Getting to an Evaluation Plan:
A Six-Step Process from Engagement to Evidence

A Workbook

November 2017

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Getting to an Evaluation Plan:
A Six-Step Process from Engagement to Evidence

Brittany S. Iskarpatyoti, MPH, Beth Sutherland, PhD, and **Heidi W. Reynolds, MPH, PhD**

A Workbook

November 2017

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE
Evaluation

CONTENTS

ACKNOWLEDGMENTS

ABBREVIATIONS

DHS	Demographic and Health Survey
FSW	female sex worker
GAC	Ghana AIDS Commission
HIA	HIV Impact Assessment
IBBSS	Integrated Biological and Behavioral Surveillance Survey
KAP	knowledge, attitude, and practice
KP	key population
M&E	monitoring and evaluation
MARP	most at-risk population
MSM	men who have sex with men
NSF	National Strategic Framework
NSP	National Strategic Plan
PEPFAR	United States President's Emergency Plan for AIDS Relief
PIP	program impact pathway
PHQA	Public Health Questions Approach
RCT	randomized control trial
SDG	Sustainable Development Goal
TWG	technical working group
UNAIDS	Joint United Nations Programme on HIV/AIDS
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

As population and program-level health data become more ubiquitous, stakeholders should be intentional in sharing and aligning information resources to answer larger evaluation questions. The collection, analysis, and use of evaluation data, especially when the data are owned by different stakeholders, should be intentionally planned. An evaluation plan defines and organizes program activities and connects them to outputs, outcomes, and impacts; identifies existing and planned data sources; prioritizes evaluation research questions; and determines the roles, responsibilities, and timelines for answering the research questions. The process of developing an evaluation plan, in cooperation with a group of stakeholders, fosters collaboration, a sense of shared purpose, and transparency, thereby ensuring that stakeholders are on the same page about the purpose, use, and users of evaluation findings. A written plan that is developed, agreed upon, and adopted by stakeholders is one of the most effective tools for ensuring that time and resources are not wasted, and that information is available to answer a program's most important questions.

The purpose of this workbook is to provide practical advice and activities to facilitate the preparation of a written evaluation plan that is in line with best evaluation planning practices, as outlined by the Joint United Nations Programme on HIV/AIDS (UNAIDS) HIV Monitoring and Evaluation Reference Group, United States President's Emergency Plan for AIDS Relief (PEPFAR), and the United States Agency for International Development (USAID) (USAID, 2011; UNAIDS, 2010b; PEPFAR, 2015). This workbook describes a six-step process for developing a written evaluation plan: (1) engage stakeholders; (2) know your program; (3) know your evaluation needs; (4) select the evaluation design; (5) draft the evaluation plan; and (6) ensure use. Extensive field testing of this process assures that that users will successfully produce a complete evaluation plan that is wholly owned by the stakeholders who participate in the planning process.

Although this workbook was developed in the context of evaluation planning for HIV programs, many of the stakeholders involved in the pilot and field applications were from other sectors, including the military, police, education, prisons, and social work; donor and implementing agencies; and civil society and advocacy groups. The variety of users to date suggests that this process can be applied successfully in other health areas and sectors.

ABOUT THIS WORKBOOK

Purpose of This Workbook

The purpose of this workbook is to provide practical advice and activities for the preparation of a written evaluation plan. An evaluation plan clarifies the actions needed to assess a program by linking program activities with intended outcomes, and defines when and how those actions will be carried out (United States Centers for Disease Control and Prevention [CDC], 2011; Better Evaluation, n.d.).

Each chapter of this workbook is composed of two parts: (1) a description of one of the six steps for the development of a program evaluation plan; and (2) companion worksheets to facilitate the process and organize information. Sample documents and additional resources are provided in the appendices. By following the six steps and using the companion worksheets, programs can successfully work with stakeholders to develop a written evaluation plan.

When to Use This Workbook

This workbook can assist in contexts where (1) national/regional monitoring and evaluation (M&E) plans, strategies, and operational plans exist, but no specific evaluation plan has been developed; or (2) programs have an M&E plan, but would like to evaluate the effectiveness of the program; or (3) where there is a need to design an evaluation to answer priority research questions. The six-step process can help identify evaluation priorities in the context of limited resources and where multiple stakeholders have interest in using evaluation results. By preparing an evaluation plan, your program or organization can decide what information you and your stakeholders really need, leverage and strengthen existing investments in data collection and use, and keep you from wasting time gathering information that is not needed.

Although evaluations are often conducted at the end of programs, they should be planned at the start because they rely on data collected throughout the program implementation period, with the collection of baseline data being especially important. This workbook can be used before or while a program is being implemented; however, the earlier you develop an evaluation plan and begin to implement it, the better off your program will be and the greater the benefits will be at the end.

This workbook is designed to assist in developing an evaluation plan. It is not intended to serve as a complete resource on how to implement program evaluation. Rather, it may be used along with other evaluation resources, such as those listed in the Resource Section (Appendix B). Moreover, it is assumed that national/regional strategic health plans and the associated operational plans are available as resources. This workbook is not intended to support the development of a national health strategy or operational plan, nor does it prioritize activities in a program. Although the six-step process can identify gaps in monitoring data that may be helpful for evaluation, this workbook will not assist in the preparation of a monitoring plan.

What is an evaluation?

The purpose of an evaluation is to provide evidence of how and why programs are or are not working in practice (MEASURE Evaluation, 2016). Evaluations measure how well program activities have met expected objectives and/or the extent to which changes in programs are associated with and/or can be plausibly or probabilistically attributed to the program (Frankel & Gage, 2007, rev. 2016).

What is an evaluation plan?

An evaluation plan is a written document that organizes program activities and connects them to outputs, outcomes and impacts; identifies existing and planned data sources; prioritizes evaluation research questions; and determines roles, responsibilities and timelines for answering the research questions (CDC, 2011; Better Evaluation, n.d.).

What about monitoring?

