions, resources, and activities—both public and private actors—whose primary purpose is to promote, restore, and/or maintain health (USAID, 2015a). USAID’s understanding of the primary purpose of the HS excludes activities where health improvement is incidental to other purposes. An HS encompasses actors at the national, state, district, and community levels (World Health Organization [WHO], 2006). It comprises stakeholders from the private and government sectors, community outreach workers, educators, researchers, patients, and health workers.

These definitions of HS give HS planners and managers discretion to describe the characteristics of a particular HS, such as its major stakeholders, functions, relationships, interdependencies, and causes and effects that result in better health status in a particular context. The HS is usually a combination of public- and private-sector institutions, with the government having the mandate to lead and regulate HS performance.

The HS is embedded in a larger socioeconomic and demographic context and is affected by it. For example, a country’s economy determines how much of the gross domestic product can be spent on health. This has two major implications for HS development and performance. First, although HS decision makers do not have the ability to influence many of the socioeconomic determinants of health (SDH), they are still responsible for addressing the health inequalities owing to SDH. For example, often HS policy aims to reduce the financial, cultural, and geographic barriers to health services. Second, not only do HS interventions affect health; other interventions in the SDH context do, too.

An HS needs health professionals, commodities, medicines, organization of health services, and a governance structure that is dependent upon a health information system (HIS) to plan, manage, and monitor these activities as well as to create financing to sustain HS functions. These activities—service delivery, human resources for health, health information, medical products, vaccines and technologies, financing, and leadership and governance—are core functions of the HS (Figure 3). These core functions (also called building blocks) have been proposed as an organizing structure for advocacy of HSS (WHO, 2007). However, descriptions of the core functions alone do not constitute a description of an HS; for that, one must also describe “the multiple relationships and interactions among the building blocks—how one affects and influences the others, and is in turn affected by them—that convert these building blocks into a system” (de Savigny & Adam, Eds., 2009: 31). In addition, the core HS functions are multidimensional and at times a core function is considered a subsystem of the broader HS, with all other core functions subsumed under it.⁵

⁵ For example, the Health Metrics Network (HMN) has described multiple information systems under HIS (HMN, 2008). The Performance of Routine Information System Management (PRISM) framework is based on technical, behavioral, and organizational components (Aqil, Lippeveld, & Hozumi, 2009). Systems for Improved Access to Pharmaceuticals and

Multiple HS theories and frameworks (Van Olmen, et al., 2012; Gruskin, et al., 2012; Mikkelsen-Lopez, Wyss, & Des, 2011; Shakarishvili, et al., 2011, 2010; Atun, de Jongh, Secci, Ohiri, & Adeyi, 2009; de Savingy, et al., 2009; Kruk, & Freedman, 2008; WHO, 2007; Mossialos, et al., 2007; Roberts, et al., 2004; Murray & Frenk, 2000; Roemer, 1993; Kleczkowski, Roemer, & Van Der Werff, 1984; Evans, 1981) help us understand the HS components, their relationships, and their linkages to HS performance. However, each HS framework has its own methods for M&E. For the sake of brevity, we will not discuss them in detail here but instead allude to them in the guide, where needed.

What Is Health System Performance?

There is no one definition of HS performance. Murray and Frenk (2000) describe HS performance as centering on three fundamental goals: improving health, enhancing responsiveness to the expectations of the population, and assuring the fairness of financial contributions. WHO (2007) concentrated on the intermediate results as performance criteria: access, coverage, equity, efficiency, quality, safety, and sustainability. In *USAID's Vision for Health Systems Strengthening*, performance is conceived both as HS outcomes (financial protection, essential services, equitable population coverage, and responsiveness to people's expectations) and impact—reduction in maternal, child, and HIV-related mortality. For USAID, good HS performance is important, because it helps to ensure that people have financial protection and access to essential high-quality services for the prevention of disease, promotion of health, treatment, and care. Good-performing HS also reach underserved, marginalized, and high-priority groups, ensuring clients' dignity, choice, safety, and protection from stigma.

It is assumed that by addressing weakness in one or more core functions and managing their interactions better, HS performance will be enhanced.

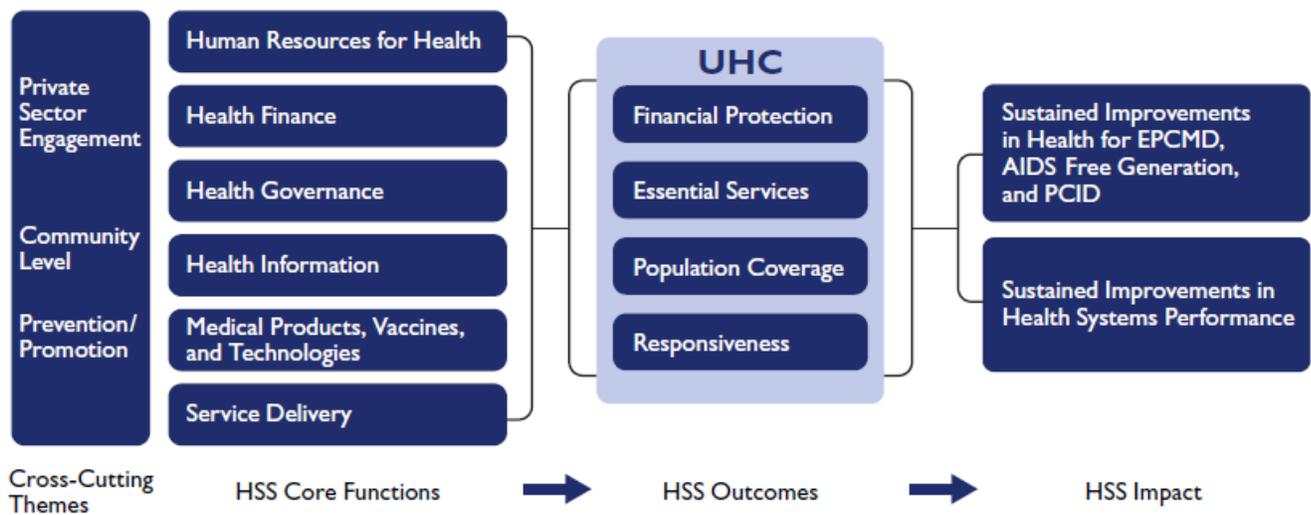
The *Vision* document states that improvements in the core functions of HSS and their interactions will lead to better HSS outcomes. Universal health coverage (UHC) means that all people can receive needed health services of sufficient quality to be effective without enduring financial hardship. UHC is one of the focal areas of the Sustainable Development Goals (SDGs). In *Vision*, the broader concepts of UHC subsume the HS outcomes of financial protection, essential services, population coverage, and responsiveness, which are defined as follows:

- Financial protection means that the out-of-pocket costs to households of accessing high-quality essential health services must neither keep people from using these services nor impoverish them.
- An essential package of high-quality life-saving prevention, promotion, treatment, and care services must be available to and used by all who need it.
- Population coverage means that people who are poor, underserved, marginalized, and vulnerable must have the same access to essential health services as anyone else.
- Responsiveness deals with people's nonmedical experiences of an HS and assures the dignity, confidentiality, autonomy, quality, and timeliness of services.

The expected impact of progress toward these HSS outcomes is sustained improvement in HS performance and sustained improvements in health to achieve the goals of preventing child and maternal deaths and controlling the HIV/AIDS epidemic.

Services Program, (Hafner, Walkowiak, Lee, & Aboagye-Nyame, 2017) is another example of a core function treated as a subsystem.

Figure 3. Framework of USAID’s vision for HSS



Quality, efficiency, and sustainability are embedded in health system performance. The effects of emerging diseases such as Ebola on HS are drawing increased attention to a new dimension of health system resilience. Resilience refers to the capacity of an HS to absorb disturbances (such as epidemics and natural disasters), responding, recovering, and adapting in order to continually provide needed health services to the population.

What Is Health Systems Strengthening?

HSS refers to the strategies, responses, and activities designed to sustainably improve country HS performance (USAID, 2015a). Verifiable improvements in HS performance will be evidence that an HS is growing stronger. Strengthening does not occur uniformly within the core functions or across regions and districts within a given country. HSS activities deploy resources specifically to improve one or more of the HS core functions in a sustainable fashion. USAID’s operational definition of HSS draws the boundaries based on the intent of USAID’s efforts and resulting patterns of resource allocation (USAID, 2015a). For its part, WHO (2007: 4) defines HSS as “...improving [the] six HS building blocks and managing their interactions in ways that achieve more equitable and sustained improvements across health services and health outcomes.”

Both definitions point to sustainable improvement in HS performance; WHO’s explicitly targets HS core functions and their interactions for strengthening an HS. The HSS intervention in one function tends to affect HS performance in interaction with one or more HS functions. Thus, one can safely assume that HSS interventions have a systemic dimension and have implications for designing the HSS MEL activities. Systems thinking helps in designing MEL activities.

Systems Thinking for HSS

Understanding how all HSS core functions work together is important in designing projects and activities as well as in monitoring and evaluating them. In 2009, WHO published *Systems Thinking for Health Systems Strengthening*, which states that an HS does not consist of its components alone; it is the “multiple relationships and interactions” that convert the individual components into a system (de Savigny & Adam, Eds., 2009). The components of an HS interact to produce health policies and programs, laws and regulations, organizations and management systems, and financing arrangements, which, in combination, result in improved health outcomes. These activities occur at all levels within a country’s infrastructure: national,

regional, district, community, and household levels (Hotchkiss, Stillman, Hutchinson, & Connor, 2006). All HSS interventions will affect the entire system, not just one part of the system.

The premise of *Systems Thinking for Health Systems Strengthening* (de Savigny & Adam, Eds., 2009) is that we must understand a system before we can strengthen it. Many interventions fail because they did not account for the system's complexity and the unintended consequences of interventions. A logical first step is to conduct an HS assessment, in the absence of one. An HS assessment provides information on the performance of an HS and helps explain how the entire system works so that HSS interventions can be planned. USAID's Health Finance and Governance Project (HFG) has assembled a manual that can help the reader conduct an HS assessment, if funds are available (USAID, HFG, 2015a). The Office of Health Systems also offers a health system benchmarking tool that provides the HS historical data for low- and middle-income countries (Aqil, Ekanem, & Ettinger, 2016). The tool could be used for situational and comparative analyses, planning, and benchmarking selected countries' health systems. The clustering function matches countries on similar socioeconomic characteristics, thus making comparison more suitable (<https://idea.usaid.gov/global-health-tools.html>).

USAID promotes a systems approach to achieve sustainable results through locally owned development, and encourages the missions to incorporate in their work the agency's local systems framework (LSF) (USAID, 2014b). "Local systems" refers to those interconnected sets of actors—governments, civil society, the private sector, universities, individual citizens and others—that jointly produce a development outcome. The "local" in a local system refers to actors in a partner country. The LSF says that a project's design should focus on "5Rs": resources, roles, relationships, rules, and results. This framework is also good for M&E, documenting such changes from an HS intervention as those occurring in staff roles, relationships, and organizational rules—useful information for deciding whether resources should be increased or shifted. Applying the LSF to an HS, one can observe that the governance function of the system deals with rules, roles, organization of service delivery, and relationships, as well as the supervision of other HS functions. The "resources" in a local system can be traced to financial, human, and information system functions of the HS.

Systems thinking involves assessing how an HSS intervention will affect other core functions of the HS. There is a need to consider the effects of one intervention across all major subsystems of the HS. This is part of the uniqueness of HSS activities that needs to be addressed when designing MEL for HSS interventions.

The Implications of HSS for M&E

HSS MEL is not only about tracking the improvement in HS performance but also about measuring changes in the targeted HS function(s) and their effects on other related HS functions. This has the following implications for the M&E of HSS projects and activities:

- First, because of the systemic nature of HSS interventions, direct and indirect effects on HS outcomes and impact must be measured. This means identifying whether the HSS intervention in a given HS function would affect the outcomes directly or indirectly through other HS functions.
- Second, causal linkages are more difficult to establish when HS core functions are treated as stand-alone subsystems and emphasis is placed on improving them without linking them to overall HS performance. For example, the following are the outputs of individual subsystem functions and stand alone:
 - Increased production and retention of the health workforce under human resources
 - Improved data quality and use of information under the HIS
 - Domestic resource mobilization under health financing
 - Availability of medicines under pharmaceutical system strengthening

The performance of each subsystem can affect the other subsystems (e.g., information system data affect performance through governance structures to make changes in other relevant functions, or the pharmaceutical system affects service quality to improve service coverage). **Stand-alone HSS interventions that are not linked with other HS functions and HS performance are not based on systems thinking.**

- Third, these HS subsystems/functions are managed by the system’s governance structures. USAID’s LSF states that any system change affects the governance of the system, which requires clarity of relationships among organizational members and outside stakeholders, organizational rules of engagement, individual roles and responsibilities among people and other functional units, and availability of resources to carry out the system’s functions or subsystems. Without changing or adjusting subsystem governance, the overall system is unlikely to perform better, regardless of the time and resources devoted to one HSS intervention.
- Fourth, the interactions of HS intervention with other HS functions can produce unintended effects and mechanisms need to be in place to detect them.
- Fifth, understanding the context in which the HS is embedded is crucial, because the context affects the HS and is affected by it as well. Therefore, contextual monitoring is crucial for teasing out the context effects. We must consider the following:
 - When HSS interventions are focused only on one or two HS functions, the rest of the functions could be considered contextual factors affecting the targeted functions.
 - There might be other HSS interventions within the HS that affect the areas where the activity is located. It is important to try to tease out the contributions of other HSS interventions.
 - Other socioeconomic interventions could affect the HSS intervention. Therefore, this information need to be collected to control for these interventions’ effects.

Given these implications, HSS MEL differs from other the MEL of other projects, because it captures not only the effectiveness of one project, activity, or intervention, but also systemic changes occurring in other HS functions, and it generates evidence by controlling for those changes that can confound observation of the HSS intervention’s effects.

Given these implications, the following recommendations for HSS MEL are in order:

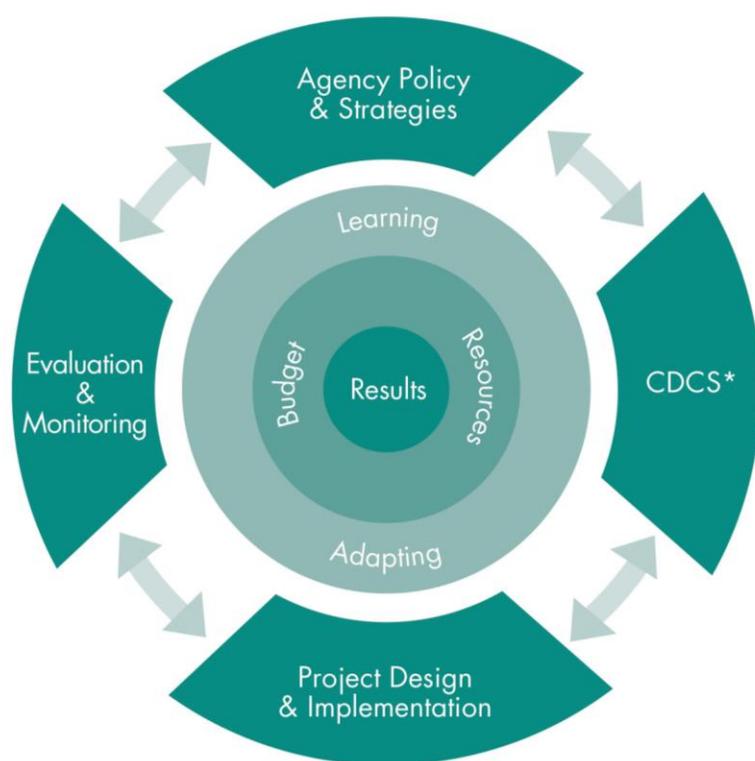
1. Include context and system-wide monitoring in MEL plans that are already monitoring performance.
2. Monitor outcomes to explore causal linkages.
3. Monitor for unintended consequences.
4. Promote the use of data by stakeholders.

These recommendations are elaborated in the chapters that follow.

Linking HSS MEL with USAID’s Program Cycle

USAID has a well-established program cycle (Figure 4). Using the agency’s policy and strategies, each country mission develops a country development cooperation strategy (CDCS). This strategy is results-oriented and focuses investments on key areas that shape a country’s overall stability and prosperity in partnership with the host country and its communities. “The CDCS is a five-year strategy that focuses on USAID-implemented resources, including non-emergency humanitarian and transition assistance” (USAID 2015b:21). Projects in different fields are developed based on CDCS. The program cycle shows MEL as the last step to gauge progress toward the CDCS goal, objectives, and results and evaluate the effectiveness of the interventions in creating change in the target population. MEL results are used for learning, adapting, and developing USAID’s policies. Thus, the program cycle continues.

Figure 4. USAID program cycle

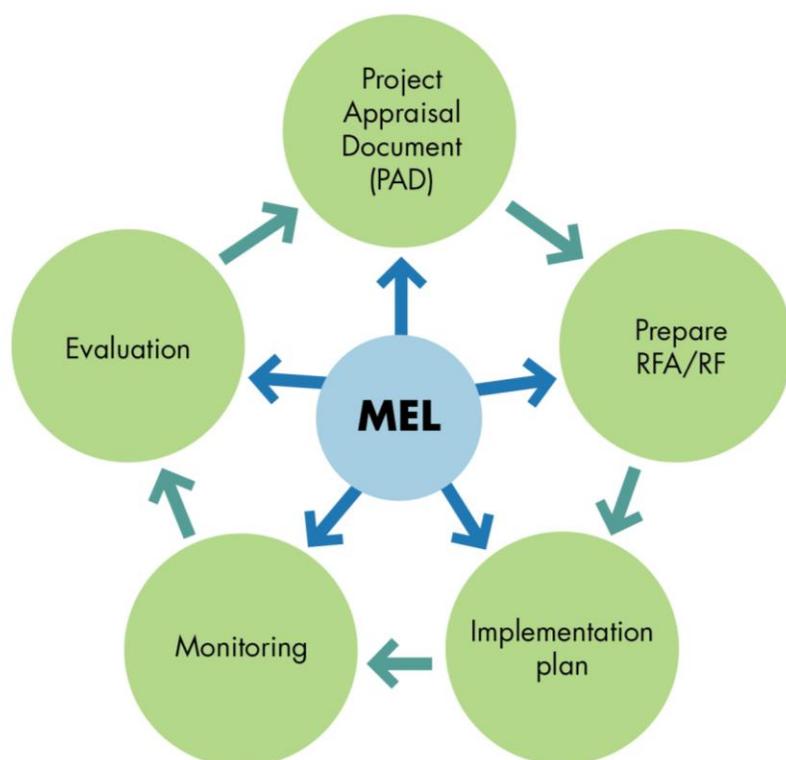


*Country development cooperation strategies

Although MEL is depicted as the last step, planning for MEL must begin when designing a CDCS and an HSS project, because decisions need to be made on how the learning and adapting will occur and be documented at each step of the cycle. Therefore, MEL activities need to be planned for each step of the program cycle.

This guide is organized according to the USAID program cycle, which is adapted and repeated in the development of a specific HSS project or activity (Figure 5).

Figure 5. Monitoring, evaluation, and learning cycle



Key Messages

- Defining concepts of HS, HS performance, HSS, and systems thinking are important for creating boundaries for HSS MEL.
- HSS is inherently systemic. Therefore, HSS MEL has the following requirements:
 - HSS MEL needs to identify and measure direct and indirect causal linkages of HSS intervention with performance and impact.
 - HSS MEL needs to detect the systemic changes in other HS functions as well as changes in organization and in people’s relationships, roles, rules, and resources.
 - To understand systemwide effects, HSS MEL must account for interaction effects among HS interventions in one function and other HS functions, including unintended outcomes of interactions.
 - Contextual monitoring needs to be incorporated to exclude alternative explanations for change.
 - Outcome monitoring must be used to explore causal pathways and linkages.
 - HSS MEL needs to be flexible to adjust to a project’s emerging needs.
- HSS MEL is based on USAID’s program cycle, which emphasizes what is needed for HSS MEL under each phase (plan, implement, monitor, evaluate, and learn across these phases).

CHAPTER 3. MONITORING, EVALUATION, AND LEARNING SECTION OF USAID’S PROJECT APPRAISAL DOCUMENT

Introduction

This chapter will focus on key considerations when developing the PAD for an HSS project focusing on MEL. We will discuss the minimum essential activities that are needed to prepare it:

- Articulate and review the theory of change (TOC).
- Prepare and review the RF diagram or narrative.
- Select and review illustrative outcome and performance indicators.
- Select and review the project’s context and other indicators as part of a systems-thinking approach.
- Guide and review the selection of sources of data collection, including a geographic information system (GIS).
- Prepare a logical framework (LF).
- Discuss and review preliminary evaluation questions and their implications.
- Discuss how learning will be incorporated in the project.
- Identify how sustainability of targeted results and outcomes will be measured.

Scenario 1

What are the factors or steps to consider when developing the MEL sections of the PAD?

The mission is designing an HSS project. As a member of the design team, you are tasked to prepare the MEL sections of the PAD. The HSS project is meant to improve the quality and availability of maternal and child health services, by strengthening leadership and governance at the district levels and below. The project will involve several HS components, such as improving the availability and quality of the health workforce and increasing the use of information for decision making.

Articulate and Review Theory of Change (TOC)

Many believe that HSS projects are complex, difficult to implement, and unable to show concrete results. The prevailing attitude is that an HSS project focusing on one or more functions of the HS may improve the targeted HS function(s) but that change may not necessarily contribute to improving HS outcomes (coverage, equity, quality, responsiveness, and financial risk protection). These prejudices can be countered by evidence of changes in HS functioning that is empirically connected to HS outcomes and, where possible, improvements in health impact.

The first step toward producing such evidence is to develop a TOC for the HSS project **to identify causal linkages and facilitate development of an RF and a MEL plan. The TOC can be revised at any stage of the project, if needed.** The TOC for an HSS project must reflect the systemwide effects discussed in Chapter 2.

The following questions should be addressed in the PAD design process:

- Does the project’s design state the problem being addressed or goal to be achieved?
- Does the project’s design describe the technical components that will come into play to reduce the magnitude of the problem, solve the problem, or achieve the project’s goal?

- Does the project’s design explain what changes will occur as result of the project’s activities and where?
- Does the project’s design describe how changes will be achieved because of the project’s strategies or intervention?
- Does the project’s design illustrate pathways between an intervention and the changes it is supposed to bring about?
- Does the project’s design provide a TOC narrative illustrating the linkages between interventions and impact, either directly or indirectly (cause and effect)?

These questions clarify what the project will achieve (goals and objectives), how the project will achieve them, and what conditions need to be met to move from lower- to higher-level results, which is helpful in developing the project’s RF. Below is an example of a TOC narrative for Scenario 1, focusing on improving the availability and quality of the health workforce and increasing evidence-informed decision making:

Example of a TOC Narrative

This project is based on the premise that **if** staff have the skills of task sharing and task shifting, **then** their ability to perform multiple tasks and share responsibilities will increase, **because** the learned skills will prepare them to meet workload needs, staff shortages, and emergencies. **If** we improve staff skills in data collection, the use of information (analysis, interpretation, and relative utility of information), **then** the staff will feel empowered to use those skills to make planning and management decisions about delivering high-quality services and associated resources, **because** informed decision making will increase the facility’s efficiency and effectiveness and improve services. This, in turn, will increase client satisfaction and facility use, which will lead to improved health status. **If** we delegate authority to make management decisions at the facility level for task sharing and task shifting and make facility staff accountable for achieving service coverage targets, **then** the staff will feel more empowered to make management decisions and feel more responsible for achieving service coverage targets. This will happen **because** leadership will have conveyed the message that they are flexible and willing to adjust governance structure when needed to keep the staff commitment and morale high for better service delivery. **If** task sharing and task shifting—a human resources for health (HRH) intervention—is not combined with delegation of authority, accountability (governance), and skills to track targets using existing data (information system), **then** it is very unlikely that staff will use the learned skills and improve clients’ use of services, **because** the interventions are linked to support one another to produce better HS outcomes. **If** instead these interventions are implemented as planned, **then** the staff, in response to increased use of services, may request more supplies (medicines and vaccines) and create demand for a budget increase at the facility or higher levels, **because** the intervention in one HS function tends to affect other HS functions, creating systemwide effects.