Monitoring data usually address a different set of questions compared to those used in evaluations. Program monitoring requires the collection of routine data for indicators that are used to modify inputs and activities (Frankel & Gage, 2007, rev. 2016). While monitoring data may be helpful in an evaluation by elucidating where and why expected outputs are not occurring, they do not always provide information about impact, outcomes, or the causal contribution of the program.

Development of This Workbook

This workbook was developed by MEASURE Evaluation, and was informed by experiences in operationalizing UNAIDS' *Strategic Guidance for Evaluating HIV Prevention Programs* in multiple countries and contexts. The Rwandan Biomedical Centre successfully used the process to engage stakeholders and develop an evaluation plan that became part of the Government of Rwanda's official five-year strategic plan for HIV. The Ghana AIDS Commission (GAC) convened stakeholders to develop an evaluation plan (GAC, 2013) that was successfully used to advocate for funding for an Integrated Biological and Behavioral Surveillance Survey (IBBSS) and a process evaluation (Reynolds, et al., 2014). The results of the IBBSS subsequently informed USAID's HIV programming.

This workbook was also used in the Dominican Republic by the USAID mission to organize and conduct a strategic information gap analysis with existing HIV programs and partners. Last, the approach was applied in Namibia, where the national key populations (KP) technical working group (TWG), under the Ministry of Health and Social Services and with support from MEASURE Evaluation and USAID, worked to identify and address priority information needs for vulnerable populations.

MEASURE Evaluation finalized this workbook based on these experiences. Illustrative examples of and learning points from these field experiences are provided throughout the workbook.

Intended Users

This workbook is written for program planners, managers, implementers, and M&E managers and staff with M&E responsibilities who want to plan an evaluation that responds to priority questions agreed to by multiple stakeholders. A facilitator who is independent of the program may also be employed. The facilitator can act as an objective party to negotiate the direction of the evaluation, the allocation of resources, and to keep the planning process on track. This workbook is written for programs that want to plan an evaluation that responds to priority questions agreed to by multiple stakeholders.

The workbook was developed and framed in the context of a national HIV program; however, the activities can be conducted on a smaller scale with programs in any sector. Many of the stakeholders involved in pilot field applications came from other sectors, including the military, police, education, prisons, and social work; donor and implementing agencies; and civil society and advocacy groups. The process of developing an evaluation plan and selecting evaluation methods and designs is the same, or at least similar, across other health areas and other sectors. The variety of users to date suggests that this six-step process can be applied successfully in other health areas and sectors.

INTRODUCTION

Background

Global efforts to control the AIDS epidemic cannot succeed without effective HIV prevention and treatment (PEPFAR, 2014a). An array of approaches tailored to local epidemics exists to respond to the diverse needs of people at risk of infection. An increasing focus is being placed on maximizing the effectiveness of combination prevention programs and making progress in the achievement of universal coverage of HIV care and treatment in an era of flat HIV funding (PEPFAR, 2017; Vermund & Hayes, 2013; UNAIDS, 2015; PEPFAR, 2011; UNAIDS, 2010a). There is an urgent need to gather evidence on the most cost-effective strategies to avert HIV infections in critical populations and in diverse settings, and simultaneously increase the proportion of those who are already infected on treatment (Creese, Floyd, Alban, & Guinness, 2002; Walker, 2003). National and regional HIV programs should maximize the effectiveness of their HIV prevention strategies by conducting high-quality evaluations that document programming results. USAID evaluation policy specifically mentions the need for evaluation to achieve "two primary goals: accountability to stakeholders and learning to improve development outcomes" (USAID, 2011). PEPFAR's evaluation standards of practice are reflected in this workbook and its appendices, from the emphasis on stakeholder engagement, to the identification of resources and data collection plans, to planning for data dissemination and use (USAID, 2011; PEPFAR, 2015; PEPFAR, 2014b).

Attention has recently shifted to concerns about appropriate methodologies for feasible yet rigorous evaluations of national HIV programs, especially of prevention programs, which consist of complex packages of interventions of known effectiveness (UNAIDS, 2010b). These programs are unevenly distributed and applied, and are often strategically focused on specific populations, each with their own sociocultural epidemiological context (UNAIDS, 2014b; PEPFAR, 2016). Many of these populations are difficult to recruit and study because of issues related to stigma, criminalization, or social vulnerability (Hart, Iskarpatyoti, Mandal, & Thomas, 2016). Another area of concern is the enhanced use of data for decision making to optimize the efficiency of HIV programming and program sustainability as the response to the epidemic shifts from an emergency response to a long-term strategy for the response (Nutley & Reynolds, 2013; Segone, 2009).

A data-driven approach to HIV prevention remains a top public health and development priority (USAID, 2017). As bilateral and multilateral agencies look to the post-2015 agenda, the importance of strategically targeting geographic areas and populations to achieve the greatest impact for investment is central (PEPFAR, 2016). The Sustainable Development Goals (SDGs) expand on the Millennium Development Goals (Sachs, 2012). Health is a central component of the SDGs. Improved health is a goal in and of itself, with 13 targets, and is recognized as being affected by, a contributor to, and an outcome measure of many of the other economic, social, and environmental SDGs (World Health Organization, 2016). The SDG agenda is more ambitious than the Millennium Development Goals' agenda and, by necessity, collaboration and resources needed to measure the SDGs are equally significant. To that end, USAID, the World Bank, and the World Health Organization convened a global summit in 2015, entitled Measurement and Accountability for Results in Health, to define a new health measurement and accountability strategy. Central to this strategy is aligning stakeholders in support of the collection, analysis, and use of data to increase the success and sustainability of public health programs (Handley, Boerma, Victora, & Evans, 2015).

For HIV-specific programs, PEPFAR 3.0 has adopted ambitious targets needed to reach epidemic control (PEPFAR, 2014a). PEPFAR emphasizes targeting for testing, treatment, suppression of viral load, and adherence to treatment (PEPFAR, 2014a). Treatment is the linchpin of prevention. Additional prevention strategies, applied in a combination prevention program, should be implemented as they are central to the goal of reaching HIV-positive persons with testing and treatment, and to promoting adherence to treatment. HIV interventions should demonstrate success and cost-effectiveness in reaching these goals, which underscores the importance of quality evaluation (PEPFAR, 2014a).