A TOC should not be rigid but rather should be updated as more is learned about the relationships it expresses. A process for updating the TOC should be described in the MEL plan. It is not necessary at the preparation stage of the PAD to develop a TOC for each activity within the project. However, we encourage you to develop those TOCs in the implementation phase, to facilitate a more detailed MEL plan.

Prepare and Review a Results Framework

An RF is a tool for summarizing the key elements of a project's technical areas and the TOC underlying the project's design.⁶ It is a reference for preparing the MEL section of the PAD. Therefore, as either a member of the team that is developing the HSS PAD or as a reviewer of the PAD's MEL section, you should make sure that the project's RF has addressed the following:

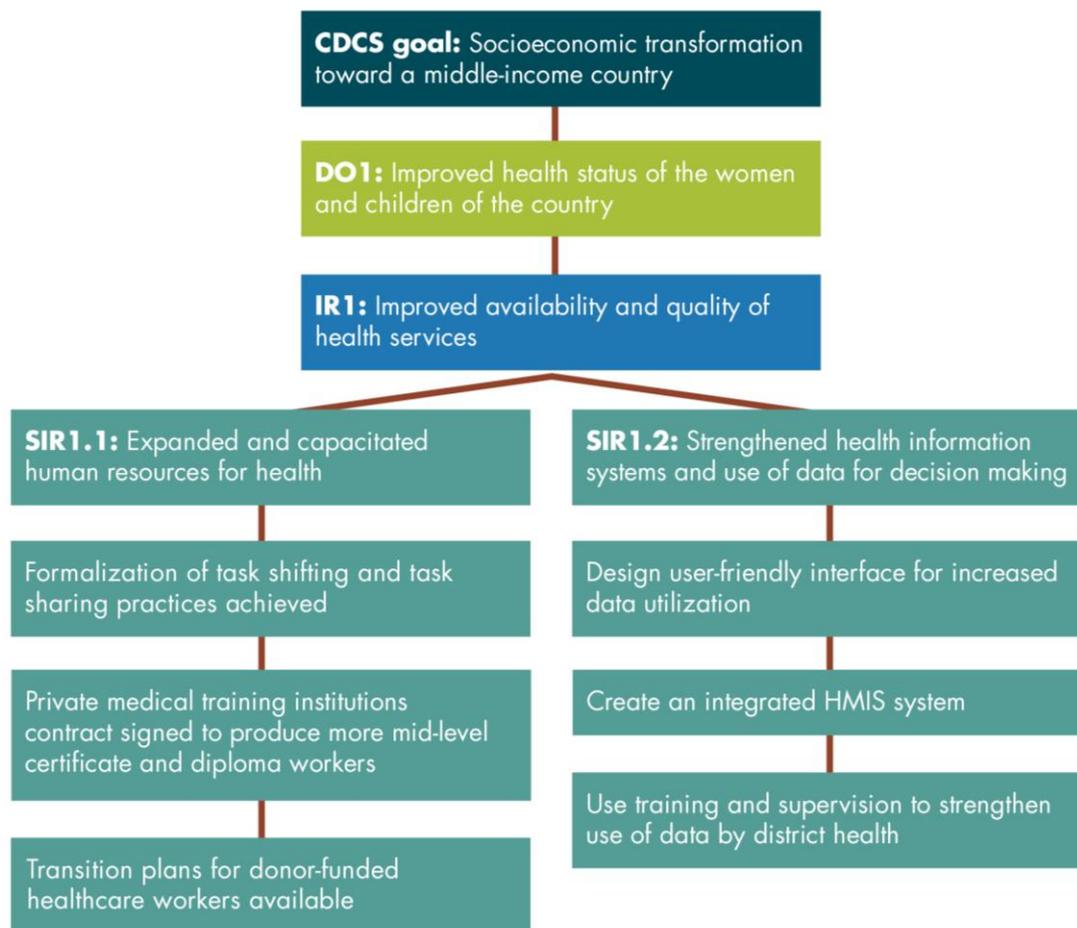
- Does the RF align itself with the CDCS RF?
- Does the RF cover all major technical areas with descriptions of intermediate results?
- Are the results reflective of major project activities under these technical areas?
- Does the RF reflect the project's TOC?
- Are the selected results appropriate and achievable?
- Are the results under each technical area described in a hierarchy—in other words, whether achieving the lower-level results would lead to achievement of higher-levels results?
- Does the RF show linkages between technical areas—in other words, pathways for achieving results and interactions between the activities described, based on the TOC?

Figure 6 provides an example of an RF for the HSS project presented in Scenario 1. In this example, the CDCS goal is advancing the socioeconomic transformation from low-income to middle-income country status. One development objective is the improved health status of women and children in this country. The intermediate result (IR) is improving the availability and quality of health services. Sub-IRs could be expanded and capacitated human resources for health, a strengthened HIS, and use of data for decision making. An increase in the number of trained health workers will improve both the availability and quality of health services. A strengthened HIS could help identify geographic areas with the greatest health disparities and guide policy and managerial decisions to improve health outcomes.

Let us look more closely at the RF depicted in Figure 6 and review whether it meets the cited criteria of a good RF. What weaknesses can you identify?

⁶ A project's results framework is sometime called a logical framework to distinguish it from the CDCS results framework.

Figure 6. Sample health system strengthening results framework

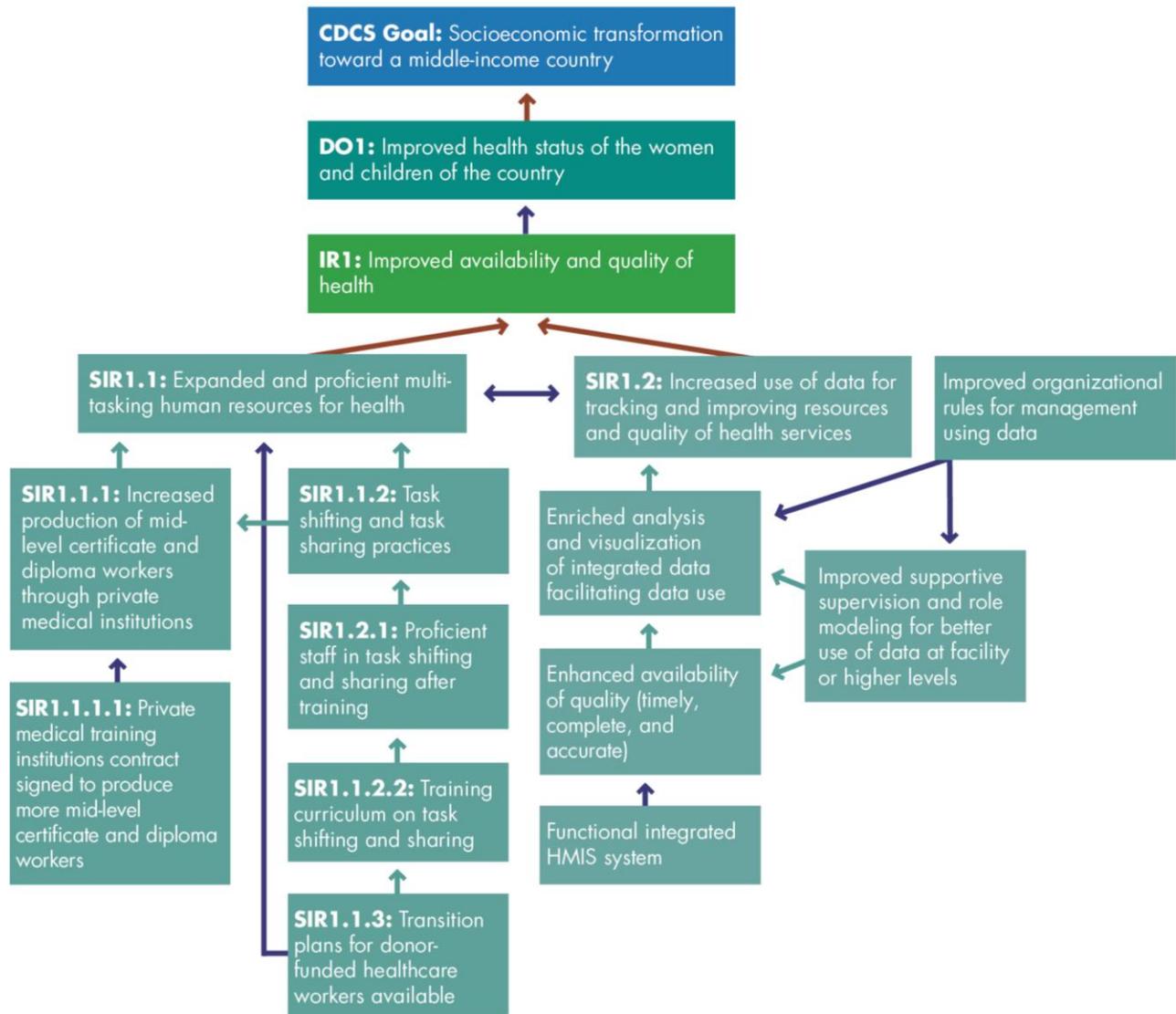


Key: CDCS (country development cooperation strategy); DO (development objective); IR (intermediate result); SIR (subintermediate result); HMIS (health management information system)

In this RF, there is some alignment between project results, the CDCS goal, and the development hypotheses. However, there is room to improve the description of results as well as linkages within and across sub-IRs, to make the TOC more explicit. What results should we achieve under the technical area of strengthened HIS that would lead to improved availability and quality of health services as well as contribute to better human resource management? What subresults will be achieved under the human resource component that would affect IR1? Are the results achieved dependent on one another, or could they be achieved in the absence of the others? Figure 7 shows an improved RF.

We understand that further improvements in this RF are possible based on details of the interventions. However, note that the description of results has improved, we have identified that results are achieved through various activities, and we have shown some results as dependent on others. In addition, we have shown that improved use of information would contribute not only to better-quality services but also better management of human resources, and a multitasking, proficient staff would also improve the use of data to improve services. This RF also assumes that task shifting and sharing as well as increasing the healthcare workforce is part of the donor transition plan. Thus, pathways for achieving results and their interactions are highlighted, making the project’s TOC more explicit.

Figure 7. Improved health system strengthening results framework



Key: CDCS (country development cooperation strategy); DO (development objective); IR (intermediate result); SIR (subintermediate result); HMIS (health management information system)

Select and Review Illustrative Performance Indicators

To monitor a project's progress toward the identified results, a limited number of illustrative performance indicators are selected under each technical area. The selected indicators are usually global standardized indicators but may include USAID's required Performance Plan and Report (PPR) indicators, to facilitate global comparison. Box 1 discusses some criteria that can be used for selecting indicators. The reason for keeping these indicators to a minimum is to allow partners who bid to implement a project to align performance indicators with their innovative strategies and to add more indicators relevant to their strategies.

Box 1. Criteria for selecting health system intermediate results indicators

1. Think about broader health element measurements at the population level (e.g., knowledge and practices) and the facility level (e.g., quality of care, facility use, and use of information).
2. Use the RF results to select relevant indicators from the broader HS outcomes (service coverage, equity, financial risk protection, and responsiveness).
3. Where possible, select standard outcome or impact indicators, as defined by the Department of State and USAID Standard Foreign Assistance Indicators (<http://www.state.gov/documents/organization/207793.pdf>).
4. Review internationally recognized outcome or impact indicators that may not be standard ones, such as out-of-pocket expenditure as a percentage of total health expenditure.

The last two considerations are important for creating comparability among the projects funded by USAID and other donors. And remember that that SMART (specific, measurable, achievable, realistic, and time bound) criteria are applicable to any selected indicator.

Please refer to Health Systems Strengthening: A Compendium of Indicators (Diana, Yeager, & Hotchkiss, 2017a) and to the Global Reference List of 100 Core Indicators (WHO, 2015) to help select appropriate and relevant indicators for your HSS project.

Here are some examples of indicators for Scenario 1:

- Percentage change from baseline of selected maternal and child service coverage in target areas after five-year lifespan of the project (population level)
- Percentage change in maternal and child services coverage between bottom and top quintile income groups (population level)
- Percentage change in household out-of-pocket expenditure as a proportion of total health expenditure (population level)
- Percentage change from baseline of health staff complying with quality-of-care standards of selected health services (facility level)
- Percentage change from baseline of use of information for evidence-based management decision making, as depicted in records at facility and higher levels in target areas (facility level)
- Percentage change from baseline in proficiency of health personnel in conducting multiple tasks and sharing tasks (facility level)

These indicators reflect results at the IR and sub-IRs level. However, if the design team decides to have more indicators reflecting lower-level results, then some examples could be as follows:

- Number of policy and organizational processes established to assure implementation of task shifting and sharing
- Percentage of facilities reporting on time, complete, and accurate data or percentage increase in timeliness, completeness, and accuracy at district or higher levels
- Percentage change from baseline in availability of health personnel by cadre category after five-year lifespan of the project

Therefore, when revising and reviewing the outcome or impact indicators in the PAD, ask:

- Does the selection of indicators meet the four criteria described?

Action: If not, revise or ask the design team to revise the selected indicators.

Select and Review Systemwide Effects and Indicators

By definition, HSS is about bringing systemic changes. The rationale for collecting information on the systemwide effects of HSS interventions is to identify changes in other functions as well as in people's roles, rules, relationships, and resources; discover whether changes in HS outcomes were caused by interaction with other HS functions; and recognize whether the intervention has a direct or indirect effect on HS outcomes through other HS functions. There are different ways to identify, develop, and track systemwide changes.

1. If you believe that the intervention in a selected function will affect other health system functions, then you can use the standardized indicators reflecting change in that HS function, for example, change in stockout of medicines/vaccines; density of health workforce per 10,000 population or geographical distribution; data quality and information use; quality of care; or reduction in out-of-pocket expenditures. Use the compendium of HSS indicators to select appropriate and relevant indicators.

You can work with the design team to identify potential effects on other HS functions, which are closely linked to the main HSS core function that this specific project addresses. For example, HIS is a support function of the governance/management structure. Therefore, you can identify a possible effect in the governance function (such as improved use of information in decision making to bolster accountability) based on improvement in HIS and then specify how evidence-informed decision

making and accountability (governance) will affect other HS functions (mobilization of more resources (HRH, pharmaceutical, budget), task shifting, retention (HRH), and outreach or community-based services (service delivery) to improve performance.

You can also ask, “Because of an intervention in one HS function, what changes are needed in other HS functions to improve performance?” If the answer is “no changes,” then clearly that intervention will have a direct effect on performance without interfering with other HS functions. However, if the answer is “yes, there will be changes,” then brainstorm with other design team members about what those changes entail. Will those changes also require changes in formal relationships, roles and responsibilities, rules of engagement, and resources to affect performance?

You can also involve MEL colleagues from the mission and Washington headquarters to assist in identifying systemwide changes and indicators, if needed.

2. It is to be noted that many emerging systemwide changes or unclear causal linkages could not be tracked through systemwide indicators, as discussed in the complexity section (Chapter 1). Therefore, it is better to put a condition in the A&A solicitation document that the selected IP will be required to employ at least one outcome monitoring technique (see Chapter 5 for more details) to track systemwide changes.
3. Due to space limitations in the MEL section of the PAD or to give more latitude to the prospective IPs, you can leave identification of systemwide effects to IPs. However, provide direction in the PAD that the IPs will be required to describe a few systemwide indicators as relevant to the proposed interventions and at least use outcome monitoring techniques to track systemwide changes. The design team could also ask the prospective IPs to provide criteria for prioritizing their selected systemwide effects, to get an assessment of their system thinking and priority setting. The design team could also specify the frequency of data collection on systemwide effects, such as every six months or yearly to reduce the cost of data collection.

Select and Review Context Information

The collection of contextual information is a requirement to explain their effects on the implementation and results of the project. However, to avoid duplication of collecting contextual information, the PAD design team can provide instructions in the A&A solicitation document that the project should take advantage of any other mission MEL mechanisms that are collecting contextual information relevant to the project. That way, the project could reduce costs and focus only on collecting the project’s relevant missing contextual information.

Therefore, when selecting and reviewing the systemwide and contextual information, make a requirement in the A&A solicitation document that prospective IPs will be responsible for the following:

- Identifying at least one outcome monitoring technique for collecting information on systemwide effects and contextual factors
- Providing evidence or lack of evidence of interaction between an intervention and another HS function, as well as mediating and indirect effects of an intervention through other HS functions
- Providing evidence, if any, on how contextual factors affect the intervention’s implementation and results

Action: If not, revise or ask the design team to revise the A&A solicitation document requirement for the IP.

Guide and Review the Selection of Sources of Data Collection

An important objective of MEL is to generate evidence that the project has produced the intended results and that plausibility of attribution exists, by excluding alternative explanations.⁷ Therefore, it is very important that data be collected from a variety of sources for triangulation and to control for confounding factors in teasing out contribution⁸ and attribution of project activities towards achievements. This section explains what to consider when choosing the sources of information for your project and other considerations related to data collection.

The considerations for selecting sources of information could be the following:

- Types of information to be collected, qualitative or quantitative
- Place and levels of organization: individual, household, facility or higher levels, IP, government, administrative records, geographical units, etc.
- Frequency of information collection: a one-time activity or at regular intervals
- New (primary) or existing (secondary) data?
- Common sources of information for the project and non-project areas?

The sources of information that would be appropriate for data collection are described in the PAD's MEL section using these criteria. Surveys are a common source of data, as is the routine health information system (RHIS), which generates data collected at public and private health facilities and institutions, as well as at community-level healthcare posts and clinics, at regular intervals. Project reports at various intervals or supervisory observation checklists are other frequently used sources of information. Project M&E staff uses at least one outcome monitoring technique (see Chapter 5 for more details) to collect information on systemwide changes. Geographic information systems and remote sensing are additional sources of data used increasingly to identify best and worst performing areas.

Data collection involves creating and maintaining databases, assuring data quality, and establishing standard operating procedures to ensure the privacy and confidentiality of respondents and collected data. These components need to be specified in the PAD's MEL section.

So, when selecting and reviewing the sources of information, ask:

- Are the sources of information appropriate and relevant to the selected indicators?
- Is the outcome monitoring technique described as a source of information?
- Are the processes of assuring data quality and respondents' privacy and confidentiality addressed during data collection and in databases?

Action: If not, revise or ask the design team to provide sources of information, including outcome monitoring technique, privacy, and integrity of data.

⁷ Attribution implies causation and involves drawing causal links and explanatory conclusions about the relationship between observed changes, whether anticipated or not, and specific activities.

⁸ Contribution implies that the activity being implemented may contribute to the observed changed but where sole attribution can't be established.

Preparation and Review of the Logical Framework

The LF is another way to summarize the MEL component of a project and is included in the PAD. It is related to the RF, but instead of describing results that reflect the changes that are expected, the LF specifies indicators that need to be tracked to measure those changes. In addition, it provides more details on project indicators of inputs and outputs and describes sources of information to collect those indicators. The LF assumes that provision of the inputs will lead to outputs, which in turn steer toward achievement of the project's objective. The LF has four columns: narrative summary, indicators, sources of data, and assumptions. The number of rows will depend on whether the RF has Sub-IRs. Table 1 is an example of an LF for Scenario 1.

Table 1. Example of a logical framework for an HSS project

Narrative summary	Indicators	Data sources/Mean verification	Assumptions
<p>Project goal: Improved health status of citizens</p>	<p>Under-5 mortality rate (100,000)</p>	<p>Demographic and Health Survey (DHS) or Multiple Indicator Cluster Survey (MICS)</p>	<ul style="list-style-type: none"> Funding for health sector will be maintained DHS/MICS will be conducted on a regular basis (every 5 years)
<p>Project purpose: Improved HS to strengthen service delivery</p>	<ul style="list-style-type: none"> Number of healthcare workers per cadre per 10,000 people Percentage of facilities demonstrating use of data for planning or budgeting 	<ul style="list-style-type: none"> HRH information system Health management information system(HMIS) Data use impact evaluation 	<ul style="list-style-type: none"> Increasing trained workforce remains a priority Continued commitment to strengthening health information system
<p>Systemwide changes</p>	<ul style="list-style-type: none"> Systemwide indicators (changes in health system functions) relevant to the project intervention Documenting systemwide changes as they occur and clarifying causal relationship between interventions and outcomes 	<ul style="list-style-type: none"> Outcome monitoring 	<ul style="list-style-type: none"> Project implementation as planned
<p>Sub-purpose 1 outputs: Expanded and capacitated human resources for health</p>	<ul style="list-style-type: none"> Vacancy rates of health worker positions in public facilities Number of new healthcare workers who graduated from a USG supported pre-service training institution by select cadre 	<ul style="list-style-type: none"> HRH information system Health facility assessments 	<ul style="list-style-type: none"> Employment packages will be competitive with private sector Training curriculum and trainers are available
<p>Sub-purpose 2 outputs: Strengthened health information system and use of data for decision making</p>	<ul style="list-style-type: none"> Primary healthcare facilities that submitted routine reports on time, disaggregated by public sector and private sector Number of planning councils using integrated HMIS indicators to develop annual health plans 	<ul style="list-style-type: none"> HMIS Monitoring reports Annual health plans 	<ul style="list-style-type: none"> Improvements in network connectivity continue; Availability of staff to analyze data and prepare report Availability of reporting forms

When reviewing the PAD, assure that LF is part of the PAD and that it aligns with the RF.

Prepare and Review Preliminary Evaluation Questions and Their Implications

The second last step in preparing and reviewing the MEL section of the HSS PAD is to determine what type of evaluation will be conducted and how it will comply with the criteria laid out in USAID's evaluation policy and ADS 203. Usually, an evaluation is planned for every large or pilot project; however, evaluation rigor and type are determined by the purpose of the project evaluation. The broader evaluation questions are these:

- Did the project implement activities as planned?
- Did the project achieve its objectives as planned?
- Did the project activities lead to observed changes in the target population compared to the control group?
- Is the intervention cost-effective?
- Is the project sustainable?

At this design stage, you must determine what type of evaluation would be needed for this project. Are we testing innovative interventions and assessing their impact, or only determining the project's effectiveness in achieving its objective without any causal attribution? Depending on the answer, select the type of evaluation to be carried out and describe it in the PAD.

Evaluations are expensive and time-consuming, and not all projects warrant ascertaining attribution, so careful consideration must be given when determining whether an impact evaluation is needed. This is a good time to seek advice from USAID staff who have experience and expertise in HSS evaluations. Consider the following criteria to determine if an impact evaluation is appropriate⁹:

- **Strategic relevance:** Is the project considered to be of strategic relevance for achieving development goals?
- **Policy development:** Are the results expected to be influential in affecting policy?
- **Innovative and untested intervention:** Is this an innovative and untested pilot project, untested project hypothesis, or an aspect of the project's TOC for which no evidence of its effectiveness exists?
- **Replicable:** Is there sufficient evidence that this type of project works well in different contexts?
- **Sustainability:** Are the end results produced by this project sustainable? Are there mechanisms instituted to sustain project results?

Any of these reasons could be used to justify an impact evaluation and should be laid out in the MEL section of the HSS PAD. If you think that impact evaluation is not needed, then you need to state the reasons in the PAD. And in that case, describe what other type(s) of evaluation will be needed.

⁹ Please refer to USAID guidance on impact evaluations for more information (USAID, 2013): https://www.usaid.gov/sites/default/files/documents/1870/IE_Technical_Note_2013_0903_Final.pdf

Learning Plan

The last step of the MEL section is providing instructions on learning plan. The PAD should outline how learning will be documented and implemented throughout the project. This includes a description of how monitoring data and knowledge from evaluations will be used to revise the TOC and adjust the project's implementation. The PAD should define how the learning from the project will fill gaps in technical knowledge and inform adjustments during implementation. The plan should also state how knowledge gaps will be explored (e.g., evaluations, periodic partner meetings, learning networks, pilot activities, use of monitoring data, and/or topical communities of practice) and how learning will be applied to manage adaptively.

Sustainability of Results and Local Ownership

Given that sustainability of targeted results and outcomes is a goal of USAID programs, you must identify means of assessing the extent to which local priorities, capacities, and resources are being brought to bear in ways that put local actors in the lead of solving development challenges. Local ownership could be measured by discrete indicators, such as involvement of local leadership and civil society in planning and management and the percentage of national budget dedicated to the health sector. Local ownership may also be included in the complementary monitoring approaches that look at how actors and relationships, rules, policies, and norms are changing within the local system in ways that indicate sustainability.

Key Messages

- Prepare or review the HSS PAD by studying the TOC, which provides the rationale for the MEL activities.
- Ensure that the HS intervention's interactions with other HS functions are depicted in the TOC and in associated results and logical frameworks.
- Check that the HSS MEL section discusses collection of systemwide changes and contextual information in addition to the required outcome/impact and performance indicators. Sources of information for each also need to be included.
- Verify that the prospective contractor will be required to identify at least one outcome monitoring technique as part of the MEL plan to track project systemwide changes.
- Describe whether an impact evaluation is needed to assess an innovative intervention, or untested ideas, or tested ideas in a new context.
- Explain learning plan, sustainability, and local ownership in MEL section.

CHAPTER 4. MONITORING, EVALUATION, AND LEARNING REQUIREMENTS IN ACQUISITION AND ASSISTANCE SOLICITATION DOCUMENTS

Introduction

Once the PAD has been approved, the project team then shifts focus to developing the A&A solicitation document for the activity. The PAD is the foundation of that document. Here we provide guidance on how to convert the PAD MEL section into the MEL section of the A&A solicitation document, to assure that prospective contractors are aware of the expectations and plan and budget MEL activities accordingly.

Scenario 2

What are the factors or steps to consider when preparing the MEL section of the A&A document? What criteria will you include to evaluate this section?

A mission has received approval of its PAD for a health-sector development project—a large project for this mission. There are several IRs, one of which involves helping the country implement healthcare financing reforms. These are aimed at achieving the government's goals of increasing the availability of financial resources for health, improving efficiency in the allocation of health resources and the use of health services, and improving both coverage and quality of care. The project, in collaboration with the government, would like to roll out the scheme to three provinces in Years 1 and 2 and later expand to six more. Given the project's size, an impact evaluation will be conducted.

MEL in the A&A Documents

The Importance of MEL in the A&A Solicitation Document

Implementing partners must be aware of the significance that USAID places on HSS MEL. Therefore, expectations should be laid out clearly in the A&A solicitation document that HSS MEL is about filling the evidence gap between HSS intervention(s) and HS outcomes and impact, by exploring causal pathways, linkages, and systemic changes using innovative M&E approaches, methods, and techniques. This expectation will ignite applicants' ingenuity to rise to the challenge and prepare the HSS MEL section well.

MEL Issues to Address in the A&A Solicitation Document

MEL Requirements

The A&A solicitation document language depends on the type of agreement. Therefore, the reader is advised to write requirements accordingly in consultation with the Office of Acquisition and Assistance. The A&A document should at least describe the components of the MEL plan that are required as part of the application. It typically states that the applicants should use both quantitative and qualitative methods to collect information. The applicants may be asked to outline their methods to establish the baseline; verify results; explore causal pathways or linkages, unintended results, intermediate outcomes, and systemwide changes; and investigate alternative explanations for achieving program results.

The A&A solicitation document, based on PAD, describes the MEL requirements. Table 2 shows some sample language used to describe the requirements for MEL in two A&A documents reviewed to prepare this guidance before the publication of the new ADS 201 (2017b). The second example provides more details than the first of what is required, including a list of illustrative indicators to be disaggregated by gender when

appropriate. The request for proposals (RFP) 2 also provides guidance on indicator selection and how the applicant must discuss the integrity, precision, reliability, and timeliness of the selected indicators. However, these examples are missing important HSS MEL considerations.

Table 2. MEL language in selected RFPs

RFP 1	RFP 2
<p>Monitoring and evaluation plan:</p> <p>The application must include an activity monitoring and evaluation plan focused on the following aspects:</p> <ul style="list-style-type: none"> • The technical application should clearly outline the broad elements/components of the monitoring and evaluation plan. • The technical application should clearly identify output and outcome indicators, end-of-program targets, and year-to-year milestones with appropriate linkages to the major activities and tasks. • The technical application should clearly describe the anticipated sources of data, methods to be used to collect data, the flow of data, and the method to be used for data analysis. • The technical application should clearly explain how information generated through this monitoring plan will be communicated to other stakeholders. • The implementing partner will provide all assistance and support to the external evaluator for the performance evaluation, as deem fit. 	<p>Monitoring and evaluation strategy:</p> <ol style="list-style-type: none"> a. A discussion of monitoring and evaluation approach clearly tied to the proposed strategy, including proposed indicators, milestones, and timetables to achieve targets tied to the results under each component of the SOW. b. A description of the monitoring strategy for tracking quality improvement (“gap closing”) in the selected set of health facilities described in Components 1 and 4. c. The Offeror shall present an illustrative performance, monitoring and evaluation plan (M&E Plan) as an attachment to the proposal, disaggregating indicator data by gender, where appropriate. Once the contract has been awarded, a finalized M&E Plan will be requested from the chosen Offeror on the basis of the accepted strategy. Illustrative indicators are provided in Attachment J.3 Annex 15 of this RFP. The proposed M&E Plan may contain additional indicators and targets proposed by the Offeror as appropriate to the proposed technical approach. d. Offerors shall also propose a system to collect, analyze, track, and report on indicator/performance data, as well as a strategy for disseminating timely information to the MOH and to USAID. e. Offerors will work with the third party external evaluator and provide all support and assistance related to impact evaluation

What is missing in these two examples in Table 2 are indicators or monitoring methods to track contextual and systemwide changes, and requirements to document learning through synthesis of collected data.

Therefore, when the A&A document is prepared or reviewed, ensure that the following are part of the HSS MEL section:

- **The collection of contextual information:** A minimum requirement for tracking contextual factors is requesting the IP to create a mechanism to document all development activities in the target areas, including manmade or natural disasters, that could influence the activity’s implementation and results.
- **The collection of systemwide indicators/changes.** In addition to performance and impact indicators, a few illustrative systemwide indicators could be described, or ask the applicant to provide a list of systemwide indicators reflecting changes in other health system functions that would be appropriate to track during the project’s life (see more in Chapter 3).
- **The outcome monitoring approaches.** The A&A document should require the applicant to propose at least one outcome monitoring approach (see more in Chapter 5) to identify changes that

could not be identified a priori and clarify unclear cause and effects relationship between interventions and impact.

- **Record learning.** The IP is asked to synthesize the performance, contextual, and systemwide indicators that will document learning in the required reports.

Action: If not, revise or ask the relevant team to revise the instructions for the prospective IP.

MEL Questions

The A&A solicitation document may ask applicants to lay out a learning agenda with illustrative questions that will be explored during activity implementation and address gaps in technical and implementation knowledge. The MEL learning agenda and questions will depend on the specific activity. However, some of the generic and illustrative questions could be as follows:

- Did the activity contribute to unique lessons in advancing the state of the art technical knowledge?
- Did the activity devise a unique methodology to capture interactions among the health system functions and their relationship to health system outcomes/impact?
- Is the activity theory of change valid for achieving its intended results?
- Did the activity identify causal pathways between intervention(s) and health system outcome/impact?
- What are the unique management lessons learned during implementation of the activity?

Using Scenario 2 as an example of an activity, some of the specific MEL questions could be the following:

- To what extent did the adoption of financial management guidelines affect other HS functions, and how? Were any unintended consequences noted?
- Has the strategy increased the amount of revenue available to healthcare facilities?
- Has the readiness to provide healthcare services improved along with quality of care?
- Has service coverage improved?

Impact Evaluation

Based on the PAD, the A&A solicitation document should clearly state whether a third party or external evaluator will conduct the impact evaluation. It would be helpful for the prospective applicants to know the rationale of the impact evaluation and its design. Applicants may be advised that they are expected to work with the external evaluator in the design and implementation of the program to ensure that it meets the evaluation needs. This is particularly important for an impact evaluation, where it is critical to set up proper treatment and control groups that are needed to determine a counterfactual. A SOW for an external evaluation can be attached if one has been developed.

Specific Issues

MEL should be represented in several sections of the A&A solicitation document and it should be cross-referenced through all sections that may contain MEL language. The specific sections and MEL issues that need to be addressed are listed below:

- **MEL budget.** One option is to request a lump sum to be set aside for both MEL and collaborating, learning, and adapting. Another is to require that 5 percent to 10 percent of the budget be allocated for MEL. Some MEL costs that can be included in the budget are MEL staff, data collection training, data collection, data analysis, data quality assurance, report writing, and dissemination.
- **MEL staffing in an A&A solicitation document.** At least one MEL position should be considered essential and required in the IP's staff. If the IP does not list a dedicated staff person, then they should indicate if they plan to work with an independent firm to provide MEL expertise.

- **MEL deliverables.** This can be a draft MEL plan as part of the proposal, as described above. The full MEL plan for the activity is due within 90 days of receiving an award and this should be stated in the request for applications (RFA) or RFP.

Reviewing MEL Sections of Proposals

HSS MEL requirements described in the A&A solicitation document serve as criteria for assessing whether the A&A requirements were addressed appropriately in the MEL section of the proposal.

Key Messages

A&A solicitations documents may include the following MEL requirements:

- Collect performance indicators, systemwide indicators, and context information.
- Conduct outcome monitoring to explore intermediate outcomes, systemwide changes, and unintended consequences.
- Explore causal pathways between the HSS intervention and its outcomes or impact.
- Assess the intervention's interaction with relevant HS functions.
- Explain learning plan, sustainability and local ownership.
- Describe roles and relationship of project IP with IP conducting impact evaluation, if there is one.

CHAPTER 5. REVIEW OF THE MONITORING, EVALUATION, AND LEARNING IMPLEMENTATION PLAN

Introduction

This chapter deals with reviewing an IP's submitted project MEL plan to ensure that it fulfills the A&A agreement requirements and that all components needed are in place to create evidence that an HSS activity is contributing to HSS. The chapter explains the steps that should be carried out in reviewing the components of a MEL plan: the TOC and RF, the HSS monitoring plan, the data collection and management system, establishing a baseline, promotion of the use of collected data, and evaluation. The HSS monitoring section is further divided into the indicators, systemwide effects, context tracking, and sources of data, including a section on outcome monitoring approaches (depicted as complementary monitoring ADS 201) to link project activities with outcomes during implementation. Throughout the chapter, emphasis is placed on what needs to be included in each section of the MEL plan, what questions to ask while reviewing the sections, and what action to take if questions are not addressed.

Scenario 3

What should be in the MEL plan to assure that evidence on HSS achievements is gathered?

A mission in a sub-Saharan African country awarded a contract to an IP for a project to improve maternal and child health in all health ministry facilities in the ABC region of the country. This region has some of the lowest rates of antenatal care (ANC) attendance and institutional deliveries, and maternal and infant mortality are higher than in other regions. One of the main components of the activity is the creation of quality improvement teams, consisting of health workers at each facility. The goals of the improvement teams are to review performance gaps in maternal and child health service delivery, identify the gaps' causes, and develop appropriate interventions using the given resources. One of the required indicators for reporting in the activity agreement was, "Number of ANC visits by skilled providers from USG-assisted facilities."

Reviewing the MEL Plan's Objectives

The first step in reviewing the MEL plan is to assure that the objectives are aligned with the M&E requirements specified in the IP's contract or cooperative agreement. The A&A document provides only a broad framework of activity tasks and illustrative indicators. Therefore, the IP is expected to go beyond what is laid out in the A&A solicitation document based on the activity being implemented.

One objective of the MEL plan is to describe the IP's strategy for generating evidence on the causal linkages between the HSS activities and HS outcomes and on the interactions among the HSS activity and the other HS functions, exploring unintended results and systemwide effects, and investigating alternative explanations for achieving program results. Another objective of the MEL plan should be to identify the main questions addressed through evaluation. Last, the MEL plan should identify a learning component explaining how the activity's implementation and special studies will fill the gaps in knowledge about the TOC of the activity or HSS in general. Illustrative learning topics should be identified.

Therefore, the main questions to ask when reviewing the objectives are as follows:

- Does the MEL plan have clear and well-stated objectives, which align with the M&E requirements of the contract or cooperative agreement?
- Does the MEL plan have objectives related to exploring causal linkages and systemwide effects?
- Does the MEL plan describe objectives related to the type of evaluation to be conducted?
- Does the MEL plan explain the learning plan?

Action: If your answer to any of the questions is “no, without any explanation,” then you must go back to the IP and ask them to add and or modify these sections.

Reviewing the TOC

The second step in reviewing the MEL plan is to assess the adequacy of the TOC and the RF. As mentioned previously in Chapter 2, the TOC is based on the development hypothesis that describes how the development results are expected to be achieved and any critical assumptions that are made in the developmental hypotheses. The TOC is critically important, because it is the conceptual basis of the activity’s RF, design and implementation, learning, adapting, and M&E. The sub-IRs in the activity’s RF reflect condition(s) that need to be met to achieve the higher-level results. Therefore, there is a connection between the subtasks for achieving lower- to higher-level IRs, but there are also linkages between the sub-IRs to achieve the IRs and higher-level goals. These linkages can provide some guidance for exploring how intervention activities interact.

Whereas the A&A agreement typically incorporates USAID’s current thinking on the TOC or development hypothesis, the MEL plan is an opportunity for the IP to revisit the activity agreement and improve upon the TOC, based on the specifics of the activity. This TOC, including causal linkages, should be based on existing theories and frameworks for HSS and existing evidence. Moreover, the TOC should be tailored to the country context and should consider how the proposed investments by USAID and others, including the health ministry and other international health partners, collectively lead to the overall goals and objectives of HSS. See Box 2 for more information about TOC for individual subactivities.

Therefore, when reviewing the TOC in the MEL plan, ask:

- Does the MEL plan have a TOC that describes steps that lead to HSS results?
- Does the TOC specify systemwide changes, such as effects on HS functions and their interactions?
- Is the TOC justified, based on existing HSS theories, frameworks, and evidence?
- Does the TOC have contextual variables that could affect expected results?
- Does the TOC present the critical conditions when evidence is there and make assumptions when there is no evidence to link activities with results?
- Does the MEL plan call for updating and adjusting the TOC as data are collected?

Action: If your answer to any of the questions is “no, without any explanation,” go back to the IP and ask it to revise the TOC.

Box 2. TOC and subactivities

One USAID-funded activity in an East African country has subactivities, such as introducing or strengthening health insurance and other types of financing reforms, building leadership and management capacity at the district and facility levels, supporting facility accreditation in health centers and hospitals, strengthening HIS, and promoting the role of the private sector throughout the HS. It is possible to develop one TOC to capture all subactivities and their major tasks. However, many details may be missed. In that case, the IP should be encouraged to develop a separate TOC for each subactivity. This is particularly important for those where there is less certainty about the pathways to the expected results.

Reviewing the RF

The next step is to review the RF, which depicts the causal chain of results that are articulated in the TOC. The IP's contract or cooperative agreement provides the RF, but like the TOC, it should be updated to reflect the subactivities being implemented. The MEL plan is an opportunity for the IP to revisit the RF to reflect any changes that were made to the TOC.

Therefore, when reviewing the RF, ask:

- Is the RF well aligned with the TOC?
- Are there any differences between the RF in the contract or cooperative agreement and the RF in the MEL plan? Are those differences justified, based on the TOC and existing evidence and country conditions?

Action: Communicate any identified gaps in the RF to the IP so that the IP's MEL staff can address them in a revised RF before the MEL plan is approved for implementation.

Reviewing Monitoring Information as per the A&A Agreement

This section is about reviewing whether the IP has included all the required changes as per A&A agreement (Chapter 4). These changes are reflected and described as:

- Performance indicators
- Systemwide changes and unintended consequences
- Contextual tracking

Performance Indicators

In the HSS monitoring plan, the IP describes inputs and outputs under specific activities and subactivities that are related to results and sub-results, besides required outcome and impact indicators. Input and output indicators should be part of the MEL plan for two reasons. First, they help in tracking progress and in moving from low- to higher-level achievement. Second, they provide useful information on whether the activities and subactivities were implemented as planned, thus excluding implementation failure¹⁰ as one explanation why an activity did not show expected results.

Table 3 defines the types of indicators and presents some examples that could be used for Scenario 3.

¹⁰ Implementation failure refers to activities that were not implemented as planned, resulting in no desired or expected changes. Such a failure need to be separated from the theory and measurement failures that arise because either the activity's design or hypothesis was faulty or the measurement tools were not good enough to pick up the desired changes.

Table 3. Indicator definitions and examples from ANC quality improvement activity

Indicator type	Definition	Example
Input	Measure resources, both human and financial, devoted to a program or activity	Number of training materials developed
Process	Look at the ways in which goods and services are provided	Number of training session conducted; Number of pregnant women referred from the community to health facility for ANC; number of pregnant women with high risks referred from health facility to hospital
Output	Refers to the access to and quality of interventions such as health services as well as the readiness of providers to deliver these interventions	Number of staff trained; Proportion of pregnant who received two doses of tetanus vaccine (TT2) during ANC visits; Proportion of women with at least 4 ANC visits; % of health providers complying with quality standards of maternal services
Outcome	Measure of the broader results achieved through the provision of goods and services at the population and system levels	Percentage of women who could describe at least two pregnancy risks; Percentage of deliveries assisted by skilled birth attendants; improvements in client satisfaction with care; HS outcomes include: financial risk protection (reduced out of pocket health expenditures), health equity, responsiveness, and efficiency
Impact	Refers to indicators of the ultimate impact of HS, improved health status and reduced mortality and morbidity, or disability-adjusted life years)	Improved maternal and infant mortality
Systemwide	Refers to changes in the HS, which could take two forms: (a) systemwide indicators are mainly operationalized as changes in HS functions; (b) changes in rules, roles, relationships, resources, and behavior among the systems' stakeholders or any other changes outside of these categories	Percentage of health provider positions filled vs. allocated at the facility/hospital levels in target areas; timeliness, accuracy and completeness of HMIS data in target areas; number of management rules (policy changes) for implementing intervention in target areas; involvement of civil society and local leadership in health decision making facility and higher levels; changes in budget for sustaining the project activities Emerging perceived changes in rules, roles, relationship, resources, and behaviors of the stakeholders or any other changes not depicted by these categories. Usually, these changes are tracked through outcome monitoring
Contextual	Refers to changes in the HS context mainly related to socioeconomic, demographic, and political determinants of health lying outside of HS, natural or manmade disasters, other development projects affecting health, or other health projects supported by other donors or government over which the current project has no control	Per capita GNI/GDP; percentage of people living in extreme poverty level; literacy level, urban/rural divide; percentage of people having access to safe drinking water and sanitation facilities; presence of development projects, including other health project in the area; any natural disaster affecting project activities

Indicator Reference Sheet

The monitoring plan should attach in the annex the indicator reference sheets (IRS) for all selected (performance or systemwide) indicators. The IRS provide indicator definitions, sources of data, and calculation instructions.

Box 3. Types of systemwide indicators and effects

1. Systemwide indicators that assess change(s) in other HS functions. Here are some examples of standard global quantitative indicators for depicting changes in HS functions and subsystems:

Medical products, vaccines, and technologies	<ul style="list-style-type: none">• % reduction in tracer medicine/vaccine stockout• % increase in the rational use of medicines
Human resources for health	<ul style="list-style-type: none">• % improvement in human resource production, deployment, and retention
Health information	<ul style="list-style-type: none">• % improvement in timeliness, completeness, and accuracy of HMIS
Health financing	<ul style="list-style-type: none">• % increase in the health budget at national and subnational levels• % increase in allocated health funds and expenditure
Health governance	<ul style="list-style-type: none">• Presence of civil society organizations participating in HS decision making at national and subnational levels
Service delivery	<ul style="list-style-type: none">• % of health providers complying with a national standard of quality for a specific health service

Systemwide effects on health system functions are reflected in these quantitative indicators. These or some other indicators (see Compendium of Indicators) could be selected based on their relevance to the project and be part of the MEL plan.

2. Systemwide changes that depict shifts in the rules of organizational management and governance brought by the intervention affecting other health functions and service coverage. Each activity will bring its own set of organizational rules, and the MEL plan should indicate that they will be tracked through outcome monitoring.

3. Systemwide effects that capture changes in system stakeholders' behaviors, roles (responsibilities), relationships, and resources, and their effects on HS performance. The MEL plan should describe that outcome monitoring will be used to track systemic changes.

4. Systemwide effects that capture unintended changes. The MEL plan should describe that outcome monitoring will be used to track unintended consequences.

Systemwide Changes/Unintended Consequences

As noted in Table 3, some systemwide changes that relate to health system functions are measured using standardized indicators and are categorized as systemwide indicators. Other systemwide changes that could not be depicted through standardized indicators and that emerge during monitoring could be grouped into three categories as systemwide changes in: (1) organizational structure/governance/management; (2) stakeholders' behaviors; and (3) unintended changes (Box 3).

To avoid data burden and costs, the MEL plan should describe that other systemwide changes should be tracked using outcome monitoring techniques discussed later in the chapter. The systemic changes should be collected quarterly, or if that is not financially feasible, every six months. These changes can also be used to explain the systemwide and performance indicators collected on a regular basis, as well as exploring causal relationships and influence of the contextual factors.

Tracking Contextual Factors

These indicators measure factors outside the control of USAID that could affect the achievement of expected results (ADS 201). The contextual data should be collected by project managers and supervisors and be kept in the project database for reporting. These data can be used to explain their influence, if any, on the activity's implementation, process, outputs, and outcomes. Some examples of context indicators that could affect an HSS activity are in Table 3. They can be included in the MEL plan, reported at the beginning of the activity, and tracked as implementation progresses. These indicators may be grouped in the following categories:

- Existence of other health activities being implemented by the government and other partners that are similar or that aim to achieve similar results
- Presence of other development projects in areas that may have indirect influence on the expected results of this activity, such as road construction, agricultural activity, water and sanitation, income generation, and micro-credits
- Existence of any national communication campaign that promote messages related to the activity
- Presence of any natural and manmade disasters or events that could potentially affect the activity's implementation

The MEL plan should also address how local ownership and activity sustainability will be measured. This could be through discrete indicators or through qualitative approaches and will depend on the activity and the kind of data available at the local level.