PEPFAR 3.0 and USAID support the fast-track targets put forward by UNAIDS: by 2020, 90% of all people living with HIV will know their status; 90% diagnosed will receive antiretroviral therapy; and 90% on treatment will achieve viral suppression (PEPFAR, 2014a; UNAIDS, 2014b).

THE TREATMENT TARGET

90%	90%	90%
diagnosed	on treatment	virally suppressed
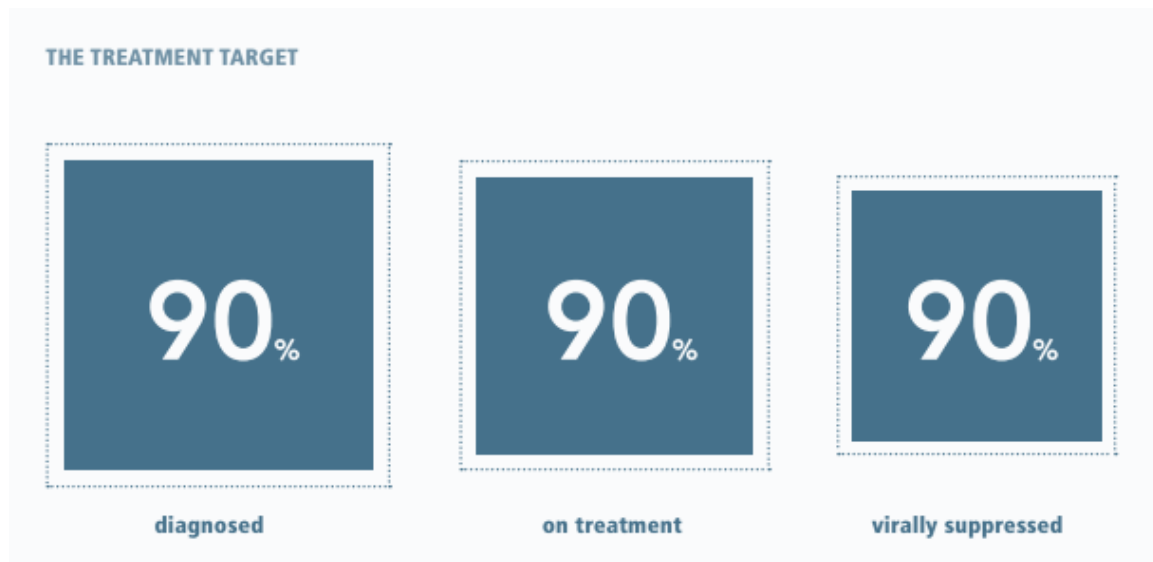
The 95-95-95 goals by 2030 can only be met if the 90-90-90 targets are achieved (UNAIDS, 2014a). The global emphasis on prevention and treatment are inextricably linked with and rely on rigorous implementation science and program evaluation to demonstrate the effectiveness of programming in working toward these ambitious goals.

In 2010, the UNAIDS-led Monitoring and Evaluation Reference Group published the *Strategic Guidance for Evaluating HIV Prevention Programmes* (UNAIDS, 2010b). The document responds to the need for practical evaluation guidelines using appropriate methods, and evaluation approaches that are unified with M&E systems and grounded in realities of the field. As HIV prevention programs implement more complex intervention packages, at scale, and where randomization is not possible, UNAIDS offered the following recommendations:

- Describe the program impact pathway: document the logical progression and relationship of the strategic program elements and their causal relationships.
- Determine what decisions need to be made and whether an evaluation is needed and feasible: identify key questions about the program, and decisions that should be made and when.

- Select appropriate measures to assess program effects: determine how to judge the effectiveness of program components.
- Assess program implementation and program effects: use complementary data collection activities to answer evaluation questions.
- Focus on actionable results: using the Public Health Questions Approach (PHQA) to HIV M&E, organize data collection and analysis to gather the right information and interpret it correctly.

This workbook reflects these recommendations, and operationalizes them into actionable steps.

Overview of the Process

Professional standards for program evaluation have been prepared, used, revised, and adapted for nearly a half-century. The purpose of this workbook is not to reiterate those standards, but rather to apply them by providing a clear, systematic process for developing a program evaluation plan. This workbook describes the process in six steps:

1. Engage Stakeholders

2. Know Your Program

3. Know Your Evaluation Needs

4. Select the Evaluation Design

5. Draft Your Evaluation Plan

6. Ensure Use

The six steps are not always as linear as presented here; they are often implemented through an iterative process. You may need to revisit a step and/or complete other discrete steps concurrently.

Before You Begin

Select the level at which you want your evaluation plan to focus. Do you want to focus on a specific intervention? Do you want to evaluate a package of interventions that a program is implementing? Or maybe you want to look at multiple programs under a national health area? Perhaps you want to look at the effectiveness of these interventions in a specific population or geographic area? These questions will determine the scope of your evaluation, and will affect the stakeholders you engage, the program you describe, and the evaluation design you choose.

STEP 1. ENGAGE STAKEHOLDERS

Identify Stakeholders

The success of a program's evaluation depends on strong commitment from program leaders and from government (if conducted at the national or subnational level). A commitment from these groups for evaluation can ensure that adequate resources are secured, stakeholders actively participate, and information from the evaluation is used (MEASURE Evaluation, 2011).

Stakeholders play an important role in the process of developing an evaluation plan. They are much more likely to buy into and support the evaluation if they are involved in the process from the beginning. Their perspectives can enrich your evaluation plan by clarifying roles and expectations for the program. They can also help:

- determine and prioritize key evaluation questions
- pretest data collection instruments
- assure the inclusion of ethical considerations
- facilitate data collection
- implement evaluation activities
- increase the credibility of analysis and interpretation of evaluation findings
- ensure that evaluation results are reported and used

The exact composition of stakeholders is up to you, but their individual roles and expected contributions should be explicitly defined and agreed to in the evaluation plan.