Therefore, when reviewing the indicators in the activity's MEL plan, ask:

- Does the MEL plan have the outcome indicators required by the activity agreement? Does the MEL plan have the indicator reference sheets, filled out properly?
- Does the MEL plan propose additional indicators (processes/outputs) to capture effects of specific HS intervention(s) on HS outcomes?
- Does the MEL plan have appropriate systemwide and contextual indicators?
- Does the MEL plan describe outcome monitoring mechanism(s) for identifying/capturing systemwide and unintended changes on a quarterly or six-monthly basis?
- Does the MEL plan describe how sustainability and local ownership of the activity will be measured?

Action: If the answers to these questions are not affirmative or satisfactory, ask the IP to revise the MEL plan, addressing missing information.

Reviewing Sources of Information: from What to How and Why

In the previous section, we learned what information is needed as part of the MEL plan. Now, we will describe how and why it will be collected. The “how” describes the sources of information that will be used to track the performance and contextual indicators. Some information may not be known at the start (e.g., systemwide changes and unintended changes). In these cases, the IP will describe a mechanism to collect information on emerging changes not captured through indicators. The “why” serves the purpose of creating evidence of change(s), by using multiple sources of information to triangulate information and validate results and by excluding alternative explanations that changes occurred only in response to the intervention(s).

Each indicator that has been specified in the MEL plan and the TOC should be associated with appropriate source(s) of information for those indicators. Thus, one needs to check whether various sources of data are aligned with types of selected indicators. Outcome and impact indicators, such as access to services, service utilization/coverage, responsiveness, and changes in health status, are collected through population surveys, civil and vital registration, and census. Health facility infrastructure, types of services, staff, and quality of care are collected through facility surveys. Table 4 summarizes examples of types of indicators and their appropriate sources of information, which can be used to review alignment between indicators and relevant sources of information in the MEL plan. These should be adapted to suit the project or activity.

Table 4. Types of indicators and their sources of information

Types of indicators*	Sources of information
Impact: Diseases, deaths, births nutritional status, disability	<ul style="list-style-type: none"> • Census, civil, and vital registration, population survey and surveillance • Routine information systems such as a district health information system (DHIS)
Outcomes: Quality of care, service coverage, equity, efficiency, financial protection, responsiveness	<ul style="list-style-type: none"> • Population-based household survey • Facility surveys; project specific surveys • DHIS, national health accounts (NHA)
Outputs Facility utilization rate, medicine/vaccine stock-out, density of human resource by cadre, # of staff trained, out of pocket expenditure, types of service provided, facility readiness	<ul style="list-style-type: none"> • Human resources information system (HRIS) • NHA • DHIS • Office/facility and project records
Processes # and types of training conducted, change in patient scheduling, service delivery organization, data quality checks, changes in governance and accountability processes	Office/facility and project records
Inputs Development of training materials, availability of supplies, equipment, staff, and finances	Office/facility and project records
Contextual indicators <ul style="list-style-type: none"> • Socioeconomic determinants of health • Epidemic/natural disaster surveillance • Other development projects, including health • Natural/manmade disasters 	<ul style="list-style-type: none"> • National and districts surveys • Special studies • Mapping of donors and their interventions • Ministries/planning/districts office records • Project documentation

<p>Systemwide indicators</p> <ul style="list-style-type: none"> • Stockouts and rational use of medicine/injection • Timeliness, completeness and accuracy of data and information use • Absenteeism, retention rate, vacant vs. allocated position • Timely payment of salaries, budget allocation • Decentralized or devolved decision making • Facility vs. community-based services ratio • Preventive vs. curative services ratio 	<ul style="list-style-type: none"> • Office records • Facility survey • HMIS/logistic management information system (LMIS)/financial information system • Key informant interviews • Focus groups of activity participants
<p>Systemwide changes</p>	<ul style="list-style-type: none"> • Outcome monitoring techniques • Key informant interviews • Focus groups of activity participants • Stakeholder interviews including with supervisors/managers
<p>Unintended changes</p> <p>These are emerging changes which were not intended by the project/activity and which could be positive or negative. These unintended changes can occur at the output, outcome, and impact levels. Sources of information can be included key informant interviews, focus groups, or document reviews.</p>	<ul style="list-style-type: none"> • Key informant interviews • Focus groups of activity participants • Stakeholder interviews including with supervisors/managers

The MEL plan specifies the source of information in alignment with indicators, but this practice could be improved upon by stating other ways the data could be used to help improve our understanding of an HSS intervention. The MEL plan should explicitly state how each source of information would be used to create evidence of expected change in selected indicators. This should go beyond collecting indicators and reporting. Box 4 presents some examples of how data can be used.

Box 4. Additional uses of data collected in the MEL plan

1. Triangulate data using different sources of information to validate the information collected on one or more indicators: for example, validating HMIS information through survey data or research data.
2. Extend existing data sources—Demographic and Health Survey (DHS), DHIS, MICS, NHAs, human resource information system (HRIS—to collect information from the activity areas to similar nonactivity areas, which require minimal additional costs. This extension of data collection from existing sources allows comparisons between the activity and nonactivity areas having similar characteristics. This comparison would allow creating evidence by excluding some of the alternative explanations for observed change. Therefore, the MEL plan should state what indicators will be compared between the activity and nonactivity areas, by specifying the sources of information.
3. Create sources of information to collect contextual data on other interventions and events that may affect activity outcomes.
4. Identify sources of information and data collection methods to collect information on unintended changes from key informants, staff, and supervisors.
5. Analyze data from different sources to control for confounding factors and create a relationship between interventions and results, excluding alternative explanations and generating evidence.

Thus, in an MEL plan for Scenario 3, we could find the following statements on the rationale for sources of information, as follows:

- ANC data will be collected through activity reporting, DHIS/HMIS, and survey data to show progress over time from baseline.
- The reported ANC coverage for activity and nonactivity areas will be compared through existing sources of information, such as DHIS/HMIS/HRIS, assuring that nonactivity areas have similar characteristics as activity areas. This comparison will give confidence that change is due only to the activity and possibly excluding a secular trend.
- Reported compliance with quality-of-care standards will be triangulated, by using information sources such as record review, client exit interviews, and direct observation.
- To exclude alternative explanations of results, the activity will collect contextual data (such as other interventions or events outside of the activity, through such sources as administrative records from government sources and activity records) that might affect the activity: for example, a national communication campaign encouraging women to seek care or a natural disaster affecting supplies or delivery of services.
- The activity will collect information on systemwide and unintended changes from stakeholder, staff, and supervisor interviews.

Therefore, when reviewing the sources of information in the activity's MEL plan, ask:

- Does the activity's MEL plan specify sources of information aligned with the selected indicators?
- Does the MEL plan identify types of comparisons between activity and nonactivity areas to be carried out for specific indicators, using existing sources of information for creation of evidence of change?
- Does the MEL plan describe sources of information to record contextual information?
- Does the MEL plan describe sources of information to record systemwide and unintended changes?

Action: If the answers to these questions are unsatisfactory, ask the IP to revise the plan to address the missing information.

Data Collection Methods

Many data collection methods could be used to collect evidence of HSS effectiveness. The type of methods you choose will depend on the types of questions you want to answer. The costs and benefits of data collection options should be considered when indicators are chosen to track changes over time during the HSS process. Data collection methods should be consistent for comparison over time. The IP should also state what methods it proposes to use to collect data on unintended changes.

Data Collection Frequency

The activity MEL plan should clearly state how frequently data for each indicator will be collected. The frequency of data collection for a specific indicator depends on whether the target or change could be achieved in a specific period or if change is affected by seasonal variations. Some information is collected often (monthly), such as the number of children vaccinated, pregnant women receiving ANC care, and cases of malaria or pneumonia reported by facilities. Other information, such as deaths in facilities, could be collected on a quarterly basis, when there are more data to report. Table 5 provides an example of selected indicators, their methods of data collection and sources of information, and frequency of reporting. This information is also part of the activity's LF.

Table 5. Example commodities indicators, source, method, and frequency of collection

Indicator	Source	Method	Frequency
Percentage of facilities with stock-out	<ul style="list-style-type: none"> • HMIS • Supervisory report 	<ul style="list-style-type: none"> • Report review • Observation 	quarterly
Stock adjustments	LMIS	Report review	monthly
Delivery coverage	LMIS	Report review	monthly
CYP	HMIS	Report review	annual

As pointed out in the indicator section above, unless specific health-function indicators are available for tracking systemwide changes, we cannot specify systemwide changes a priori. Therefore, we recommend that stakeholder interviews and supervisory and other administrative/activity records should be used to document these systemwide changes on a quarterly or six-monthly basis.

Therefore, when reviewing data collection methods and frequency, ask:

- Does the IP provide a reasonable description of the data collection methods that will be used to collect the HSS monitoring indicators?
- Does the IP provide a rationale for the frequency and method of data collection and the source of the data?
- Does the IP provide a rationale for the frequency and method of data collection on systemwide effects and unintended changes along with the source of the data?

Action: If the methods and frequency of data collection and the data sources of information are not mentioned in the MEL plan, ask the IP to revise the plan and address the missing information.

Outcome Monitoring Techniques

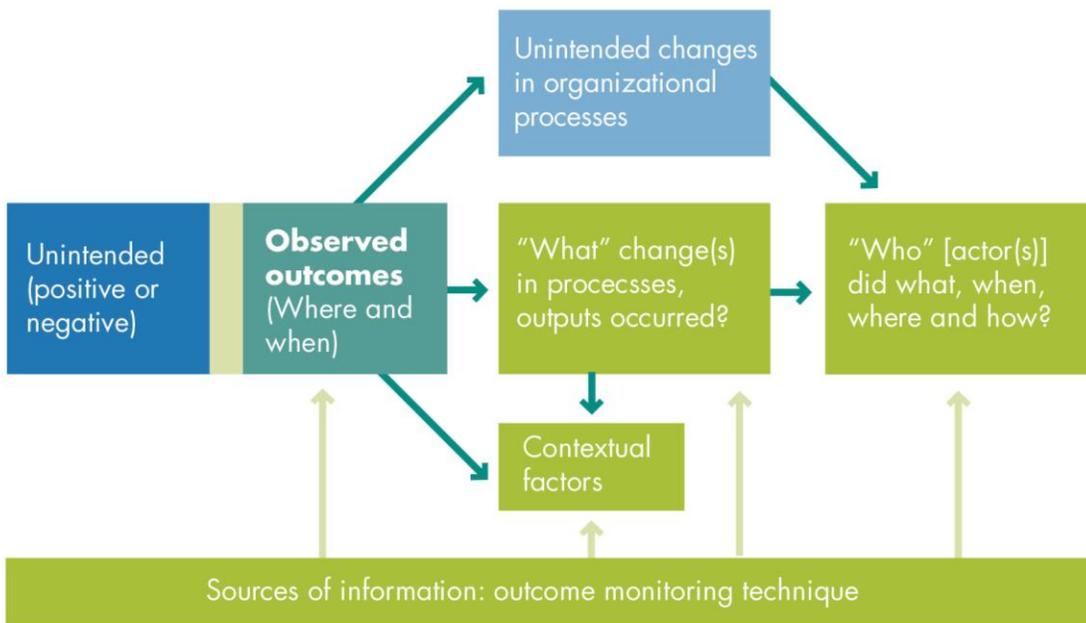
We described outcome monitoring in Chapter 1 as a collection of monitoring techniques based on qualitative methods, used for identifying emerging systemwide changes and clarifying causal relationships. In subsequent chapters, we continued defining systemwide indicators and changes, along with unintended changes and context tracking. We recommended using at least one outcome monitoring technique to collect systemwide changes and exploring causal linkages by elaborating what outcome monitoring entails. **The current section describes the basic steps of outcome monitoring and the various techniques that could be applied for outcome monitoring.**

Qualitative researchers have been developing methods and techniques to fill the gaps of quantitative methods in exploring causal linkages between development projects and their impact (Table 6). This is especially the case when these projects affect multiple components and levels, context is dynamic, and uncertainties exist about the project’s ability to produce the intended effects. These techniques are mostly employed during project implementation to identify emerging changes and their related factors to understand causal linkages. They account for nonlinear causal relationships as well. Since the major emphasis is on exploring outcomes and their causal linkages, these techniques are often referred to as outcome monitoring techniques in the literature, although outcomes are usually described as change in target beneficiaries, who could be people (e.g., change in behavior/practices), organizations (change in management practice), or changes in a component of a system (such as a health system function). **When the object of interest and unit of**

analysis is the system, outcome monitoring in the system could easily be labeled “systemwide monitoring.”

These techniques can explore two types of questions in HSS monitoring. First, what changes did the beneficiaries observe in the past six months (or any selected period) associated with the HSS activity? The changes could be in beneficiaries’ behaviors, in organizational processes and performance, or the context in which the project is embedded. Second, what factors caused the observed changes? These questions create linkage between changes and causes that may or may not be related to the project. However, they generate evidence during the project’s implementation on the efficacy of the project activities in bringing desired or unintended change. Figure 8 describes how the outcome monitoring process unfolds, collects, and synthesizes information. After identifying the systemwide outcomes, the underlying factors are traced backward by asking who, what, where, when, and how the outcome emerged.

Figure 8. Outcome monitoring, creating causal linkages between processes and outcomes



An important aspect of outcome monitoring is that it captures all emerging changes in the system’s organization, in stakeholders, and in beneficiaries, which are important but are not depicted by the indicators specified in advance. The beneficiaries’ participation in identifying changes not only empowers them to have a say in the results of the activity but also increases their ownership of the activity and political will to sustain it. Thus, outcome monitoring not only generates evidence that could be used for adaptive management or modification of the activity but also ensures ownership and sustainability. Table 6 lists examples of outcome monitoring approaches. One of these approaches could be part of the activity’s MEL plan for exploring the intervention’s causal linkages to outcomes. Table 6 provides some information on each technique, types of questions explored under each technique, and strengths and weaknesses.

We recommend that the USAID staff gain required knowledge and expertise in these techniques in order to supervise IP MEL teams to employ these techniques, collect and analyze information, and report accordingly. This will facilitate following new guidance on context and outcome monitoring for projects with multiple components, complexity, and limited clarity of causal linkages, as is the case with HSS projects.

Table 6. Selected outcome monitoring techniques

Method	Types of questions	Description	Strengths	Weaknesses
Ground truth solutions: www.groundtruthsolutions.org	Is the activity meeting the needs of the beneficiaries?	This hybrid method combines traditional social science models of participation with an approach to data collection and analysis adapted from the customer satisfaction industry. The approach is to ask very few questions (from 5 to 10), and to ask them frequently.	Method can capture change reliably, inexpensively, and frequently. Data quality improves as beneficiaries see results from their participation.	Not appropriate for all types activities; most examples come from disaster response.
Most significant change (MSC): http://www.mande.co.uk/docs/MSCGuide.pdf	What is the most significant change since we last visited? Do the outcomes vary greatly among beneficiaries? What will be the most important outcomes?	Tool for collecting, discussing, and selecting stories about the significant changes that people experience as a result of your programs. It involves people at different levels of an organization discussing the stories and then selecting the stories they consider most significant. It does not use indicators to measure progress.	Can be useful in situations where the outcome is uncertain and the approach is participatory.	It is very subjective and there is the possibility of bias toward success stories and popular views.
Outcome harvesting: http://usaidlearninglab.org/sites/default/files/resource/files/Outcome%20Harvesting%20Brief%20FINAL%202012-05-2-1.pdf	What happened? Who did it (or contributed to it)? How do we know this? Is there corroborating evidence? Why is this important? What do we do with what we found out?	Outcome Harvesting does not measure progress toward predetermined outcomes or objectives, but rather collects evidence of what has been achieved and works backward to determine whether and how the project or intervention contributed to the change.	Corrects the common failure to search for unintended results; has verifiable harvested outcomes; ties the level of detail provided in the descriptions directly to the questions defined at the outset of the process.	Skill and time are required to identify and formulate high-quality outcome descriptions. Only those outcomes that the informant is aware of are captured; participation of those who influence(d) the outcomes to be harvested is crucial.
Outcome mapping: http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/5058.pdf	Create consensus on desired change; results are measured by the changes in behavior, actions, and relationships of those individuals, groups, or organizations with whom the initiative is working directly and seeking to influence.	Outcome mapping (OM) is an approach to planning, monitoring, and evaluating social change initiatives developed by the International Development Research Centre (IDRC) in Canada (Earl, et al., 2001). At a practical level, OM is a set of tools and guidelines that steer project or program teams through an iterative process to identify their desired change and to work collaboratively to bring it about.	When tackling complex problems, building capacity, working in partnership, promoting knowledge, and influencing policy, a deeper understanding of social factors is critical; and to embed reflection and dialogue.	Consensus building take time and skills.

<p>Rapid outcome assessment (ROA): http://www.odi.org/publications/6800-rapid-outcome-assessment</p>	<p>Did this activity contribute to the changes observed? What other changes can be attributed to this activity?</p>	<p>The ROA is a learning method to assess and map the contribution of a project's actions on a particular change in policy or the policy environment. It is a flexible and visual tool that can be used in conjunction with other evaluation tools and methods.</p>	<p>Allows for the direct participation from different types of stakeholders to try to tease out the effects of the activity; can be done during and after an activity.</p>	<p>Can be costly and is time-intensive; a 3-step approach, and the second step is convening workshops.</p>
<p>Contribution analysis (CA): http://betterevaluation.org/sites/default/files/ILAC_Brief16_Contribution_Analysis.pdf</p>	<p>Has the program made a difference? How much of a difference?</p>	<p>An approach for assessing causal questions and inferring causality in real-life programs. It offers a step-by-step approach designed to help managers, researchers, and policymakers arrive at conclusions about the contribution their program has made (or is currently making) to particular outcomes.</p>	<p>Useful in establishing plausible causality when an impact evaluation can't be done.</p>	<p>May not provide statistical generalizability.</p>

The activity’s MEL plan should use at least one systemwide outcome monitoring technique to accomplish the following tasks:

- Gather data on the changes in stakeholders’ knowledge and behaviors, system processes, outputs, and outcomes
- Document how those changes occurred
- Exclude alternative explanations of causal linkages
- Verify evidence from different sources

Therefore, when reviewing the MEL plan, ask:

- Does the MEL plan describe at least one systemwide outcome monitoring technique, which captures causal linkages between the intervention and outcomes and excludes alternative explanations?
- Does the MEL plan describe how information on casual linkages is verifiable from different sources?

Action: If the MEL plan does not mention a systemwide outcome monitoring technique, ask the IP to revise the plan and address the missing information.

Setting up Data Collection and Management

Now it is time to review the MEL plan’s data management section. We will focus on the following six areas:

1. Data collection process, entry, and cleaning
2. Data quality assessment
3. Maintaining privacy and confidentiality of data
4. Analysis plan
5. Open data policy
6. Roles and responsibilities

Data Collection Process, Entry, and Cleaning

The MEL plan should state how the data will be collected, entered, and cleaned and the mechanism for feedback and data use: in other words, the activity’s information system. The plan should provide steps to develop data collection instruments and pretest them. The MEL plan should describe the data entry program and database with inbuilt validation to minimize data entry mistakes and a timeline for data cleaning and reporting.

Use of unique identification numbers to identify units of analysis—facilities, clients, regions, or districts—should be used. The IP should set up an algorithm to create identification numbers, to make clear what the numbers are identifying. For example, let’s say you are collecting data for clients in a facility in one region. You can assign the region’s code to be 80 and the facility’s code to be 22, and then assign a client number, 234. So, if anyone in the project saw the ID number 8020234, they would be able to identify the region and facility. Identification numbers are used both for quantitative and qualitative data.

Data Quality Assessment

The MEL plan should specify that it will carry out data quality assessments for activity indicators that are reported to USAID headquarters after the first year and after three years, as USAID’s data quality assessment (DQA) policy stipulates (ADS 200). The following chapter offers more discussion on DQAs.

Maintaining Privacy and Confidentiality of Data

The MEL plan should describe how it will maintain the privacy and confidentiality of data at the following level:

- During surveys, key informant interviews, focus groups (e.g., data collected in privacy; presence of informed consent; confidentiality agreement)
- Transmitting information from the field to headquarters (secure coded transmission system)
- During data cleaning, analysis, and reporting (limiting access to data)
- Making survey data available in public (such as removing personal identification information)

Analysis Plan

The analysis plan should describe what kind of analyses will be conducted. Here are the minimal things the data analysis plan should describe for evidence generation:

- How data will be used for analysis. Careful consideration should be given to this at the beginning. For example, will data from various sites be merged for final analysis? This aggregating of data should influence how data entry is set up, underscoring the importance of having experienced data analysts on the MEL team to ensure that data entry is conducted properly.
- Types of comparisons:
 - From baseline and target over time
 - By types of services over time, and by districts
 - By activity and nonactivity areas, etc.
 - How triangulation will be used to understand the monitoring of results
- Exploring association and causes:
 - Using quantitative data: teasing out effect by controlling for confounding factors using various statistical techniques assessing interaction and casual pathways analysis
 - Using qualitative data: case studies; systemwide outcome monitoring approaches such as most significant change (MSC).

Open Data Policy

All USAID operating units, including the agency's worldwide missions, must ensure that USAID-funded data are centrally cataloged and made available to the public by default, with limited exceptions. This is in keeping with USAID's evaluation policy (USAID, 2011a) and the USG's Open Data Policy (Executive Office of the President, Office of Management and Budget, 2013). The MEL plan should describe what data (survey; special studies) will be made available in the public domain with consent from all stakeholders, based on these policies.

Roles and Responsibilities

The plan should also indicate which staff member is responsible for each of the MEL functions. All staff members should have the expertise and experience to do their jobs. One or more staff members should be designated the MEL director and MEL manager. If the organization has a MEL skills gap, the MEL plan should explain how it will be managed. This could mean that key staff will receive MEL training or that an external MEL group will be consulted. This part of the MEL plan should also outline responsibilities for reporting and delegation. For example, who will be responsible for conducting DQAs? Who will be responsible for regular review of data coming from subcontractors? Who will be responsible for problems in data quality?

Therefore, when reviewing the data collection and management section of the MEL plan, ask:

- Does the MEL plan describe data collection process, entry, and cleaning?
- Does the MEL plan describe data quality assessment and its frequency?
- Does the MEL plan describe means to maintain clients' privacy and the confidentiality of data during data collection, transmission, analysis, and reporting?
- Does the MEL plan describe an analysis plan with details on unit and types of analyses?
- Does the MEL plan state triangulation of specific indicators through various sources of information for validation of information collected?
- Does the MEL plan describe what data (survey; special studies) will be made available in the public domain with consent from all stakeholders, according to the open data policy?
- Does the MEL plan describe roles and responsibilities of MEL staff and management?

Action: If the questions are not answered appropriately in the MEL plan, ask the IP to revise it and address the missing information.

Establishing Baseline and Targets

All indicators must have baselines and targets. Some of this information may not be available within the first 90 days of the award, but the method for collecting baseline data should be laid out (see Box 5 for some methods to establish baselines). The MEL plan can be updated once all the baselines have been conducted and all targets set. The plan should also indicate if any of the data can be reported by sex.

Let's look at the ANC activity described in Scenario 3 and assume that the following indicators were selected:

- Number of women attending at least one ANC visit by a skilled provider at a USG-assisted facility
- Number of women attending at least four ANC visits by skilled providers at USG-assisted facilities
- Number of skilled deliveries at USG-assisted facilities
- Quality of maternity services offered in USG-assisted facilities

Let's assume that three of these required indicators are routinely collected as part of the national HMIS system. It was determined that in the previous year in the selected region, 52 percent of pregnant women attended at least one ANC visit, 29 percent completed at least four ANC visits, and 28 percent of deliveries were skilled deliveries. These values can be used as baselines for these three indicators, but no baseline values exist for the quality of services provided, because this is not routinely collected. Therefore, the IP must collect data on this indicator to establish a baseline value. To do so, the IP proposes to conduct a medical record audit in a sample of facilities to determine the percentage of adherence to clinical standards for ANC care.