The options for the types of stakeholders to involve are: program managers and staff; national/regional government representatives; partner organization representatives; clients/service recipients; funders; community representatives; and volunteers. Given that a single evaluation cannot answer all possible evaluation questions raised by diverse and, sometimes, competing groups, it is critical to identify select representatives (about eight to ten) from key stakeholder groups. Key stakeholder groups are generally those that are interested in the program and would use the results, those who implement the program, and those who are served by the program. The process of choosing and engaging stakeholders should be transparent, to facilitate buy-in of the evaluation results from people/groups who did not participate in the evaluation planning process.

The Ghana Experience

Using the *Know Your Epidemic, Know Your Response* approach (UNAIDS, 2007), the GAC determined that the HIV epidemic in Ghana was heavily concentrated in KPs with high risk sexual behaviors, such as female sex workers (FSWs) and men who have sex with men (MSM).

Because of this concentrated prevalence, the GAC created a national TWG on KPs, comprised of important government and nongovernment stakeholders in the country, including representatives of the MSM and FSW communities. This TWG has been in operation since 2010 and has developed many of the background documents used to inform the evaluation plan in Ghana.

By tapping into an already functioning stakeholder group, MEASURE Evaluation leveraged established trust to ensure positive participation and engagement of the communities.

Engage Stakeholders

Stakeholder engagement is central to the evaluation process; however, it may be difficult to get everyone in the same room at the same time. One or more of the following approaches can be used to engage stakeholders as you develop your evaluation plan:

- Conduct a series of meetings: stakeholders are brought together for a series of in-person meetings to share experiences and knowledge, and to come to consensus on aspects of the evaluation plan. This approach requires resources to convene stakeholders in a single location, conduct detailed logistical planning, and secure the services of a strong facilitator (or someone in your program with strong facilitation skills) who can move the conversation along and ensure the optimal use of time. Because distractions can be minimized in a facilitated meeting, conversations and consensus-building may be more productive and efficient in-person compared to a meeting conducted virtually.

- Achieve virtual consensus: stakeholders come together through an electronic medium (such as email, Skype, and/or web forums) to share experiences and knowledge, and to come to a consensus. This approach reduces the amount of resources needed to bring people together; however, consensus may be more difficult to reach. Facilitators need to be dynamic enough to keep the conversation moving, stakeholders engaged, and the discussion on-track.

- Conduct individual consultations: a consultant meets with stakeholders individually or in small groups, and synthesizes their input. This is a quick way to develop a plan; however, because stakeholders are not aware of each other's input, misunderstandings or competing priorities may impede consensus building.

All stakeholders, including funders, government officials, and program implementers have varying, and sometimes, competing interests. Useful evaluations are not about special research interests or what is easiest to implement; they are about what information will be used to improve the program and make long-term decisions.

The Ghana Experience

The development of the evaluation plan in Ghana was a participatory process involving the GAC, other partners in Ghana working to prevent HIV in KPs, and MEASURE Evaluation. MEASURE Evaluation facilitated the overall process, engaging stakeholders throughout to gather information needed and ensure activities were completed in a timely manner.

MEASURE Evaluation organized three two-day, in-person meetings over 12 months to gather information and stakeholder input, and a final meeting of the TWG to approve the evaluation plan. Specific details about this process are described elsewhere (Reynolds, et al., 2014). The process respected the GAC's leadership and facilitated its ability to exercise its leadership. By employing an external facilitator, all parties working on KP programs and M&E research in Ghana could meet and discuss the full range of activities taking place in a structured, yet participatory way. MEASURE Evaluation was then able to objectively synthesize the input to develop a comprehensive evaluation plan.

A similar process was used in Namibia with its KP TWG.

The Dominican Republic Experience

The USAID mission in the Dominican Republic sought to conduct a gap analysis of current M&E efforts against the already-identified strategic information needs of existing HIV programs and partners. MEASURE Evaluation conducted independent consultations with stakeholders from USAID, CDC, and implementing partners to understand ongoing M&E activities. MEASURE Evaluation synthesized the information using tools from this workbook. The resulting gap analysis was presented to the groups to plan to address the identified data gaps.

At this point, you have:

✓ Identified a target program
✓ Organized program stakeholders

WORKSHEET 1.1. STAKEHOLDER IDENTIFICATION EXERCISE

Adapted from MEASURE Evaluation, 2011 and CDC, 2011

Your stakeholders could include program managers and staff, national/regional government representatives, partner organization representatives, clients/service recipients, funders, community representatives, and volunteers. However, given that a single evaluation cannot answer all possible evaluation questions raised by diverse and, sometimes, competing stakeholder groups, it is important to identify select representatives from key stakeholder groups. To begin the selection process, list all possible stakeholders, with corresponding comments about the possible roles for each person, related to the evaluation and the use of evaluation results; the type of stakeholder (government, funder, partner organization, client, etc.); and their availability during the planning process.

Person/Group	Type of stakeholder	Role related to the evaluation	Availability during process	Additional comments

Review the list of possible stakeholders and prioritize each person/group based on the information/input they could provide or how they may be affected by the evaluation results. You could rank each stakeholder as "high," "medium," or "low" or you could rank them in numerical order (i.e., from "1" to "n"). You should also consider the diversity of the stakeholders; ensure that you have representatives from government, funders, private organizations, and clients/service recipients, and make rankings within these subgroups. It is recommended that you have a total of eight to ten representatives in your stakeholder group.

Note: There may be key stakeholders who are not members of your stakeholder group(s) and who still have an important role in the evaluation. It is essential that the selection and engagement processes for stakeholders are transparent to facilitate buy-in of the evaluation results from those who do not participate in the evaluation planning process.

STEP 2. KNOW YOUR PROGRAM

The second step in preparing your evaluation plan is to know what it is you will evaluate. This section helps you map the logical progression of your program's planned and ongoing activities, their causal relationships, and their intended results.