Box 5. Methods to establish baselines

Baseline values can be determined using existing sources or by collecting baseline data as part of the project and activity start-up.

- **Existing sources:** Baseline values can come from government data—whether annual health statistics reports or through data collected from the RHIS. Other sources of baseline values are population-based or facility surveys (i.e., DHS, MICS for health indicators, and Service Provision Assessments and Service Availability Mapping for facility information) and data collected by another IP or bilateral organization. The important thing to remember is that the data for subsequent data collection must come from the same source or be collected using the same methods and tools.
- **Collecting baseline data:** When the baseline data are not already available, then it is necessary to collect them. The types of baseline surveys will depend on the type of data (quantitative versus qualitative) that you are trying to collect as well as the unit of analysis (i.e., households, individuals, women, facilities, districts, and government departments). Some examples of methods to collect baseline data are as follows:
 - **Quantitative data:** questionnaires, checklists, and tests (e.g., you want to establish a baseline for the percentage of district HMIS that can read and interpret a graph correctly)
 - **Qualitative data:** interviews, focus groups, direct observations, document analysis, and case studies

Once baseline values are established, it is important to set targets. Targets need to be realistic, taking into consideration resources that are available and based on evidence but also ambitious: setting low targets is not useful. Targets can be set various ways, such as the following (USAID, 2010):

- **Examining historical trends:** A baseline value is just the starting point. If data are available that allow you to see how an indicator has changed over time (five or ten years), then that can help you set a realistic target.
- **Soliciting expert judgments:** This involves soliciting opinions from experts who are knowledgeable about the program area (whatever specific HSS core function the activity covers) and the country context to provide guidance on what is possible or feasible with respect to an indicator and country setting. For Scenario 3, maternal health experts could be convened to discuss the targets for the quality improvement program.
- **Investigating research findings:** Research and evaluation findings from similar programs in other settings may help you choose realistic targets. This may be more difficult for some HSS activities that have not been extensively studied, in contrast to some other program areas (e.g., reproductive health) that have ample evidence to guide realistic target setting. This underscores the importance of building the evidence base for HSS.
- **Comparison with similar programs:** Comparing the achievements of similar programs can also be helpful in setting targets. This may involve looking at the HSS work that is being done beyond USAID.

An absolute target is not possible, because there are always variations in performance. Thus, it is always good to create variation boundaries of lower and upper limits of performance or control limits. These boundaries ensure that all stakeholders conform to the performance requirement and assist us in identifying special cases where performance is below or above the control limits. This can help us learn from what is happening in these cases and improve (see Box 6).

Therefore, when reviewing the target selection section, ask:

- Does the target setting describe who provided expert judgments?
- Is the target setting based on existing research findings and does it identify these references?
- Does the target setting use historical comparison and identify the source of data?
- Does the target setting use benchmarking or comparison with similar programs and identify those programs?

Action: If answers to these questions are negative, ask the IP to provide the answers in the revised MEL plan.

Box 6. Control limits for targets

Control limits are established for targets. These are usually set to allow 80 percent of the units of analysis (districts, facilities, etc.) to remain within the variations and distinguish those that are constantly falling below or above variation (control) limits. By identifying those that fall below and above the control limits, we learn what makes these facilities perform higher or lower than the majority and use that knowledge to improve performance. If most of the units are performing above or below the control limits, then it could mean that targets and variations were set too low or too high. Then, the target and its variations must be adjusted.

Reviewing the Approach to Promote the Use of Collected Data at Different Levels

You have reviewed how the IP would collect, manage, analyze, and report data at regular intervals. However, many IP plans do not describe how the project stakeholders at different levels can and will use this information. Collaboration among the stakeholders is key not only to successful implementation but also for sustainability. The activity managers and government officials at district or higher levels and facility staff all need to share these data for decision making and learning. Here are some key issues to consider:

- Is a process in place to develop the capacity of the beneficiaries to strengthen their use of information?
- How can variations in performance among facilities/districts/regions be used to promote learning and improve HS performance at these levels?
- Is the collected information answering the main monitoring questions and generating evidence on the effectiveness of the activity?
- Is the host government developing new policy or program or changing processes to streamline implementation?
- Is there a process to document decision making at different levels?

The MEL plan should be able to describe an approach to strengthen the use of information, given the limited resources at hand.

Therefore, when reviewing the approach to promoting the use of information, ask:

- Does the MEL plan have a section on promoting the use of information?
- Does that section provide a plan for capacity building of the host government officials and managers?
- Does the section explain how the information will be used at the community, facility, and higher levels?
- Does the section explain how the activity will assess the use of information?
- Does the section describe the consequences if the collected information is not used?
- Does the section discuss linking the use of information to HS functions and performance?

Action: If the answers are unsatisfactory, ask the IP to address the missing information.

Reviewing the Evaluation Plan and Additional Studies

We will spend more time discussing evaluations in the final chapter of this guidance document, but there is also an evaluation section in the MEL plan. It asks the IP to describe any process or impact evaluations it plans to conduct. The IP should state some possible research questions to investigate as well as possible research methods to address the research questions. The IP may also suggest special studies or operations or implementation research in the MEL plan.

Therefore, when reviewing the evaluation section, ask:

- Are the topics and evaluation questions in alignment with the USAID mission or headquarters learning agenda and referenced appropriately?
- Are the justifications for the topics of special studies accompanied by a literature review?
- Are the priorities described among the listed topics along with priority-setting criteria?
- Are the stakeholders (such as the health ministry, USAID CDCS, donors, and the private sector) identified as having suggested the topics for special studies?
- Are planned evaluations (midterm, end-of-the-project, or impact) and their schedules described?
- If an impact evaluation is planned, are the details provided?

Action: If the questions are not answered satisfactorily, ask the IP to address the missing information. If an impact evaluation is described, check who will conduct the evaluation and when the evaluation protocol (including the data collections instruments) will be submitted for review.

Key Messages

- Ensure that the activity's TOC and RF are aligned with each other. Encourage the development of a separate TOC for each subactivity, to facilitate better implementation and MEL.
- Confirm that the performance, systemwide, and context indicators are described in alignment with results and have appropriate sources of information.
- Verify that systemwide effects, unintended results, and causal linkages are captured using at least one systemwide outcome monitoring approach.
- Confirm that the collection, management, quality assessment, triangulation, and validation of data are described in the MEL plan.
- Ensure that an analysis plan is available to generate evidence in a way that excludes alternative explanations from various sources.
- Check that the use of information is part of the MEL plan.
- Determine that baseline and target setting is part of the plan for tracking progress.
- Ensure that the plan includes the rationales for types of evaluations and special studies.

CHAPTER 6. MEL IMPLEMENTATION: BUILDING EVIDENCE THROUGH MONITORING

Introduction

After the MEL plan is approved, it is necessary to ensure that it is implemented as described, the data are analyzed, and observed changes are reported. This chapter is divided into two sections. The first section deals with reporting performance data, systemwide effects, contextual data, unintended results, and the synthesis of data to create evidence of change during the quarter. The second part of this chapter deals with accountability mechanisms, sharing data with other IPs and DQA by the mission staff, or the mission hiring a third-party contractor or consultant.

Scenario 4

Think about what you would like to review in the monitoring report presented in the scenario?

An IP working in southern Africa is implementing activities to build capacity for pharmaceutical management and services. This includes the development of accredited training programs to increase the number of qualified pharmaceutical professionals and updating existing training curricula to include HIV and AIDS pharmaceutical and supply chain management. Stakeholder analysis identified the need to have a midlevel cadre of pharmacy assistants who will be placed at lower-level clinics to focus on stock management. The curriculum includes modules on HIV and TB management, family planning, and other primary healthcare components.

Monitoring Report

Mission staff overseeing the activity should receive a monitoring report documenting its progress. At minimum, this report should contain the following sections—information that can help you guide the IPs on how to prepare monitoring reports and review the reports once submitted.

Performance Tracking

The performance monitoring based on inputs, outputs, and outcomes indicators is paramount to answer the following questions:

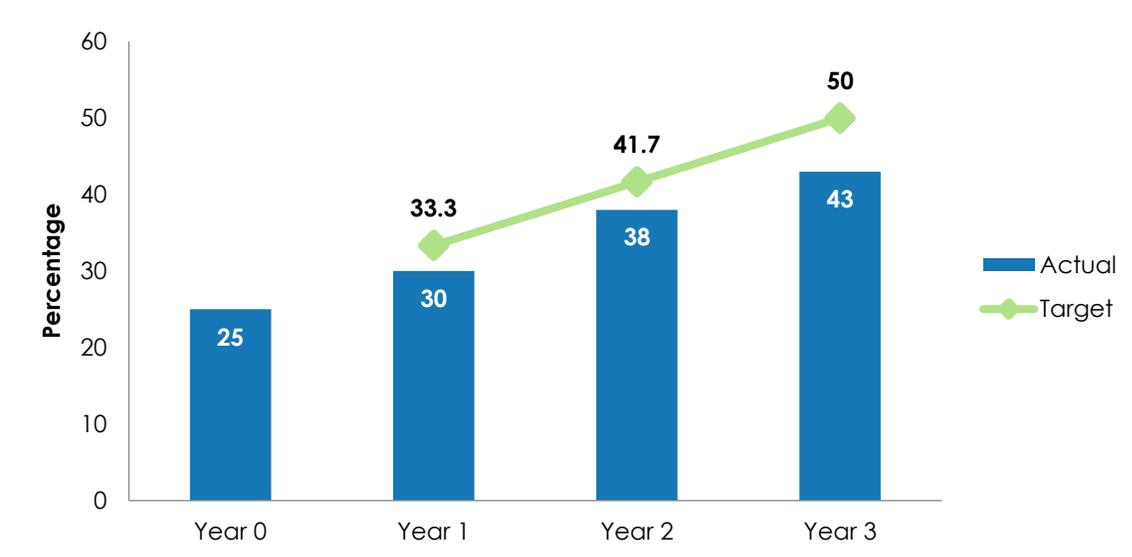
- Are services meeting the monthly or quarterly targets for the reporting period and over time? In other words, is their overall performance improving?
- Are most of the services' monthly or quarterly performance targets within the control (upper and lower) limits for the reporting period and over time? Where is performance outside the control limits and why?
- Does performance vary by districts or regions?
- Is there a difference in performance between activity and nonactivity areas (if included for comparison) or between performance in activity areas and the national or regional performance average?

Of the many ways to present performance data, visual illustrations—tables, graphs or charts, and maps—are typically the most effective. Some IPs also use dashboards to help track performance.

Let's look at how the first question (Are the monthly/quarterly targets for various services being met over time?) could be reported. For example, let's assume that one of the performance indicators for the

pharmaceutical strengthening activity described in Scenario 4 is “Percentage of public health facilities with certified pharmacy personnel” in four districts of region ABC of a country. The target was for 50 percent of the facilities in all four districts in Year 3 to have at least one certified pharmacy staff member. In our example, the baseline in Year 1 was 25 percent of facilities. The target was equally divided for each year. Figure 9 presents the activity’s achievement for the entire region. Despite consistent improvement over time, the activity failed to achieve the yearly targets. Is that because the targets were set too high, or because resources to achieve the targets were not provided? Are there some other reasons for not achieving the target? We need further analysis to answer this question.

Figure 9. Improvement over time to meet the target of 50% of facilities having at least one trained pharmacist in Year 3



Let’s now look at how we can report the second question (if performance is within the upper and lower limits of the target). If we compare only actual performance with the target, we might misinterpret performance. Variations are allowed in performance based on historical norms or expected outcomes. A control chart presents set variations in performance. It is not routinely used for monitoring, but this guidance recommends its use to observe variation in performance more clearly.

Let’s assume from Scenario 4 that we want to improve availability of contraceptive methods in the target area, where stockouts often led to disruption in use. The baseline was 25 percent of facilities reporting availability of all contraceptive methods and the target was set for 50 percent in one year—a 6.25 percent increase per quarter. Figure 10 shows that although the target was not met for the second and third quarters, the percentage of facilities having all contraceptive methods was within the lower control limit. Availability of methods improved over time at a slower pace than expected. If we compare only against the absolute target, we would conclude that “availability of methods” did not achieve the target. Allowing variations, we see that availability of methods was within the lower control limit of targets, suggesting that there may be some systemic issues to address to improve performance.

Figure 10. Percentage of health facilities reporting availability of contraceptives, target versus actual, Quarters 1–3

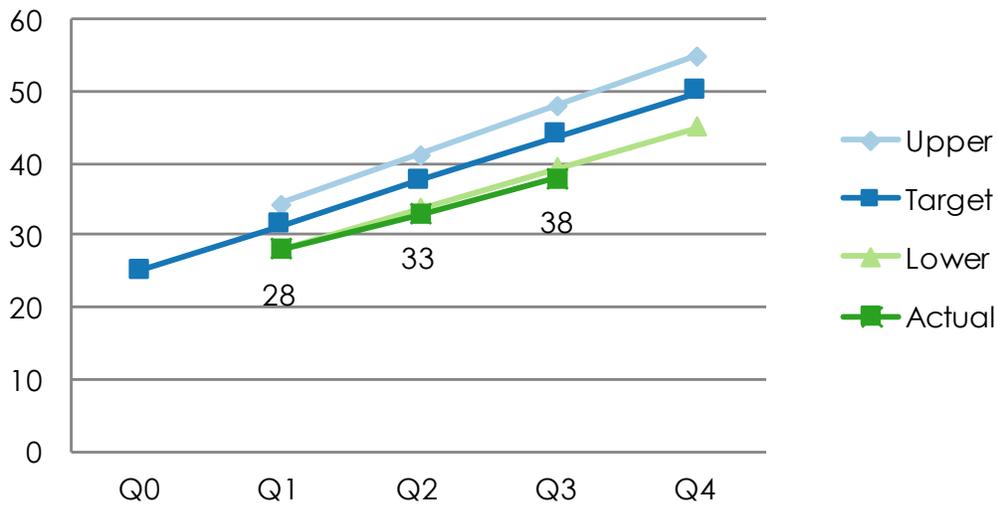
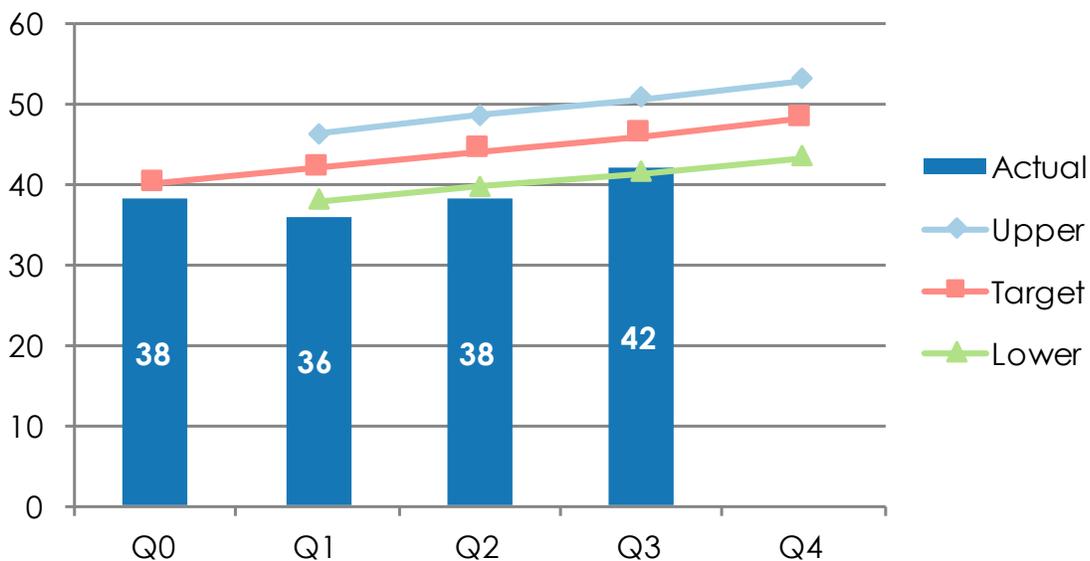


Figure 11 looks at women who attended at least four ANC visits (ANC4). This example is a more effective visualization of comparison, because it presents the actual performance values as bars. The baseline value for ANC4 coverage was 40 percent, and initially, ANC4 coverage decreased in the first quarter before it returned to its baseline value in the second and then improved in the third. It is possible that project initiation disrupted services, but improvements were observed in the second and third quarters. However, coverage reached only the lower limit of the third quarter’s target.

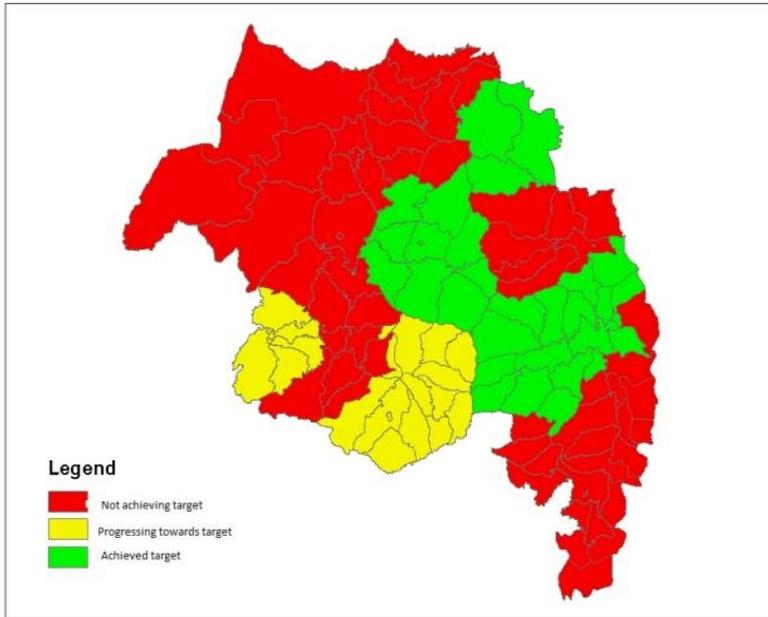
Figure 11. Tracking ANC4 coverage over one year



Differences observed in the activity’s performance by district or region is the third question. The activity’s overall performance data are usually presented as the average level of performance within activity areas. However, this presentation masks variations in performance across geographic areas, which can be useful in

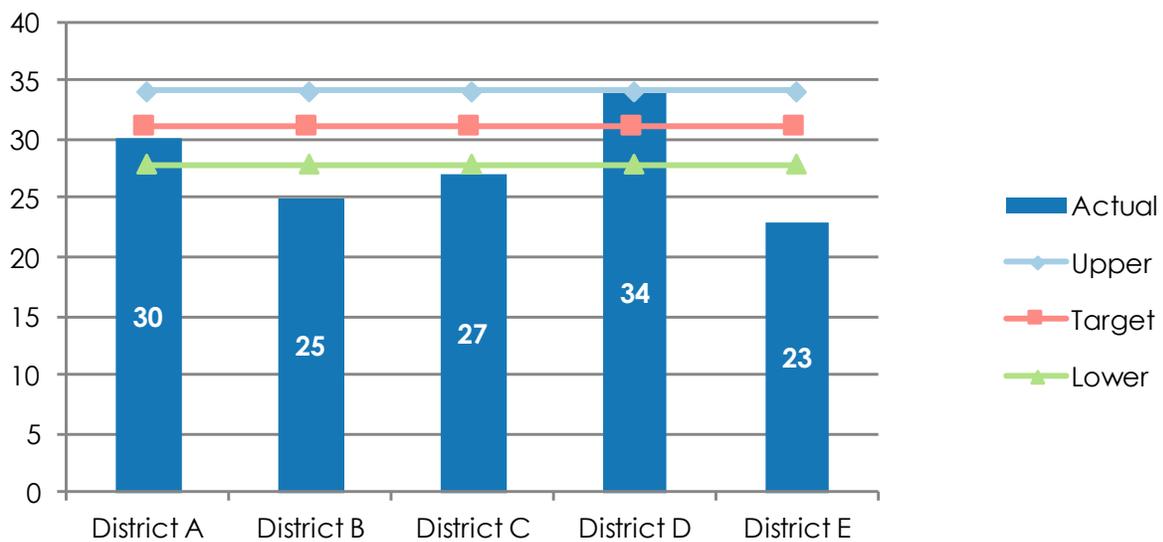
identifying areas of low and high performance, setting priorities, and allocating resource. A map can display these variations (Figure 12).

Figure 12. Achievement of targets across the region



A graph can also help us observe the variation in performance in five districts (Figure 13). We can identify which districts are performing below the lower limits, but we don't know why this is happening. We need additional data to fill in the picture—to be discussed later in this chapter, in the section on outcome monitoring.

Figure 13. Comparing contraception coverage among five districts in the first quarter



The final question deals with differences in performance between activity and nonactivity areas or at the national and regional levels. Comparing performance in activity areas and similar nonactivity areas can provide some evidence that a change is occurring because of the activity. This assumes that other alternative explanations are also excluded and that the nonactivity areas have similar socioeconomic and demographic characteristics to make the comparison valid. The same can be said about comparing the activity area's performance with the provincial or regional performance average (after excluding the activity areas). Figures 14 and 15 illustrate how this information could be reported, using coverage of two tetanus toxoid injections (TT2) and postnatal care (PNC) coverage. Figure 14 shows that TT2 and PNC coverage improved over time in the activity areas in comparison with nonactivity areas, indicating that the activity might have contributed to that increase. Figure 15 shows that TT2 and PNC coverage were lower than the provincial averages, indicating that the activity areas were worse off than the rest of the region at the baseline—probably the reason why they were selected for the activity. However, the activity areas showed some improvement in services in the first quarter and the provincial averages remained the same, indicating that the activity contributed to the change, while no change was observed in other areas of the province, because the activity was not implemented there.