Collect Background Documents

A document and literature review can help you understand the history, philosophy, and functioning of the program you plan to evaluate and the context in which it operates. The documents can be hard copy or electronic, and may include internal records, program logs, funding proposals, external reports, and strategic materials. Different documents provide different information. You may want to review existing M&E plans that list the data and indicators the program expects to collect and measure, operational strategies that describe the methods the program uses to reach its objectives, and reports that present the activities the program implements to achieve its results. If you intend to evaluate a large national or regional program, you will also want to review relevant national and regional plans/strategies. Investigate what types of documents exist and determine which ones you think will help define and answer your evaluation questions. The following are illustrative documents that may be useful in developing your evaluation plan:

- National/regional HIV[1] M&E plans
- National/regional HIV[1] strategies
- National/regional data reports (such as IBBSS, PEPFAR Health Impact Assessments [HIAs], Demographic and Health Surveys [DHS], etc.)
- Program M&E plan(s)
- Program strategy(ies)
- Program operational plan(s)
- Program report(s)
- Literature review on current epidemiological trends and treatment and/or prevention priorities

Compose a Program Impact Pathway

A Program Impact Pathway (PIP) (or program logic model) is a systematic way to organize and present the relationship between planned activities and measurable objectives in a specific context (UNAIDS, 2010b). A PIP is generally composed of a program's planned work (resources/inputs and activities) and its intended results (outputs, outcomes, and impact) (Figure 1).

---

[1] Or other sector of interest

Figure 1. A simple PIP
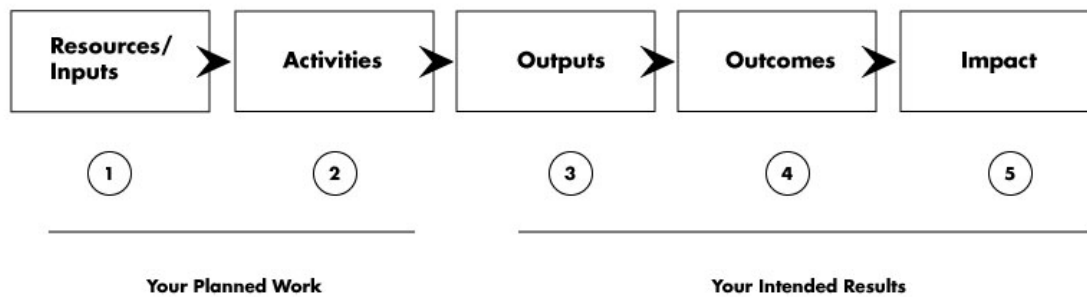
The PIP draws on existing evidence and experience within your health area. It is developed through an iterative process of discussion, analysis, and justification of the components, causal relationships, and feedback loops. For example, a program may already have an operational plan that lists planned activities and there may be related national/regional strategies that describe desired results. However, not all activities, outputs, and outcomes may be reflected in the background documents you collect. For a clear, comprehensive inventory of planned and ongoing activities and their measured or expected indicators, stakeholders and decision makers need to be engaged.

The PIP framework presented here is linear, but the sociopolitical framework in which many health and development programs—especially those working with marginalized or vulnerable groups—are embedded is complex. Explicit recognition of these complexities can provide much-needed context to illuminate the relationships between a program's implementation and its impact.

Appendix A1 provides an example of an HIV program's PIP. Note that PIPs can be structured using numbered lists in columns to aid discussion, or rows to order and show relationships among components, and/or box and arrow formats to illustrate causal linkages among components. To identify and categorize the elements, a description of each element, relevant examples, and potential sources follow. It may be easiest to start with your intended impact and work backwards, if you are working with existing plans. Alternatively, consider the activities you have planned and move forward from there to define your intended impact.

The elements of the PIP are:

Resources/inputs: Identify the available resources for your program. This helps you determine the extent to which you can implement the program and achieve your intended goals and outcomes. List the resources that you currently have to support your program. Second, list all resources you will need for a successful program, whether you have them in hand or not. You may wish to separate resources under the headings "need" and "have." If you intend to raise additional funds for the program during the program implementation timeframe, account for them under "activities."

> Examples of inputs: staff; facilities; materials; funds
>
> Sources of information: stakeholder knowledge; operational plans

Activities: Activities are the actions needed to implement the program—what you do with program resources to achieve measurable results. It is often helpful to group related activities together. The types of activities depend on your program's size and how you administer it. For a large, complex

program, there may be several types of activities (e.g., training, promotional activities); smaller programs may implement just one or two types of activities. Each type of activity will require specific activities to implement it. The main types of activities will have already been defined in a national or regional operational plan. Information from this plan should be supplemented with input from stakeholders, who may be more knowledgeable about what is happening or needed in the field.

> Examples of types of activities: events; trainings; workshops; promotions; data collection

> Sources of information: stakeholder knowledge; operational plans

Outputs: Outputs are the direct products or results of program activities. They are not the changes you expect the program to produce, but rather, steps along the way to your intended results. They are usually expressed in terms of the scope, reach, and coverage of the program—whether the program was delivered to the intended audiences at the intended "dose." An output statement does not reveal anything about quality. The evaluation will assess the quality of your outputs.

> Examples of outputs: number of classes taught, meetings held, materials distributed; program participant rates; total hours of service delivery

> Sources of information: Operational plan; National/regional strategy

Outcomes: Outcome measures represent the actual changes that occur or the difference a program makes on individuals, groups, families, organizations, systems, or communities that are directly related to its goal(s) and objectives. You may have summary outcomes, or you may want to break them out into short-term and long-term outcomes.

Short-term outcomes are results you expect to achieve after one to three years of program activity. They are specific changes in such things as people's attitudes, behaviors, knowledge, skills, or health status that are the result of program activities. They are usually expressed at an individual level among program participants.

> Examples of short-term outcomes: New knowledge; changed opinion/values; increased skills; changed motivation; changed attitudes; changed aspirations

> Sources of information: National/regional strategies; strategic plan

Long-term outcomes are results you expect to achieve in four to six years. Long-term outcomes are also specific changes in things like attitudes, behaviors, status, or practices expected to result from program activities. These usually build on the progress expected by the short-term outcomes.

> Examples of long-term outcomes: Modified behavior; changed policies; changed practices; changed social action

> Sources of information: National/regional M&E plan

Impact: Impact refers to the results expected after seven to ten years of program implementation, the social change your program is working to create. Impacts are the kinds of organizational, community, or system-level changes expected to result from program activities, which may include improved conditions, increased capacity, and/or changes in the policy arena.

Examples of impact: Changed health condition; changed human condition; changed civic condition; changed economic condition

Sources of information: National/regional M&E plan

ACTION: *Use worksheet 2.1: Program Impact Pathway Template to organize the program's identified inputs, activities, outputs, outcomes, and impact.*

The Ghana Experience

With the assistance of the Most At-Risk Population (MARP) TWG, the GAC developed a National Strategic Plan for Most At-Risk Populations, 2011-2015 (GAC, 2011c), and a MARP Operational Plan Framework (GAC, 2011b) to achieve the goals and objectives outlined in the National Strategic Plan (NSP) for HIV and AIDS (GAC, 2010). MEASURE Evaluation used these documents to develop a logic model (PIP) for the national KPs program.

Information on "activities" was drawn from the Operational Plan. "Outputs" were derived from the MARP Strategy. "Outcomes" were drawn from the NSP and the MARP Strategy. "Impact" was taken from the NSP. This information was organized by MEASURE Evaluation into the PIP/logic model, and presented to stakeholders at the first in-person meeting. Using existing documents to construct the PIP was efficient; however, the model was incomplete, especially in terms of key program outcomes.

Working in small groups, stakeholders were asked to review the PIP and provide input on the process, output, and outcome measures, and the indicators and data sources for those measures. This input was integrated into the draft PIP.

Participants reported that the process of reviewing the PIP/logic model was helpful to better understand the theory behind specific services in the national program and in their own organizations.

The Namibia Experience

The Namibia KP TWG was involved in preparing the KP section of the National Strategic Framework (NSF), which defines the output, outcome, and impact indicators for which KP programs would be responsible. No logic model specifically geared for the national KP program had been developed.

MEASURE Evaluation met with individual stakeholders to understand what programmatic activities were currently occurring or planned. These were then consolidated into overarching strategies and tied to the NSF output, outcome, and impact indicators. MEASURE Evaluation presented the draft PIP to the KP TWG, at which point the stakeholders modified the strategies and added indicators that they felt were missing from the NSF framework. This input was integrated into the draft PIP.

At this point, you have:

✓ Identified a target program
✓ Organized program stakeholders
✓ Collected background program information
✓ Composed a program impact pathway

WORKSHEET 2.1. PROGRAM IMPACT PATHWAY TEMPLATE

A Program Impact Pathway (PIP) illustrates the relationship between planned activities and measurable objectives in a specific context. Begin by identifying the components of your program and organize them in the template below.

Inputs	Activities	Outputs	Outcomes (short term)	Outcomes (long term)	Impact
Resources needed for program implementation	Actions needed to implement the program	Direct results or products of program activities	Results expected 1-3 years after an activity is under way	Results expected 4–6 years after an activity is under way	Future health change(s) your program is working to create

PIPs can be organized using numbered lists in columns to aid discussion, or rows to order and show relationships among components, and/or box and arrow formats to illustrate causal linkages among components. Appendix A1 provides sample PIPs.

STEP 3. KNOW YOUR EVALUATION NEEDS

Evaluation needs are inextricably linked to the program's planned activities and expected results, and are limited by the types of questions your program can realistically answer and wants to answer. This section describes how to use an analytical approach (known as the PHQA) to map the programmatic approach and expected outcomes, identify the program's gaps in knowledge, and determine available data sources.

Identify Questions, Gaps, and Additional Information Needs

The PIP identifies the relationship between planned activities and measurable objectives in a specific context. To measure progress in the PIP, you need different sources of information and different research or evaluation activities. Therefore, you need to make an inventory of ongoing and planned research and M&E activities (be sure to distinguish between planned activities, and planned and funded activities), and map them to the components of your PIP. Engagement with stakeholders can help facilitate this process, by clarifying details of research/M&E activities and how they fit into the evaluation plan. Note any PIP component(s) and indicator(s) that are missing or are not measured by a distinct data source. A literature review of primary research and best practices should be used to supplement this input. A template and examples are provided in worksheet 3.1 and Appendix A2, respectively.

> **ACTION:** *Use worksheet 3.1: PIP Gap Analysis to organize and identify gaps in data sources, indicators, and data availability.*

Map M&E Activities to the Public Health Questions Approach

At this point in the evaluation planning process, you have mapped the interventions your program is implementing, where, for and by whom, and on what timeline. You have formed a common understanding across your stakeholder groups about the expected outputs, outcomes, and impact of your program, if implemented as planned. At this stage, you need to understand how this information can be leveraged to answer larger public health questions. For example, you may want to know what factors are contributing to the public health problem. You might want to know what additional interventions could or should be implemented. Or your stakeholders might have more direct questions about a program's response: Is the program working? Is the program reaching enough people in the right places with the right interventions to realistically achieve expected public health outcomes and impacts?

The Public Health Questions Approach is a way to organize the data collection and analysis methods to understand where a program lies in the analytical process (Figure 2) (UNAIDS, 2010b). Categorizing ongoing and planned research/M&E activities, and identifying where a program is on the PHQA can help define the purpose, methodology, and scope of your evaluation plan. As you identify your research questions and anticipated results, note the programmatic questions, data gaps, or information needs that are not addressed.