Figure 14. Comparison of services between activity and nonactivity areas

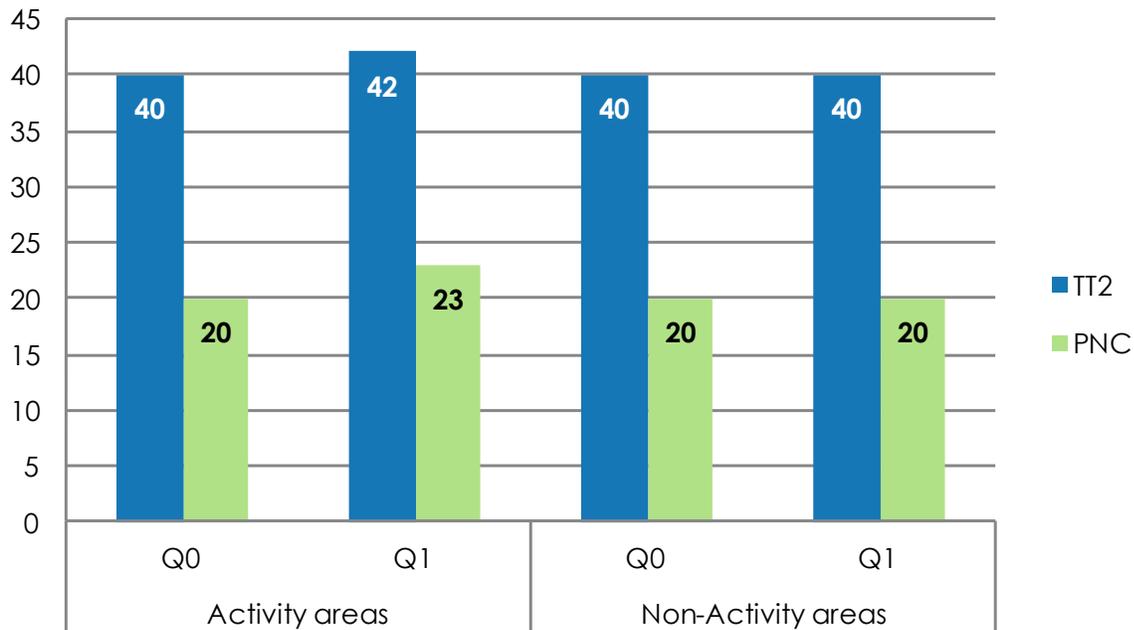
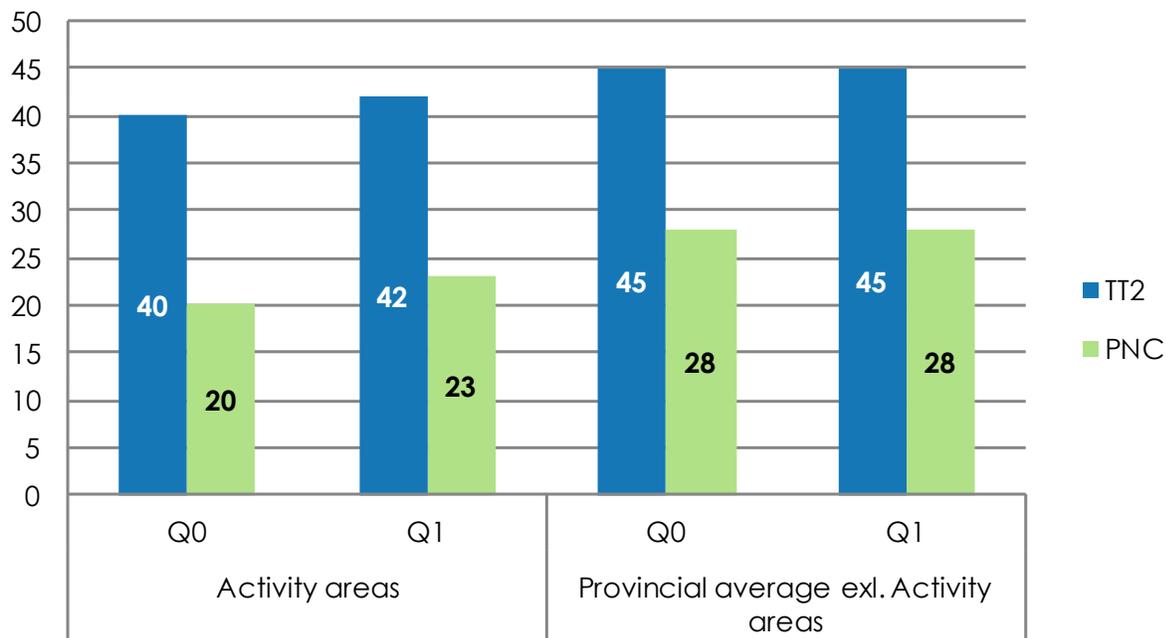


Figure 15. Comparison of services with provincial average, excluding activity areas



Therefore, when reviewing a quarterly performance report, ask:

- Does the monitoring report describe performance over time against targets?
- Does the report illustrate variations through control charts on services against target ranges?
- Has the IP reported variations among districts or regions in the quarterly reports?
- Does the report compare activity areas to nonactivity areas and activity areas to provincial or regional averages excluding the activity areas?

Reporting Systemwide Quantitative Indicators

If the systemwide quantitative indicators and their sources of information are part of the MEL plan, then the IP can report these indicators in the same manner as the performance indicator data discussed earlier.

Therefore, when reviewing a monitoring report, ask:

- Did the report describe the systemwide quantitative indicators, if applicable?

Reporting Contextual Factors

The main purpose of reporting contextual data is to determine if contextual factors are causing observed variations in performance indicators. The reporting mechanism for contextual indicators and its frequency of use is decided by the mission and the IP. This information could be presented as yes/no binary indicators. Table 7 is one example of how an IP could present contextual indicators. Explaining the tables, the IP can provide more details on how the emerging contextual information could influence the activity's implementation and outcomes.

Table 7. Tracking contextual indicators

Contextual indicators	Q1 Y/N	Q2 Y/N	Q3 Y/N	Q4 Y/N
Existence of water and sanitation project implemented by the government				
Presence of road construction in the area				
Existence of a family planning national communication campaign				
Presence of any natural and manmade disasters or events that could potentially affect the activity (specify)				
Any other development project that can affect health outcomes (specify)				
Any other development project that can affect health outcomes (specify)				
Manmade or natural disaster (specify)				

Therefore, when reviewing the monitoring report, ask:

- Did the report describe any new factors or changes in selected contextual factor(s) of the activity during the quarter and use them to explain performance?

Reporting Systemwide Changes/Effects Using Outcome Monitoring Techniques

Systemwide changes observed through outcome monitoring can be reported in a brief narrative. The narrative should be able to show perceived/objective causal linkage(s) between the intervention and effects from stakeholders' perspectives. The narrative should describe not only what changed but also why those changes occurred. The validation of evidence for attribution or contribution should meet the qualitative standards of repeated responses, triangulation from various sources of information, and exclusion of alternative explanations. Thus, it would be possible to determine if the interventions are working as planned, make course corrections, and adapt, if needed. Box 7 presents an example of reporting using specific outcome monitoring techniques.

Therefore, when reviewing a monitoring report, ask:

- Did the report describe the systemwide changes/effects of this activity for the quarter, meeting standards of evidence generation such as observation, repeated responses, triangulation, or other means?

Monitoring Unintended Results

One of the important aspects of monitoring in HSS is to identify unintended effects or consequences of the activity, which were not anticipated during the design of the project. Therefore, the quarterly or six-monthly report should state any unintended results—especially any negative results—to ensure corrective actions.

These unintended results could easily be gathered through supervisory visits or their reporting could be part of the project reporting mechanism. The unintended effects could also be collected and reported through the

outcome monitoring. The data should be collected from the activity's beneficiaries and implementers, supervisors, program managers, and policymakers (see Box 7 for an example).

The usual questions to ask to collect information on unintended effects are:

- Did you find any changes in your work and/or in beneficiaries and/or in the HS due to the activity implementation that is beyond your expectations?
- Did you find any unusual events in your work or in the activity's beneficiaries that are due to the activity implementation?

Box 7. Example of a description of unintended results

For Scenario 4, an unintended result could be reported like this:

During supervisory site visits of three facilities, it was discovered that two pharmacy assistants who had been trained had left government service. The facility head said the assistants had gone to work for a private pharmacy, where they are getting better remuneration, causing concerns that others may leave as well. It was also reported during the site visit that the productivity of the supervising pharmacists had declined because they must spend a significant amount of time supervising and supporting the pharmacy assistants.

Therefore, when reviewing a monitoring report, ask:

- Did the report describe any unintended positive or negative changes in the project areas during the quarter?

Reporting the Use of Information

The information collected must be reported, but it should also be shared with all stakeholders for decision making. The use of information at the community level and higher shows how the different stakeholders are making decisions to plan and manage the activity’s implementation. It also reflects ownership of the data and, consequently, ownership of the activity. Therefore, the quarterly monitoring report should document what actions were taken based on the collected information. These major decisions could be reported as in Table 8 or through narrative or other means, as appropriate:

Table 8. Types of the use of information by different levels

	Planning	Management	Monitoring progress and follow-up of planned actions	Success story	Role-modeling of information use	Advocacy (using information to ask for policy/process change or resource allocation)
Community						
Facility						
Districts						
Region/national						
IP managers						

Therefore, when reviewing a monitoring report, ask:

- Does the monitoring report describe the purposes/decisions for which the collected information was used at different levels?

Synthesizing Data and Creating Evidence

So far, we have described how to report performance, contextual, and outcomes monitoring data showing achievements as well as systemwide and unintended changes. This reporting is necessary to show specific changes, but it does not provide the following information:

- A holistic picture of the activity’s progress
- Whether the findings are consistent with different sources of information
- Whether causal linkages between the activity’s intervention and observed changes can be identified
- Whether alternative explanations for observed changes can be ruled out

A synthesis of the findings could help answer these questions. As previously mentioned, monitoring does not lead to exclusive attribution conclusions but it does point to the contribution of the activity to the changes observed. Therefore, the IP should synthesize the collected information and highlight evidence that supports

linking observed changes to the activity's implementation. The reader should note that IP in the first few quarters of the activity may not provide detailed information on observed changes, as there is often a lag period before the changes due to the activity become apparent. However, as the data from performance, contextual, and outcome monitoring accumulate over time, it is possible to observe changes, trend(s), and contributions of the activity.

- The first step in synthesizing information and creating evidence is to triangulate the information with information from other sources to check reliability and consistency. Are the supervisory reports and facility reports consistent? Do the RHIS data show consistency among service coverage, availability of staff, and medicines/vaccines?
- Second, do the reported systemwide effects (changes in other HS functions) align with the TOC?
- Third, is the performance variation in the target districts consistent with the intensity and magnitude of the interventions in those districts or some other factors?
- Fourth, are the results different in activity and nonactivity areas using the same or other sources of information?
- Fifth, do the systemic changes and outcomes observed through outcome monitoring support improvement in performance data?
- Last, do the contextual monitoring data explain any variation in overall activity performance over time and variations among districts over time?

By answering these questions from the sources of data collected, the IP can create evidence of the activity's contribution to strengthening the HS and its outcomes. Examples of conclusions that might be drawn from data synthesis are presented in Boxes 8 and 9.

Box 8. Synthesizing data and drawing conclusions: pharmaceutical system

This activity is about strengthening the pharmaceutical system, by creating a new cadre of pharmacy assistants in Regions A and B in country MNOP. The monitoring data from this quarter showed that training of pharmacy assistants continued as planned and another 60 trained pharmacy attendants were added to the workforce, for a total of 200 trained pharmacy assistants.

An unintended change was observed when we reviewed HRIS data for the quarter. They showed that 30 out of 140 newly trained pharmacy attendants left their government positions. Of those who left, 90 percent were deployed in rural areas. Therefore, it is necessary to investigate why. Is rural deployment the cause? What other factors contribute to this attrition? This information can be used to improve the intervention for better retention. RHIS data also showed that facilities reporting stockouts went from 30 percent to 25 percent over two quarters, suggesting improvement in the logistics management of medicines and vaccines at the facility level.

The outcome monitoring data showed that service providers are satisfied with the pharmacy workers. They reported that medicines are dispensed correctly and there is better compliance by clients, because pharmacy assistants explain correctly how to use the medicines. The service providers also reported that they could save time in counseling clients about the medicine and use it for other activities. The contextual data showed that there were no major natural or manmade disasters to disrupt the activity's implementation. However, another donor started an HIV treatment facility in the area during this quarter. The activity will follow whether and how that new facility might affect the activity's implementation and results in the future. Given that no other projects are training and supporting pharmacy assistants, we conclude that positive changes observed in this quarter are more likely to be due to the activity.

Box 9. Synthesizing data and drawing conclusions: malaria example

This activity is about changing practices related to malaria prevention- and treatment-seeking behaviors in Region DNC of Country QRST. The number of malaria cases dropped from 100 to 75 in this quarter; there were 105 cases in the same quarter in the previous year. Pregnant women taking prophylactic antimalarials increased from the baseline of 25 percent to 27 percent in this activity's first quarter, which is within upper and lower limits of the quarter target. Comparison of malaria cases and pregnant women taking prophylactic antimalarials in activity and nonactivity areas showed the same pattern so was the case for comparing with regional average for malaria cases. The contextual information showed that the government is distributing insecticide-treated bed nets (ITNs) to households in malaria-endemic areas. The outcome monitoring data showed that more pregnant women and children are sleeping under ITNs and are getting more information on preventing and treating malaria from the community health workers and community leaders. In District A, a road was broken after a landslide and the distribution of ITNs was affected for two weeks. The data from different sources of information are consistent, showing some improvements in preventive- and health-seeking behaviors as well as more malaria cases in the part of District A where the distribution of ITNs was disrupted. Given that this activity is the only malaria prevention program in this area, it likely that these improvements are due to the activity's implementation. Monitoring over many quarters will confirm whether preventive- and health-seeking behavior for malaria persist over time and result from the activity's contribution.

Therefore, when reviewing the monitoring report, ask:

- Does the report synthesize the data from different sources to create evidence of the changes that occurred during the quarter?

Data Quality Assessments

Mission staff are responsible for monitoring the quality and timeliness of outputs produced by the IPs. Part of their role is to ensure and verify through site visits and DQAs the following:

- Activity-level performance data are accurate (e.g., disaggregation and other calculations are correct)
- Reported data meet minimum data quality standards
- Verification documentation is being maintained (e.g., photos pre-, post- and during construction of a facility, or original daily sign-in sheets with training participants' signatures or thumbprints)

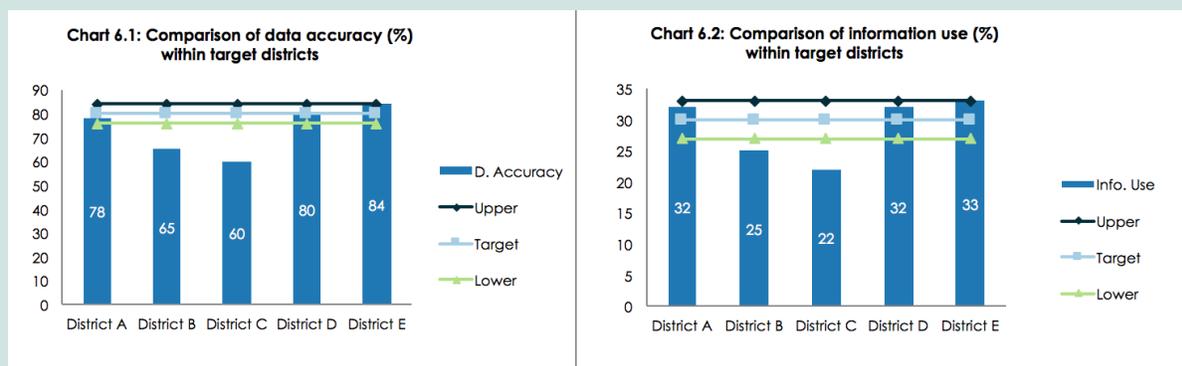
Assessing the quality of data being generated to track the activity's implementation is an important step in building HSS evidence. Poor-quality data call into question the reliability and validity of any results observed. There are two mechanisms for DQAs: mission DQAs conducted on selected indicators every three years and annual DQAs conducted by the IPs. IPs are not required to conduct DQAs, but this is a good management practice that should be encouraged. The IP will establish a data collection and reporting system, a database, and mechanisms to ensure data editing, cleaning, removal of duplicates and typographical errors, and other corrections made during implementation. We recommend that the IP conduct a DQA every six months to assure data quality. It serves two purposes: (1) continuous assessment of data quality; and (2) knowing when to act to improve data quality and set new bars.

Mission staff should ask IPs to submit DQA reports at least annually, although we recommend that the IP encourage its staff to conduct a DQA quarterly and submit the results along with the monitoring report. By reporting data quality to the mission, the IP shows commitment to data quality, improving transparency and increasing trust in the data submitted.

The IP should be able to state what has been done to address any possible data quality problems and promote information use. Box 11 provides an example of how DQA results could be reported.

Box 10. Example of a DQA report

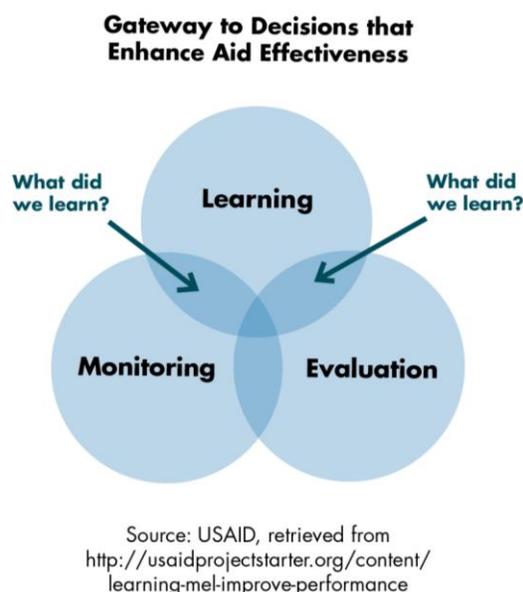
“We found the overall data accuracy at 61 percent and information use at 24 percent using LQAS this quarter. The data accuracy and information use disaggregated by districts are presented in Charts 6.1 and 6.2, which show that two of the three districts are not meeting the quarterly targets. The reasons for low data accuracy and information use were discussed at the district and activity staff meetings. It was decided to train staff on how to improve data entry and interpret data to improve their use. The results of this action will be reported in the next quarterly monitoring report.”



Learning and Adapting

Learning and adapting are important components of the project life cycle (Figure 16). The activity’s MEL data provide what is being learned on a regular basis and over the life of the activity. Besides learning that the activity is being implemented as planned and what progress it has made, the IP should report whether any innovations have been made, whether knowledge or skills that the activity has gained are being used outside of it, whether gaps in the TOC have been identified, and what attempt has been made to fill them. Is there sufficient evidence to make a course correction? The IP should share the quarterly information with the activity’s stakeholders, public and private, to get their feedback on what they have learned from the quarterly, half-yearly, and annual reports and, moving forward, what their role and responsibilities will be in achieving the activity’s goal and objectives.

Figure 16. USAID learning and adapting



Another level of learning is, “What is the relative contribution of the activity to the purpose and goal of the mission’s CDCS?” Therefore, it is expected that the activity IP, in consultation with the mission, will ask all other IPs to share their information. Sharing information helps in developing the evaluation plan for the activity and CDCS. It also helps local actors, such as government partners, understand what each partner is doing in their region, so meetings with the partners are recommended.

The process of portfolio review during project implementation may identify new learning, changes in the international development context, or problems in implementation that point to possible new directions or approaches. If so, the mission may need to add, change, or discontinue activities. Portfolio reviews are usually done once a year (before annual reports are submitted) but a mission may choose to do them twice a year.

Key Messages

An activity’s monitoring report may consist of the following items:

- Performance tracking results, including selected systemwide quantitative indicators
- Contextual monitoring results
- Outcome monitoring: systemwide effects, unintended consequences,
- Documentation of information use at different levels for policy and management decisions
- Data synthesis for evidence generation
- Data quality assessment report, when needed
- Synthesis of what has been learned from monitoring

CHAPTER 7. BUILDING THE EVIDENCE BASE FOR HEALTH SYSTEMS STRENGTHENING THROUGH EVALUATION

Introduction

This final chapter discusses HSS evaluations. It describes the uniqueness of HSS evaluations and distinguishes between impact and performance evaluations. Guidance is also provided on what to include in an HSS evaluation SOW and how to review the evaluation protocol. Management of the evaluation and the process of analysis, report writing, and dissemination of findings is described. The chapter ends with a description of how to learn from the evaluation findings.

Scenario 5

Think about the issues and steps you would take to help the mission plan evaluation in the scenario below.

A USAID mission in East Africa has just awarded a five-year activity to an IP to support a wide array of HSS interventions. Among these are strengthening the country's community-based health insurance program, by improving financial management and purchasing of health services, working with district health management teams to improve their ability to develop and implement district action plans, introducing hospital accreditation systems, and supporting the availability and use of routine health information in all districts in the country. The focal interventions are part of an HS reform process that has been ongoing for the past few years. Other IPs have also supported the government's HSS strategies, and will continue to do so over the next five years.

Importance of HSS Evaluations

Evaluation within USAID has two primary purposes: accountability and learning. In USAID's Evaluation Policy, evaluation is defined as “the systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments to improve effectiveness and inform decisions about current and future programming” (USAID, 2011a). This definition implies that a program's or project's strategy and characteristics can play a large role in determining the program's or project's outcome and effectiveness, evidence of which can improve future programming. There is a dearth of evidence to show that HSS interventions make a difference in improving HS performance and saving lives (Hatt, et al., 2015). Therefore, it is extremely important for USAID missions to ensure that evaluations are built into the design of USAID-supported HSS projects and activities, to improve the evidence base on what works (USAID, 2017b).

What Is Unique about HSS Evaluation?

The following factors make HSS evaluations unique and have implications for HSS evaluation design:

- **Mediation:** HSS interventions can be divided into distal and proximate levels to produce HS outcomes and impact (see USAID's *Vision* framework). For example, service delivery is proximate and has a direct relationship to health outcomes. However, HS functions such as governance, information, and financing are located at more proximate levels and affect health outcomes indirectly, through service coverage and quality. Thus, evaluators attempting to assess an intervention's effect should pay attention to its location in the TOC, because if an evaluation does

not consider mediating factors, it can wrongly conclude that an intervention has no effects on HS performance or impact.

- **Interaction:** Some HSS interventions in an HS core function do not act independently to change the HSS outcomes or impact alone but in interaction with the other HS functions. An interaction between two HS functions means that the effect of one of those functions on the HSS outcome or impact depends on other aspects of the HS working well. For example, medical staff depend on the availability of equipment, medicine, and vaccines for diagnosis and treatment. If the intervention is to increase compliance with quality-of-care standards, the availability of medicines or functional equipment will also affect the quality of care. Trained staff alone will not be enough to overcome these other HS function deficiencies. The evaluation should be able to identify possible interactions. This implies that it is important to include in the evaluation information on relevant functions and their interactions.
- **Systemic changes:** HSS interventions are intended to bring about systemwide changes, which means they cause changes in other HS functions. Therefore, it is important to gather data on all HS functions to assess the intervention's direct and indirect effects on the HS.
- **All things being equal (ceteris paribus):** As stated, the effect of an HSS activity on an HS function depends on the degree to which other HS functions are working well, which may vary across and within regions and districts. Because there may be significant variations in HS performance within countries, an impact evaluation should address this issue by including randomly distributed intervention and control groups. This underscores the importance of collecting information on all HS functions and the system's context, so that an assessment of attribution or contribution can account for specific factors that may influence the overall results.
- **The necessary changes having been made (mutatis mutandis):** During an activity's implementation, stakeholders or donors make adjustments to fill obvious gaps, such as lack of staff or salary supplementation for additional work on the project. Other types of midcourse corrections can also be made. These changes can influence an activity's implementation, outcomes, and impact. Therefore, the evaluation should be designed to document changes that are not part of the original plan, to tease out their effects or to control for them somehow in the data analysis.
- **Attribution versus contribution:** HSS activities are not research projects. Because of their multilevel and multivariable nature, the causal linkages between intervention and impact may be unclear, multiple, or indirect. In addition, there may be a substantial time lag between when the activity is implemented and when anticipated HS outcome and impact occur; thus, assessing attribution¹¹ is challenging. Plus, it can be very difficult or costly to accurately pinpoint which of many factors are playing significant roles in the observed change. In some cases, the activity contributed to the changes observed but a causal relationship can't be established. Therefore, the evaluation planning team may think of evaluating the "contribution" of the HSS activity to HS outcomes and impact, as opposed to a study using experimental or quasi-experimental methods, which in some cases may not be feasible or practical.

¹¹ Some questions addressing attribution might be: Are the outcomes of interest attributable to the program? Are the outcomes of interest changing because of the program? Did the program cause the outcome of interest?

Types of Evaluations

Many types of evaluations can be conducted. USAID’s evaluation policy focuses on two: impact evaluations and performance evaluations.

Impact Evaluations

Impact evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to treatment and control groups provide the strongest evidence of a causal relationship between the intervention under study and the expected HSS outcomes measured (USAID, 2011a).