Figure 2. The public health questions approach
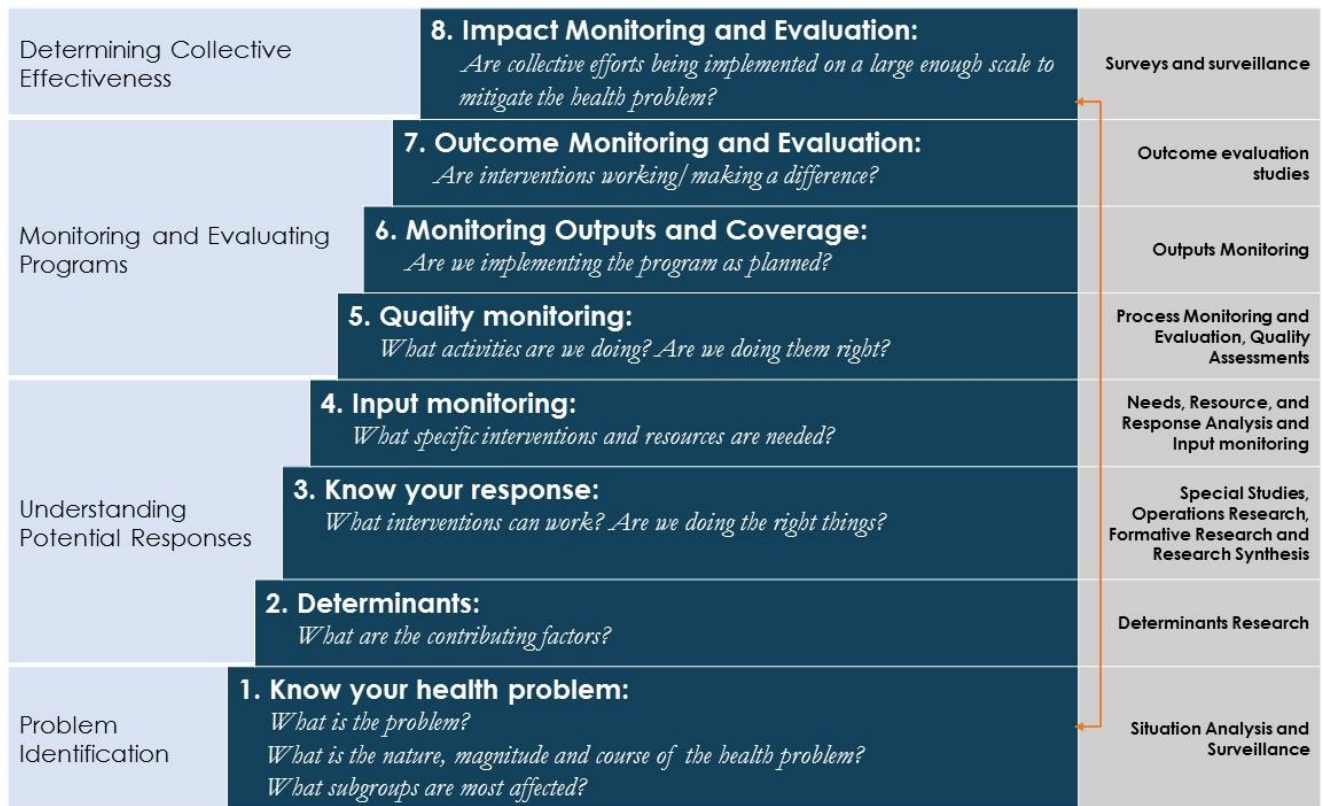
Source: Adapted from UNAIDS, 2010b.

Each step provides a foundation for the next; however, the steps are not necessarily conducted in the logical, sequential order presented here. Programs can be in different stages in the PHQA, and move backwards and forwards as information emerges or evolves. The PHQA approach is also appropriate for designing complexity-aware evaluations, by driving the consideration of what sorts of contextual factors such as social norms or stigma could affect the definition of the health problem, the selection of potential responses, or the uptake of services.

When applied at the program level, the PHQA can:

- Serve as a framework to organize ongoing M&E activities and other data sources
- Reveal data gaps
- Facilitate the articulation of information needs relative to decision points
- Facilitate planning to fill data gaps
- Obtain a thorough understanding of the health problem and programmatic response

The following sections review the questions that each step in the PHQA poses, and provide brief descriptions of the process and the data required to answer the questions. Many programs may already have information to answer the questions. Categorizing your current and planned research/M&E activities, and identifying where your program is on the PHQA can help you define

the purpose, method, and scope of your evaluation plan. Additional guidance may be found in the UNAIDS *Strategic Guidance for Evaluating HIV Prevention Programmes*.

Know Your Health Problem

What is the problem? What is the nature, magnitude, and course of the health problem? What subgroups are most affected?

Knowing the extent of the health problem for the populations and in geographical areas of the country is key for planning the right mix of prevention strategies. To have an impact on disease incidence, program efforts should be directed to the appropriate populations and behaviors in the appropriate locations and settings. Program planners should have information on disease incidence in the country by geographic location, population, and time. This can be identified through sentinel surveillance, national and regional surveys, rapid assessments, participatory mapping of the response, and consultations with vulnerable populations and service providers.

Determinants

What are contributing factors?

Once you have identified when, where, and who is most affected, the next natural question is why and how this occurred. It is important to understand the relationship among the biomedical, behavioral, and structural drivers of risk and vulnerability. This information is gathered through "determinants research" with affected communities, including situational analysis and targeted quantitative investigations (knowledge, attitude, and practice [KAP] surveys, epidemiological risk factor studies) and/or qualitative investigations (participatory action research, ethnographic studies).

Know Your Response

What interventions can work? Are we doing the right things?

This step involves determining which interventions might work under ideal circumstances (efficacy) and in real-world settings (effectiveness). Drawing on existing information through systematic and meta-analytic reviews, program managers can elucidate what works to mitigate the health problem. However, a review of existing evidence and previous efficacy and effectiveness studies of prevention programs and interventions might not be sufficient because they often fail to address replicability, external validity, or methodologic challenges. Engagement with stakeholders and development of a PIP can supplement this information with practical program application knowledge.

Input Monitoring

What specific interventions and resources are needed?

This step focuses on identifying the specific interventions needed to reach specific populations and settings, and what resources are required to implement them. This step draws heavily on the previous three to select an appropriate package of interventions for a specific population, at scale, and delivered with a minimum level of quality to achieve desired results. A detailed resource analysis alongside your PIP can help inform this decision.

Quality Monitoring
What are we doing? Are we doing it right?