Under the OHS Vision for Health Systems Strengthening framework, the short term or intermediate effects on the beneficiaries are considered “outcomes” (access, service coverage, quality, equity, and household out-of-pocket health expenditures relative to total expenditures) while the long-term effect on beneficiaries is considered “impact” (health status and quality of life) (USAID, 2015a). Impact evaluations are recommended when we need to establish the causal link between an intervention and its impact or to test an innovative intervention that has not been tested. Examples of three HSS impact evaluations are provided at the end of this chapter to highlight the purpose, methods, and results of this type of evaluation.

Performance Evaluations

The second type of evaluation—performance evaluation—asks the following questions:

. . . descriptive and normative questions: what a particular project or program has achieved (either at an intermediate point in execution or at the conclusion of an implementation period); how it is being implemented; how it is perceived and valued; whether expected results are occurring; and other questions that are pertinent to program design, management and operational decision making. Performance evaluations often incorporate before-and-after comparisons, but generally lack a rigorously defined counterfactual (USAID, 2011a).

The purpose is to assess whether the project’s objectives and targets are achieved without ascertaining the causal link and controlling for other factors that may help explain changes in the development outcome. In other words, performance evaluations do not provide evidence of the counterfactual situation: What would have happened if the project had not been implemented?

Different types of performance evaluations can take place at different times throughout a project’s lifecycle. For example, a performance evaluation can take place during implementation (most likely at the project midterm) to understand the process of implementation and to identify any areas that need improvement. A performance evaluation can also be conducted at the end of a project to inform the development and design of the next project. For example, for the scenario described above, understanding whether the expected results were achieved in the areas of health financing, decentralization, quality assurance, and information systems can help identify the achievements of the initiatives and what type of USAID support is needed to strengthen the HS in these areas.

Planning for Evaluations

Evaluation questions and the types of evaluations (impact or performance) needed to answer these questions should be planned when the PAD is developed, because these have implications for the activity’s budgets, staffing, and implementation.

Ideally, the decision is made at the time of project design to conduct any evaluation—especially an impact evaluation, which requires careful, prospective planning, including having control groups and creating baselines for intervention and control groups. The RFPs for the activity and the impact evaluation are issued simultaneously. Contracts both for the evaluation and the activity are signed around the same time, so that the evaluators can work with mission staff and the IP to plan the evaluation. The research design for the impact evaluation is developed and intervention and control groups are selected based on the chosen design. A baseline survey that is administered prior to program implementation and at least one follow-up survey is conducted in both intervention and control groups to collect the data needed for the study.

While prospective evaluations that are developed when the activity is being designed are more likely to yield credible, rigorous evaluation results, in practice, USAID staff and activity managers often do not plan to conduct an impact evaluation at the beginning of an activity. Instead, the need for an impact evaluation may be determined midway or near the end of the activity.

This raises the question, is it possible to conduct an impact evaluation midway or at the end of the activity? Retrospective evaluations can be conducted for programs that were introduced in the past, but the research design options are limited. To generate estimates of impact, the research team must have data with sufficient coverage of the treatment and comparison groups both before and after the interventions were introduced. In this situation, the use of quasi-experimental methods may be possible (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011).

Evaluation SOW

The evaluation SOW is the first step toward planning the evaluation. It describes in detail what is expected from the evaluation team. Therefore, before you write it, we encourage you to talk with MEL advisors at the mission and headquarters levels or with an independent consultant to determine whether and what type of evaluation questions to include in the SOW to serve the evaluation objectives. Developing an evaluation SOW should be done carefully. A review of USAID SOWs conducted by Management Sciences International found that SOWs varied in the quality of the criteria they used. While some SOWs had exemplary sections, the reviewers were unable to identify one SOW that was exemplary throughout (Frumkin, Kearney, & Hageboeck, 2010). This means that more deliberation is needed in preparing SOWs, because they may influence the quality of the entire evaluation.

The evaluation SOW lays out the context, purpose, evaluation questions, and other important components, such as deliverables and budget, that evaluators need to determine in order to design a successful and high-quality evaluation. The SOW should be clear about how the issues of context and systemwide effects will be considered in the evaluation. USAID provides excellent resources on developing SOWs that ensure that these SOWs comply with ADS (links to this and another document with good practice examples from SOWs can be found at the end of this guidance).

Evaluation Questions

The formulation of the research questions is one of the most important steps of the evaluation process. In developing evaluation questions, the following issues should be considered:

- Limit the questions to the most important aspects of the activity you want to evaluate. It is not possible or feasible to evaluate all aspects of an activity.
- Using the activity's TOC as a guide, consider incorporating questions that investigate that systemwide effects of the selected activity strategy that is being evaluated.

- Include only questions that are realistic in terms of what an activity can accomplish and how much impact it can have during the study period. This is especially important for HSS impact evaluation questions. For some types of HSS initiatives, the HSS process is complex and often has lagged effects, so the impact on health or the health system may not be known until years after the activity has been implemented and evaluated.
- Include only two to three questions that can be feasibly studied given the available resources.

The evaluation questions determine the type of evaluation to be conducted. Is there a causal link between an activity’s intervention and observed changes in the beneficiaries? What is the magnitude of change resulting from the HSS activity? How much confidence do we have that observed change(s) are caused by the intervention? Is intervention X more effective and/or cost-effective than interventions Y and Z? Are the observed changes sustained after the activity ends? The questions that deal with determining causality can be answered only through impact evaluations, but other types of evaluation can be conducted. A performance evaluation can answer questions about whether the implementation occurred as planned or whether the project achieved its objectives. We assume that if the implementation occurred or the project achieved its target, it is plausible that the activity will contribute to improving health status. Illustrative questions relating to impact and performance evaluation are provided in Table 9.

Table 9. Examples of evaluation questions, by type of evaluation

Examples of evaluation questions	Type of Evaluation
<ul style="list-style-type: none"> • Did an intervention designed to integrate family planning and maternal services cause a change in contraceptive prevalence, the total fertility rate, maternal service coverage, and maternal mortality when areas with and without the intervention are compared? What are the positive and negative effects on other HS functions? • Did an intervention that aimed to improve domestic resource mobilization for HIV services cause changes in those services' outcomes (access, coverage, quality, equity, responsiveness, financial risk protection, etc.) in pilot districts? To what extent did it affect financing of other health services and HS functions? • Did a role modeling intervention on the use of information by senior management improve the quality and use of DHIS information and service coverage? • Did a task-sharing intervention among the health providers cause changes in the quality of care and service coverage? What systemic changes occurred because of task sharing? • To what extent did health decentralization at the district level improve service coverage and reduce child mortality compared to before decentralization? How did decentralization affect HS functions and outcomes? • Did an intervention designed to improve training on the rational use of medicines reduce household out-of-pocket health expenditure, adverse drug reactions, and mortality and increase quality of care and service coverage? • Did a strategy designed to improve coordination and collaboration between the public and private sectors at different levels improve progress toward PEPFAR's "90-90-90" goals¹²? 	Impact

¹² By 2020, 90 percent of those with HIV will have been diagnosed, 90 percent of those diagnosed will be in antiretroviral treatment, and 90 percent of those treated will be virally suppressed (PEPFAR, 2014).

<ul style="list-style-type: none"> • To what extent did the training of health facility staff in checking data quality achieve the target of 80 -percent data accuracy? Were any systemwide changes associated with project activity, such as changes in time allocation of health workers or improvements in service provision? • To what extent did the training of health facility staff in integrated management of childhood illnesses (IMCI) achieve its objective of 90% compliance with IMCI standards? What other systemwide changes were made to accomplish the objective? • To what extent was an intervention to improve the involvement of community members (an accountability strategy) associated with reducing staff absenteeism at the facility level? What other systemwide changes observed are owing to community involvement in facility governance, such as changes in service quality and use? • Is the community-based health insurance intervention being implemented as planned? Were the poor correctly identified? Were resources—human and in-kind—managed well? • Did supportive supervision improve immunization coverage from 40 percent to 70 percent? Were there any systemwide effects? • Was the target-setting policy for service coverage implemented at all districts and health facilities? What were positive and negative consequences of this policy at individual and system levels? 	Performance
---	--------------------

Under Scenario 5, evaluation questions could be as follows:

- How effective was the activity or selected strategic component of the activity in improving HS performance and health status?
- To what extent were the stakeholders, including the health ministry, IPs, other international partners, and health workers, were satisfied with the activity’s implementation and results?

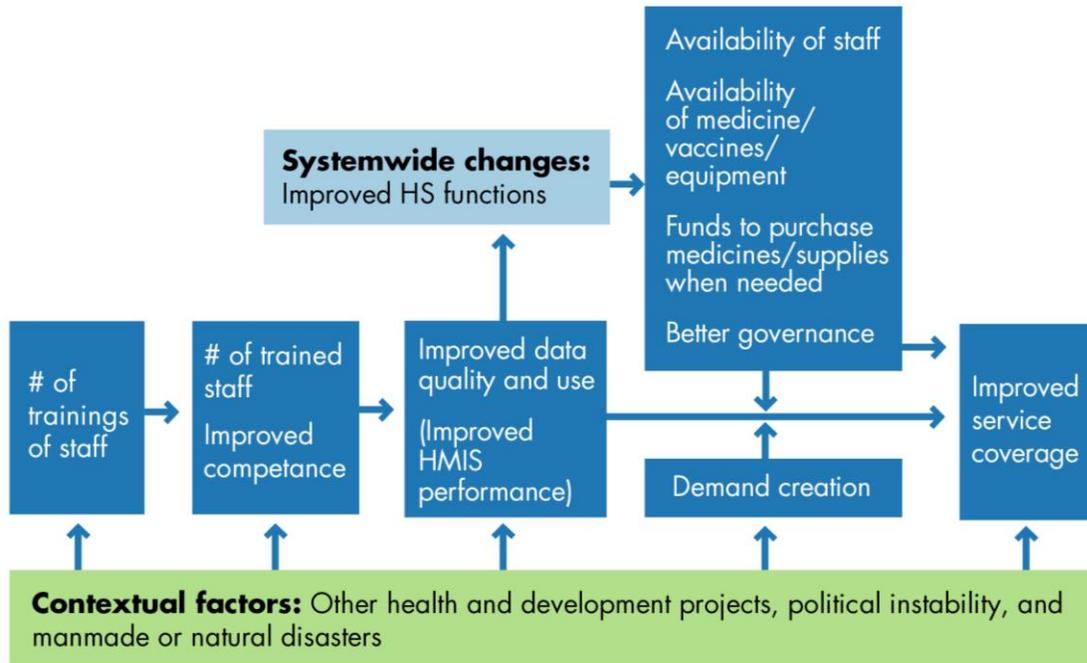
Conceptual Framework for Evaluation

USAID requires review of an activity’s TOC and RF to clarify (1) the causal pathways between the interventions supported by the activity and expected outcomes and impacts and (2) the underlying assumptions. The TOC and RF could be the basis for development of evaluation questions, but it is also recommended that the team responsible for the evaluation develop a conceptual framework, though ADS 201 does not require this. The rationale for developing a conceptual framework is four-fold. First, the activity’s TOC and RF often are not described in detail or updated based on the implementation context or modifications in the strategies. Second, the TOC and RF usually do not describe the contextual factors that may influence observed changes in the activity’s outcomes. Third, given the systemic nature of HSS, it is important to describe causal pathways between intervention and HS functions leading to HS outcomes and impact—in other words, the direct and indirect effects of the interventions on outcomes and impact. Fourth, the framework also aids decisions about research questions, the feasibility of evaluation design, what information to collect, and the methods of data collection.

The evaluation conceptual framework lists specific aspects of the activity that could be evaluated, after reviewing the TOC and RF and incorporating any modifications to the activity strategy. What are the pathways through which the intervention is supposed to create changes? What are the independent, mediating, and dependent variables, and how they are linked (or not linked) to one another to cause changes in dependent variables? Figure 17 is an example of an evaluation conceptual framework for an HMIS

strengthening activity whose main evaluation question is, “did an intervention designed to improve HMIS performance lead to improved data quality and use, and in turn, improved service coverage?”

Figure 17. Evaluation conceptual framework for HMIS strengthening



According to the framework, training staff will improve data quality and the use of HMIS data. This will then improve service coverage, through improved decision making. The framework also hypothesizes that an improved information system will produce systemic changes in other HS functions such as governance, financing, logistics management, and human resources because of the use of collected information. The sociodemographic characteristics of the services providers, demand creation activities, and contextual factors such as other health and development projects, political instability, and manmade or natural disasters could act as confounders and need to be controlled. By listing the variables and identifying their relationships, the evaluation framework facilitates selection of evaluation design, data collection instruments, and information sources.

Choosing an Evaluation Design

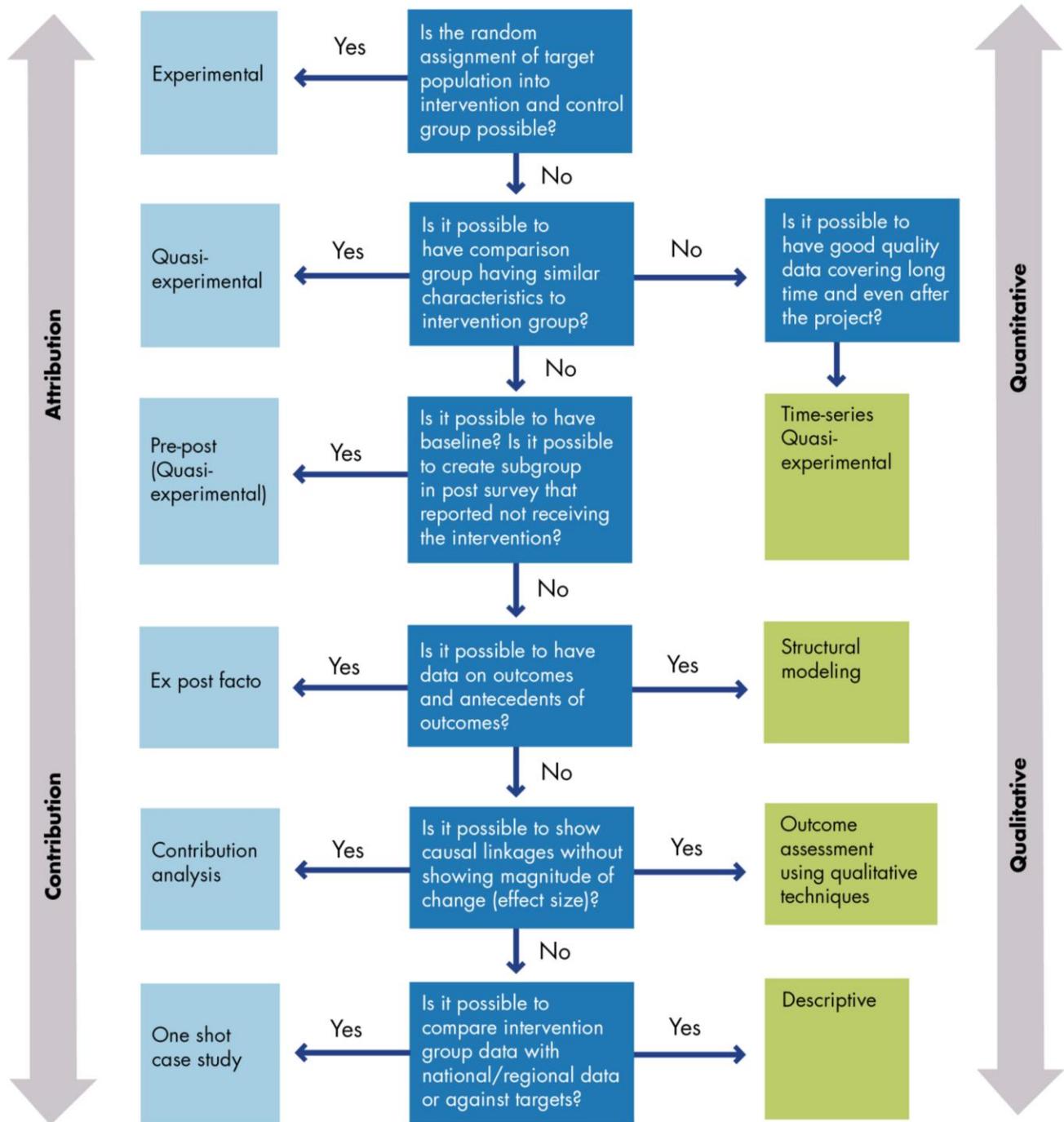
The choice of an evaluation design starts with the question, is it necessary to assess how much change in the intended outcomes (effect size) was caused by the intervention? If determining attribution is not possible, can the contribution of the activity or intervention be assessed in such a way that alternative explanations for observed changes are excluded? Even though quantitative experimental evaluation is the gold standard for assessing attribution and measuring the magnitude of change (effect size), qualitative methods—though lacking the ability to assess the magnitude of change—are useful for exploring causal linkages. Given that HSS intervention(s) affect multiple functions at different levels, it is often difficult and costly to tease out the direct and indirect effects of HSS interventions using experimental designs. Therefore, it is important to develop innovative evaluation design that addresses HS intervention complexity.

Two considerations should be at the forefront of a decision about the evaluation design. First, do the evaluation questions of interest lay on the continuum of attribution and contribution? Second, which design or method is appropriate to the evaluation question on the continuum of quantitative and qualitative methods? Figure 18 presents a decision tree for selecting an evaluation design. The evaluation design needs to meet certain conditions either to show observed changes completely attributable to the intervention (experimental) or to describe changes (descriptive) for which alternative explanations cannot be excluded. Randomization of the target population into intervention and control groups is the best solution for eliminating alternative causes. If randomization is not possible, then establishing comparison group whose characteristics largely align with those of the intervention group provides a safeguard against alternative explanation. Other resources can be consulted to help identify the appropriate quantitative evaluation design (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011; <http://usaidprojectstarter.org>).

Causal pathways and linkages can be explored with good quality contextual information and data on the processes, outputs, and performances connected to outcomes and impact, based on the TOC and using structural modeling or comparing data on the group receiving the intervention with another group that did not receive the intervention (ex post facto). Similarly, linking outcomes with an intervention could be explored with qualitative techniques (outcome assessment) or by combining qualitative and quantitative information (contribution analysis). Performance evaluation by design falls under the contribution side of the contribution–attribution continuum.

Given the complexity of HSS intervention and because the field of HSS evaluation is relatively nascent, we recommend that evaluation design should be more flexible and exploratory to describe causal pathways and linkages than to assess causality and magnitude of change (effect size). The plausibility, contribution analysis, and outcome linkage assessment are more important at this stage of the evaluation’s design for complex systems. We also recommend involvement by M&E specialists in the mission and headquarters to make the best decision on the evaluation design after conditions and costs are considered.

Figure 18. Decision tree for selecting HSS evaluation designs



Data Collection Methods

The SOW can link the data collection methods with the questions they aim to answer. This can be done using the tool “Getting to the answers matrix” (see Table 10). By providing this matrix in the SOW with broad answers will help the evaluation team come up with specific data collections methods, sources of information, or sampling or selection criteria.

Table 10. Getting to the answers matrix

Evaluation Questions	Type of Answer Needed	Data Collection Method(s)	Data Source(s)	Sampling or Selection Criteria	Data Analysis Method(s)
1.	<input type="checkbox"/> Descriptive <input type="checkbox"/> Comparative (normative) <input type="checkbox"/> Cause-and-Effect				
2.	<input type="checkbox"/> Descriptive <input type="checkbox"/> Comparative (normative) <input type="checkbox"/> Cause-and-Effect				

Source: USAID, 2011b

Data Analysis Plan

Another part of the SOW will be requesting the data analysis plan from the evaluation team, based on the matrix provided. This plan can describe how quantitative and qualitative data will be managed and analyzed and findings will be presented.

Deliverables

Deliverables such as the evaluation protocol—a description of the research methods, data collection and analysis activities, work plan, consultation and dissemination activities with partners, and the draft and final evaluation reports—will also be outlined in the SOW.

Team Composition

The SOW must specify the evaluation team’s composition. The evaluation team should be identified such that it has the appropriate methodological and subject-matter expertise and demonstrated experience in conducting high-quality evaluations. The SOW can specify the number and types of team members and background of the team. The team leader should always be an external expert, who has no affiliation with the project. You can request that the evaluation team members provide samples of their past evaluation reports as part of the application to assess the quality of their work. It is also possible to add a requirement that stakeholders need to be involved in the evaluation.

Bidding for Evaluations

Once the evaluation SOW is ready, you must decide which organization or individuals should conduct the evaluation. You could use a USAID mechanism (such as the Bureau of Global Health, Bureau of Policy, Planning and Learning [PPL], the Global Health Program Cycle Improvement Project [GH Pro], MEASURE Evaluation, the U.S. Global Development Lab (the Lab)/PPL/GH Monitoring, Evaluation, Research and Learning Innovations Program (MERLIN), or the Office of Learning, Evaluation and Research [LER] Master M&E IDIQ) to identify the organization that will conduct the evaluation or use external evaluation services. If you decide to use a USAID mechanism, share the SOW with project AORs so that they can advise on next steps, because there might be some considerations you may want to use for making a final decision in selecting a mechanism. If you will use external evaluation services, consult USAID's Office of Acquisitions to plan the bidding process.

Before sending out the SOW to evaluation mechanisms or developing the A&A document, please ensure the following:

- The purpose of the evaluation is clearly stated. The evaluation questions are described in such a way that the type of evaluation is apparent.
- A conceptual framework for the evaluation describing independent and dependent variables and their causal relationship is recommended but not required under ADS 201.
- A request for the evaluation design and its rationale is included.
- A request for context and systemwide effects in the evaluation's design is included.
- A request for a description of qualitative and quantitative methods of data collection and their rationale is included.
- The SOW states that after the award, if the method is to be an impact evaluation, the winner will submit a protocol for it.

Evaluation Design and Work Plan

Once the evaluation award has been made, an evaluation protocol should be prepared that includes a description of the key research questions, methods, sources of data, data instruments, data analysis plan, and dissemination plan. The study should be based on a systematic methodological approach that relies on facts, evidence, and data. Studies that collect primary data from human subjects should undergo ethical review from an institutional review board in the country where the research will be carried out, to ensure that the study meets appropriate ethical standards.

To review the protocol, it is important to check that the following aspects of the evaluation are described:

- Rationale for the evaluation
- Evaluation research questions
- The evaluation design and its rationale and alignment with the evaluation questions (This should include a rationale for the research design chosen and any qualitative methods that are incorporated in the design.)
- The methods section, including:
 - Randomization procedures, if needed
 - The sample size and its calculation method explained and justified
 - The sampling method
 - All the data collection instruments, translated, if necessary
- The plan to train data collectors and collect survey data

- Mechanisms to track other data, such as recruitment and training of data collectors, targets achieved in their timeline, and demographic information about participants
- The data management processes, including how confidentiality of respondents and data are assured, and processes to make the data publicly available
- The data analysis plan, including an explanation of how dropout of respondents is accounted for in the analysis
- Mechanisms and processes to meet the ethical standards
 - Ethical review by an institutional review board in the country
 - Attachment of informed consent processes (including consent forms) meeting international standards and translated, if necessary
- Final and interim reports, which may be needed if the evaluation has multiple data collection rounds
- The dissemination plan

Action: Reach out to the evaluator if any of the information is missing or is unclear.

Implementing the Evaluation

The length of the evaluation will vary with the evaluation’s type. If a prospective approach is taken, impact and performance evaluations can last beyond the life of the activity (four or five years). Retrospective evaluations are conducted in much shorter periods. If the evaluation is expected to be of long duration, you should request a progress report every six months to stay abreast of developments and challenges encountered.

Managing the Evaluation

To assure that evaluations are being carried out as planned and to monitor the progress of the evaluation process, it may be helpful to conduct routine meetings between the evaluation team, the implementation team, and USAID. These meetings can help ensure that the activities of the evaluation and implementation teams are properly coordinated; that research protocols and reports are reviewed and adhere to evaluation research standards; that data collection is carried out as planned, and that data entry and analysis are on schedule; and that plans are in place to effectively disseminate the evaluation findings.

Analysis, Report Preparation, Utilization, and Dissemination

Analysis

After data have been obtained or collected, they should be cleaned and analyzed. Sound data management practices ensure the integrity of the analysis. Therefore, monitor the following:

- Processes are in place to protect individually-identifiable data and to report any breach of confidentiality.
- Data cleaning processes are in place and described in the progress report or documentation is made available, when requested.
- Data are analyzed according to the analysis plan, which provides information on descriptive, disaggregation, and inferential analyses. Documentation is available on the formulas, variables used, and new variables constructed as well as the analysis programming file.

Action: Ask the evaluation partner to provide all missing information under analysis.

Report Preparation

Once the data analysis has been completed, an evaluation report is prepared. USAID provides guidelines, a template, and an evaluation review checklist to make sure the report meets quality standards (USAID, 2017b).

Action: Assure that the evaluation report is not made public before there is agreement among all stakeholders about the evaluation findings, conclusions, and recommendations. If there is disagreement, a statement of differences to accompany the report can be submitted by any of the parties that don't agree with the findings.

Utilization

Learning from MEL to improve performance is a key principle of the USAID evaluation policy. Mission and Washington OUs must develop a post-evaluation action plan after the acceptance of the evaluation report (USAID, 2017b).

The following steps are required as per ADS:

- Review the key findings, conclusions, and recommendations systematically.
- Determine and document whether the mission or Washington OU accepts/supports each conclusion or recommendation.
- Identify any management or other program actions needed based on the evaluation findings, conclusions, or recommendations. This may include changes to strategy, projects, activities, or other planning frameworks.
- Assign responsibility and the time frame for completion of each set of actions.
- Document the expected actions based on the evaluation, responsibilities, time frames, and completion of actions in a post evaluation action plan.

Dissemination

Evaluation results are used to promote accountability and learning, which can help in decisions to determine the future funding, design, and expansion of activities. Thus, dissemination plans must ensure that the evaluation contributes to the broader goal of learning from experience. In collaboration with stakeholders, USAID mission staff should work with the evaluation team to determine the most appropriate format in which to report the evaluation findings, as well as the timing and audience of the report. The format can be a technical report, research paper, or presentation. Results may be written up in a formal manuscript and submitted to a peer-reviewed journal. This is most common when the study finds new and innovative results, which would be of interest to a wide audience.

The audience for evaluation reports varies with the project. In some cases, evaluation results may be made public; in others, they will be kept for internal purposes. USAID policy encourages the sharing of evaluation findings, and evaluation reports must be submitted to the Development Experience Clearinghouse (DEC)—<https://dec.usaid.gov/dec/home/Default.aspx>—within three months of the conclusion of the evaluation. This makes the evaluations available to USAID staff as well as others and it is an important part of building the evidence for what works.

The protocol provides the dissemination plan for sharing evaluation results. Therefore, monitor the activities in the following ways:

- Seminars for sharing the evaluation findings in the country and at USAID headquarters
- Dissemination of major findings through professional networks, listservs, Twitter, Facebook, websites, etc.

- Distribution of the evaluation report to all stakeholders within and outside the country
- Submission of the evaluation report to the Development Experience Clearinghouse (DEC) within three months of the conclusion of the evaluation
- Submission of the data to Data Development Library, (<https://www.usaid.gov/data>) after deleting the participants' identification information, as per USAID's open data policy, "within thirty (30) calendar days after the Dataset is first used to produce an Intellectual Work or is of sufficient quality to produce an Intellectual Work."
- Reporting internally to senior management on the use of evaluation report – proxy indicator for use. The number of times the evaluation report is downloaded and disaggregated by countries and missions.

Key Messages

- Start planning an evaluation—especially an impact evaluation—in the project design phase.
- Ensure that conditions are right to allow an impact evaluation of an HSS intervention.
- Prepare a conceptual framework for evaluation based on the TOC and tied to evaluation questions.
- Confirm that documenting systemwide effects and identifying causal pathways are part of the evaluation.
- Select the evaluation type and design by considering an attribution–contribution continuum.
- Request and review an impact evaluation protocol for approval before implementing the project.
- Make sure that a communication plan for the evaluation findings is available.
- Ensure that the evaluation's scope of work (SOW) states how evaluation findings will be used.
- Involve MEL advisors from USAID's mission and headquarters levels in the design phase of the evaluation.

Summaries of Illustrative Impact Evaluations

Example 1. Paying Primary Healthcare Centers for Performance in Rwanda

Background: Rwanda instituted a pay-for-performance scheme to improve the use and quality of care in primary health centers. The program provided incentives in the form of bonuses to providers for improvements in these indicators. The aim was to improve service delivery and health outcomes by incentivizing providers to put more effort into specific activities and by increasing the resources available to finance the delivery of services. The scheme paid for 14 maternal and child healthcare output indicators and the incentives were given to the facilities to use at their discretion. Comparison facilities received an increase in their budgets equal to the average pay-for-performance payments to treatment facilities to isolate the resource effect from the incentive effect. This study evaluated the pay-for-performance scheme in Rwanda between 2006 and 2008.

Study design: This study used a prospective quasi-experimental evaluation design nested in the program rollout. The program was rolled out over a 24-month period at the district level. Treatment facilities were enrolled in the pay-for-performance scheme in 2006 and comparison facilities were enrolled two years later. Districts were grouped in eight pairs based on similar characteristics for rainfall, population density, and predominant livelihoods using data from the 2002 census. One side of each pair was then randomly assigned to the treatment group and the other to the comparison group.

Sample: A total of 166 primary care facilities were included as well as 2,158 households (13 households per facility catchment area were sampled).

Methods: They conducted a baseline survey and a follow-up survey 24 months later at each facility and in selected households. The main outcome measures were maternal health services, quality of prenatal care, and child preventive care. Multivariate regression specification of the difference-in-difference model was estimated, controlling for covariates.

Results: Pay-for-performance had a larger impact on services with higher incentives and for services that are more in the control of the provider and depend less on patients' decisions. Pay-for-performance had a large and significant positive impact on institutional deliveries (which had the highest payment rate) and preventive care visits by young children, and improved the quality of prenatal care (which was under the provider's control). No effect was observed on the number of prenatal care visits (a provider can encourage but it is ultimately the woman's decision to come in) or on immunization rates.

Conclusion: The authors concluded that financial performance incentives can improve both the use and the quality of health services.

Source:

Basinga, P., et al. (2010). Paying primary healthcare centers for performance in Rwanda. Policy Research Working Paper 5190. Washington, DC, USA: World Bank. Retrieved from https://www.rbfhealth.org/sites/rbf/files/Rwanda_P4P.pdf

Example 2. Contracting Primary Healthcare Services to Improve and Increase Coverage in Cambodia

Background: In 1998, the Cambodian government piloted the contracting of health service delivery to nongovernmental entities to explore alternative ways to improve service coverage and quality. Two ways of contracting were tested: (1) contracting-out the management of all aspects (staffing, equipment, and drugs procurement, etc.) of service provision was given to the nongovernmental entity, and (2) contracting-in, such that the nongovernmental entity was responsible only for the management of the civil service health staff. The other components of service delivery, such as recurrent costs, were handled through the regular government channels. Additionally, in this pilot, drugs and other amenities needed for the delivery of good-quality services were also adequately provided and provider salaries were adjusted to align their incentives with high-quality service delivery. Selected health indicators were coverage levels for antenatal care, assisted deliveries, full immunization, vitamin A, knowledge of modern birth-spacing, use of modern contraceptive methods, and public health facilities. Facilities received payments for meeting these targets and bonuses for indicators for which targets were exceeded.

Study design: The study used a pre- and post-intervention with controls (quasi-experimental) research design to evaluate the effect of contracting on service delivery. There were three groups in the study: (1) contracted-out districts (2) contracted-in districts, and (3) control districts. The contracted in and control districts received supplemental budgets.

Sample: Rural districts with about 100,000 to 200,000 people were randomly selected in one of the three study groups. There were two contracted-out districts, three contracted-in districts, and four control districts. All the districts in the study had comparable socioeconomic status and very poor service coverage, with no more than 20 percent of their planned health facilities operational.

Methods: Baseline household and health facility surveys were conducted before the pilot began and a follow-up study was conducted two and half years after the contracts began.

Results: Overall, service coverage improved in all districts. Service provision in contracted-out districts improved the most, followed by contracted-in districts and control districts. For instance, health service use among the poorest people in contracted-out districts improved by almost 30 percent. Additionally, out-of-pocket payments fell by about 70 percent (\$35 a year) among the poor in contracted-out districts. However, there was no change in vitamin A coverage and only a small change was observed for assisted deliveries. The authors concluded that although contracting to nongovernmental entities might be challenging for policymakers to accept, it can be equitable and effective.

Sources:

World Bank. (2003). World Development Report 2004: Making services work for poor people. Washington, DC, USA: World Bank. Retrieved from <https://openknowledge.worldbank.org/handle/10986/5986> License: CC BY 3.0 IGO."

Bhushan, I., Keller, S., & Schwartz, J.B. (2002). Achieving the twin objectives of efficiency and equity: Contracting health services in Cambodia. ERD Policy Brief 6, Manila, Philippines: Economics and Research Department, Asian Development Bank. Retrieved from <http://www.hrresourcecenter.org/node/4510>

Example 3. Using Public Policy to Improve the Use of Health Services by the Poor

Background: In 2003, Mexico introduced a set of health reforms aimed at improving insurance coverage among its populace, especially the poor. These health reforms, titled *Seguro Popular*, consisted of increased funding for the healthcare sector, improved federal supervision, added administrative responsibility for the states, and health insurance coverage and information resources for individuals. *Seguro Popular* would be an incentive to strengthen the accreditation system for health facilities, by aligning federal support to the quality of the health facility. It entitled affiliated families to a benefits package that covered 266 health interventions, 312 medications, and support for catastrophic medical expenditures related to certain diseases. The phased implementation of the program included an experimental evaluation of the reforms that were being instituted.

Study design: The study used a matched-pair cluster-randomized experiment to assess the impact of *Seguro Popular*. The phased rollout of the program enabled some health clusters to be randomized for the study's purposes. Randomization occurred at the cluster level. Each health cluster included a health facility and the population in the catchment area. One health cluster in each pair was assigned to be the treatment cluster. In the treatment cluster, a campaign was implemented to encourage all the families to enroll in *Seguro Popular*, and procedures were initiated to increase medical staff and drug availability in order to enhance service delivery in health facilities. Nothing was changed in the paired health cluster that was the comparison cluster.

Sample: The sample consisted of 36,181 randomly chosen households from 100 clusters in the survey.

Methods: The study conducted two household surveys: a baseline household survey and a follow-up household survey 10 months later. The key outcomes were household catastrophic expenditures and household out-of-pocket expenditures for inpatient care, outpatient care, medical devices, and medication. An intent-to-treat analysis aimed at estimating the impact based on the program assignment of households irrespective of ultimate compliance, a complier average causal effect analysis that measured the program's impact on compliers only, and a difference-in-difference analysis to confirm the robustness of the results.

Results: Participation in the *Seguro Popular* program reduced by 23 percent the proportion of respondents who experienced catastrophic expenditure. Among those who complied with the assignment in the program, catastrophic expenditures were reduced by 55 percent. Poor households experienced most of the reduction in catastrophic expenditures. Participants also had reduced out-of-pocket expenditures for inpatient and outpatient services. There was no positive effect on expenditure on drugs and devices detected. The program also did not increase the use of health services among participants, as had been previously observed elsewhere.

Conclusion: The *Seguro Popular* program was effective in reducing overall out-of-pocket payment and catastrophic healthcare payments for inpatient and outpatient medical care, especially among the poor.

Source:

King, G., Gakidou, E., Imai, K., Lakin, J., Moore, R. T., Nall, C., ... Llamas, H. H. (2009). Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. *The Lancet*, 373(9673), 1447–1454. Retrieved from [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(09\)60239-7/abstract](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(09)60239-7/abstract)

REFERENCES

- Aqil, A., Lippeveld, T., & Hozumi, D. (2009). PRISM framework: A paradigm shift for designing, strengthening and evaluating routine health information systems. *Health Policy and Planning*, 24 (3), 217–228. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2670976/>
- Aqil, A., Ekanem, E., & Ettinger, M. (2016). *Health systems benchmarking tool*. Washington, DC, USA: United States Agency for International Development. Retrieved from <https://idea.usaid.gov/global-health-tools.htm>
- Atun, R., de Jongh, T., Secci, F., Ohiri, K., & Adeyi, O. (2009). Integration of targeted health interventions into health systems: A conceptual framework for analysis. *Health Policy and Planning*, 25, 104–111. Retrieved from <https://academic.oup.com/heapol/article/25/2/104/641536/Integration-of-targeted-health-interventions-into>
- de Savigny, D., & Adam, T., Eds. (2009). *Systems thinking for health systems strengthening*. Geneva, Switzerland: World Health Organization. Retrieved from <http://www.who.int/alliance-hpsr/resources/9789241563895/en/>
- Diana, M.L., Yeager, Y.A., & Hotchkiss, D.R. (2017a). *Health systems strengthening: A compendium of indicators*. Chapel Hill, NC, USA: MEASURE Evaluation. Retrieved from <https://www.measureevaluation.org/resources/publications/tr-17-167b>.
- Diana, M.L., Yeager, Y.A., & Hotchkiss, D.R. (2017b). *Health systems strengthening: A literature review*. Chapel Hill, NC, USA: MEASURE Evaluation. Retrieved from <https://www.measureevaluation.org/resources/publications/tr-17-167a>.
- Evans, R. (1981).: Incomplete vertical integration: The distinctive structure of the health-care industry. In Van der Graag, J.J., *Health, economics, and health economics*.
- Executive Office of the President, Office of Management and Budget. (2013). *M-13-13, Open data policy—Managing information as an asset*. Washington, DC, USA: Office of Management and Budget. Retrieved from <https://www.digitalgov.gov/open-data-policy-m-13-13/>.
- Frumkin, M., Kearney, E., & Hageboeck, M. (2010). *Quality review of recent evaluation statements of work (SOWs)*. Washington, DC, USA: United States Agency for International Development. Retrieved from [http://usaidprojectstarter.org/sites/default/files/resources/pdfs/quality review of USAID SOWs.pdf](http://usaidprojectstarter.org/sites/default/files/resources/pdfs/quality%20review%20of%20USAID%20SOWs.pdf).
- Gertler, P., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact evaluation in practice*. Washington, DC, USA: World Bank. Retrieved from http://siteresources.worldbank.org/EXT/HDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf.
- Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168, 267–306. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2004.00349.x/abstract>.
- Gruskin, S., Ahmed, S., Bogecho, D., Ferguson, L., Hanefeld, J., MacCarthy, S., Raad, Z., & Steiner, R. (2012). Human rights in health systems frameworks: What is there, what is missing and why does it matter? *Global Public Health*. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/22263700>

- Hafner, T., Walkowiak, H., Lee, D., & Aboagye-Nyame, F. (2017). Defining pharmaceutical systems strengthening: Concepts to enable measurement. *Health Policy and Planning*, 32, 572–584. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28025324>
- Hatt, L., Johns, B., Connor, C., Meline, M., Kukla, M., & Moat, K. (2015). *Impact of health systems strengthening on health*. Bethesda, MD, USA: Health Finance & Governance Project, Abt Associates. Retrieved from <https://www.hfgproject.org/impact-hss-health/>.
- Health Metrics Network. (2008). *Framework and standards for country health information systems*, 2nd edition. Geneva, Switzerland: Health Metrics Network, World Health Organization. Retrieved from <https://www.measureevaluation.org/his-strengthening-resource-center/resources/hmn-framework-and-standards-for-country-health-information-systems>.
- Hotchkiss, D., Stillman, K., Hutchinson, P., & Connor, C. (2006). *Improving health system performance through monitoring and evaluation*. Bethesda, MD, USA: Partners for Health Reform Plus Project, Abt Associates, Inc.
- Kleczkowski, B., Roemer, M., & Van Der Werff, A. (1984). *National health systems and their reorientation towards health for all. Guidance for policy-making*. 77th edition. Geneva, Switzerland: World Health Organization.
- Knowles, J. C., Leighton, C., & Stinson, W. (1997). *Measuring results of health sector reform for system performance: A handbook of indicators*. Bethesda, MD, USA: Partnerships for Health Reform, Abt Associates Inc. Retrieved from https://www.measureevaluation.org/prh/rh_indicators/crosscutting/mgt/other-resources/measuring-results-of-health-sector-reform-for.
- Kruk, M., & Freedman, L. (2008). Assessing health system performance in developing countries: A review of the literature. *Health Policy*, 85(3), 263–276. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17931736>
- Mikkelsen-Lopez, I., Wyss, K., & Des, D. (2011). An approach to addressing governance from a health system framework perspective. *BMC International Health and Human Rights*, 11, 13. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/22136318>
- Mossialos, E., Allin, S., & Figueras, J. (2007). *Health systems in transition: Template for analysis*. Copenhagen. World Health Organization Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.
- Murray, C., & Frenk, J. (2000). A framework for assessing the performance of health systems. *Bulletin of the World Health Organization*, 78(6), 717–731. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2560787/>.
- Patton, M.Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY, USA: The Guildford Press, 94.
- Roberts, M., Hsiao, W., Berman, P., & Reich, M. (2004). *Getting health reform right. A guide to improving performance and equity*. New York, NY, USA: Oxford University Press.
- Roemer, M. (1993). National health systems throughout the world. *Annual Review of Public Health*, 14, 335–353. Retrieved from <http://www.annualreviews.org/doi/abs/10.1146/annurev.pu.14.050193.002003>

- Shakarishvili, G., Lansang, M., Mitta, V., Bornemisza, O., Blakley, M., Kley, N., Burgess, C., & Atun, R. (2011). Health systems strengthening: A common classification and framework for investment analysis. *Health Policy and Planning*, 26 (4), 316–326. Retrieved from <https://academic.oup.com/heapol/article/26/4/316/659551/Health-systems-strengthening-a-common>
- Shakarishvili, G., Atun, R., Berman, P., Hsiao, W., Burgess, C., Lansang, M. (2010). Converging health systems frameworks: Towards a concepts-to-actions roadmap for health systems strengthening in low and middle income countries. *Global Health Governance*, III, 2. Retrieved from http://www.ghgi.org/Shakarishvili_Converging%20Health%20Systems%20Frameworks.pdf
- Snowden, D.J., & Boone, M.E. (2007). A leader's framework for decisionmaking. *Harvard Business Review*, 69–76. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18159787>
- United States Agency for International Development (USAID). (2009). *Glossary of evaluation terms. Planning and performance management*. Washington, DC, USA: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/Pnado820.pdf.
- United States Agency for International Development (USAID). (2011a). *Evaluation: Learning from experience. USAID evaluation policy*. Washington, DC, USA: USAID. Retrieved from <https://www.usaid.gov/sites/default/files/documents/1870/USAIDEvaluationPolicy.pdf>.
- United States Agency for International Development (USAID). (2011b). *Evaluation statements of work: Good practice examples*. Washington, DC, USA: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/Pnadw976.pdf.
- United States Agency for International Development (USAID). (2013). *Technical note: Impact evaluations*. Washington, DC, USA: USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1870/IE_Technical_Note_2013_0903_Final.pdf.
- United States Agency for International Development (USAID). (2014a). *Acting on the call: Ending preventable child and maternal deaths*. Washington, DC, USA: USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1864/USAID_ActingOnTheCall_2014.pdf.
- United States Agency for International Development (USAID). (2014b). *Local systems: A framework for supporting sustained development*. Washington, DC, USA: USAID. Retrieved from <https://www.usaid.gov/policy/local-systems-framework>.
- United States Agency for International Development (USAID). (2015a). *USAID's vision for health systems strengthening*. Washington, DC, USA: USAID. Retrieved from <https://www.usaid.gov/what-we-do/global-health/health-systems/usaid-vision-health-systems-strengthening>.
- United States Agency for International Development (USAID). (2015b). *ADS chapter 200: Development policy*. Washington, DC, USA: USAID. Retrieved from <https://www.usaid.gov/sites/default/files/documents/1870/200.pdf>.
- United States Agency for International Development (USAID). (2016). *Impact of health systems strengthening on health systems performance and outcomes. Marshalling the evidence: A status report*. Washington, DC, USA: USAID, Bureau for Global Health, Office of Health Systems. Retrieved from <https://www.usaid.gov/what-we-do/global-health/health-systems/resources>.

United States Agency for International Development (USAID). (2017a). *ADS chapter 303: Grants and cooperative agreements to non-governmental organizations*. Washington, DC, USA: USAID, Bureau for Global Health, Office of Health Systems. Retrieved from <https://www.usaid.gov/ads/policy/300/303>.

United States Agency for International Development (USAID). (2017b). *ADS chapter 201: Program cycle operational policy*. Washington, DC, USA: USAID. Retrieved from <https://www.usaid.gov/ads/policy/200/201>.

United States Agency for International Development (USAID). (2017c). *How-to note: Prepare and maintain a Performance Management Plan (PMP)*. Washington, DC, USA: USAID. Retrieved from <https://usaidealmlab.org/library/how-note-prepare-and-maintain-performance-management-plan-pmp>.

United States Agency for International Development (USAID), Bureau for Policy Planning and Learning. (2012). *How to note: Preparing evaluation reports* (Issue brief No. 1). Washington, DC, USA: USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1870/How-to-Note_Preparing-Evaluation-Reports.pdf.

United States Agency for International Development (USAID), Health Finance and Governance Project. (2015a). *The health system assessment approach: A how-to manual*. Washington, DC, USA: USAID/HFG. Retrieved from <https://www.hfgproject.org/the-health-system-assessment-approach-a-how-to-manual/>.

United States Agency for International Development (USAID), Health Finance and Governance Project. (2015b). *Health systems benchmarking tool (HSBT)*. Washington, DC, USA: USAID/HFG. Retrieved from <https://www.hfgproject.org/usaid-health-systems-benchmarking-tool/>

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2012). *PEPFAR blueprint: Creating an AIDS-free generation*. Washington, DC, USA: PEPFAR. Retrieved from <https://www.pepfar.gov/documents/organization/201386.pdf>.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2014). *PEPFAR 3.0. Controlling the epidemic: Delivering on the promise of an AIDS-free generation*. Washington, DC, USA: PEPFAR. Retrieved from <https://www.pepfar.gov/documents/organization/234744.pdf>.

van Olmen, J., Criel, B., Bhojani, U., Marchal, B., Chenge, F., Van Damme, W., Hoeree, T., Pirard, M., Kegels, G. (2012). The health systems dynamics framework. *Health, Culture and Society*, 2(1), 1–12.

World Health Organization (WHO). (2006). *Constitution of the World Health Organization*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/governance/eb/who_constitution_en.pdf.

World Health Organization (WHO). (2007). *Everybody's business: Strengthening health systems to improve health outcomes: WHO's framework for action*. Geneva, Switzerland: WHO. Retrieved from <http://apps.who.int/iris/handle/10665/43918>.

World Health Organization (WHO). (2015). *Global reference list of 100 core indicators*. Geneva, Switzerland: WHO. Retrieved from <http://www.who.int/healthinfo/indicators/2015/en/>



MEASURE Evaluation

University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517 USA
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement AID-OAA-L-14-00004 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. TR-17-167c
ISBN: 978-1-9433-6455-8