Before you can evaluate the outcome or impact of your program, you must first establish whether and how the program was implemented. Begin by assessing what activities the program has set out to conduct through its strategies and operational plans. Next, determine the quality of these activities and the data collected from them. While outputs are generally used for program management and accountability, they can inform an evaluation of the program's quality. This information is usually gathered through routine program documentation, expenditure data, and client records.

Monitoring Outputs and Coverage
Are we implementing the program as planned?

Outcome and impact evaluations require information on how the program was implemented. This is generally provided through routine *process evaluations* that identify what is working well in program implementation and where there may be problems. This includes information on the availability of services (geographic and target population); access to services; how services were delivered; client satisfaction; and uptake of services.

Outcome Monitoring and Evaluation
Are interventions working/ making a difference?

While interventions are chosen according to the best available evidence, it is important to measure the effectiveness of activities on outcomes in your target location and in your target population(s). A baseline should be established, either through program data collection or existing data sources, before implementation (or as early as possible). If process evaluations do not reveal any problems with implementation, changes in behavioral/social/structural outcomes and disease incidence should be assessed for plausible association. You should collect and triangulate data from multiple sources, such as annual surveillance, representative population-based surveys (DHS, PEPFAR HIA), IBBSS, and special studies.

Impact Monitoring and Evaluation
Are collective efforts being implemented on a large enough scale to mitigate the health problem?

This step determines the population-level effectiveness of the program. Like the step before, data should be collected and triangulated from multiple national sources, the intent being to have a basic minimum package of comparable, consistent national data sets over time, which allow for trend analysis. Modeling techniques can elucidate how national portfolios relate to these trends.

ACTION: *Use worksheet 3.2: Ongoing M&E Activities Mapped to PHQA Template to organize research/M&E activities in the PHQA model, and note any gaps or questions that arise.*

The Ghana Experience

The first of three in-person meetings among the GAC, MARP TWG, and MEASURE Evaluation was organized to gain an understanding of the ongoing and planned M&E and research activities on KPs that were being conducted in Ghana. Participants were asked to present on their research and M&E activities, including, but not limited to, the IBBSS (GAC, 2011a), formative studies, resource analyses, and routine program monitoring. MEASURE Evaluation used this input and the MARP Strategy (GAC, 2011c) and Operational Plan (GAC, 2011b) to map ongoing and planned research and M&E activities to the PHQA model. The project also noted the anticipated results each activity intended to yield.

Questions, gaps, and information needs were noted by TWG participants and MEASURE Evaluation facilitators.

At this point, you have:

- ✓ Identified a target program
- ✓ Organized program stakeholders
- ✓ Collected background program information
- ✓ Composed a Program Impact Pathway
- ✓ Mapped ongoing and planned research and M&E activities to the PHQA model
- ✓ Identified information gaps
- ✓ Determined your evaluation needs

WORKSHEET 3.1. PIP GAP ANALYSIS

Once stakeholders have agreed on a PIP, you need to identify indicators and data sources that account for each program component described. Copy the descriptions of program components from your PIP into the second column in this worksheet. Data sources and indicators come from (but are not limited to) background documents (national HIV strategy, program strategies, or operational plans); routine program documentation (reports, logs, inventories); surveillance data; representative population-based surveys (DHS, AIDS Indicator Surveys); IBBSS; and special studies.

Logic level	Description	Indicators	Data source	Data available
Activities				
Outputs				
Outcome (short term)				
Outcome (long term)				
Impact				

Indicate information gaps by highlighting the appropriate cells.

WORKSHEET 3.2. ONGOING M&E ACTIVITIES MAPPED TO PHQA TEMPLATE

The Public Health Questions Approach (PHQA) is a way to organize the data collection and analysis methods to understand where a program lies in the analytical process. Categorizing ongoing and planned research/M&E activities, and identifying where a program is on the PHQA model can help define the purpose, methodology, and scope of the evaluation plan. As you identify your research questions and anticipated results, note programmatic questions, data gaps, or information needs that are not addressed.

Public health question step	Ongoing or planned research/data collection	Anticipated results	Questions, gaps, additional information needs
1. Know your health problem: What is the size and nature of the problem?			
2. Determinants: What are the contributing factors?			
3. Know your response: What interventions can work? Are we doing the right things?			
4. Input monitoring: What specific interventions and resources are needed?			
5. Quality monitoring: What activities are we doing? Are we doing them right?			
6. Monitoring outputs and coverage: Are we implementing the program as planned?			
7. Outcome monitoring and evaluation: Are interventions working/making a difference?			
8. Impact monitoring and evaluation: Are collected efforts implemented on a large enough scale to mitigate the health problem?			

STEP 4. SELECT THE EVALUATION DESIGN

Identify Your Research Question(s)

Now that you better understand your program, its evaluation needs, and information gaps, you need to identify your specific research question(s). In this step, you should solicit evaluation questions from the stakeholders. The stakeholder group(s) should consider:

- What do we want to be able to say about the program?
- What evidence will represent success of the program?

These questions should be considered through the lens of your PIP and the PHQA model, and weighed against the available data sources and strength of evidence produced by different study designs. For example, if the evaluation needs to unequivocally demonstrate attribution of impact to a program, then a full-scale impact evaluation design is needed, requiring commensurate investments of time and resources. However, if the program is seeking information about the scope and quality of interventions implemented, for example, in a situation where the effectiveness of the intervention being carried out has already been demonstrated in previous studies, then a less resource-intensive process or outcome evaluation may meet the needs of your stakeholders.

In step 3, you mapped your available data to the PHQA model, and came to an understanding of what evaluation questions you can currently answer based on what steps in the model have adequate information sources. You may, however, need to plan for additional research to fill the information gaps.

The Ghana Experience

Stakeholders were asked to review the PIP and PHQA model to 1) identify whether the pathways by which the HIV prevention program for the MARPs were sufficient to influence desired results; and 2) determine whether the desired results were complete and measurable with existing indicators and data sources. Gaps in M&E activities were identified, and discussion about how to fill the gaps at the national level were facilitated and recorded by MEASURE Evaluation.

As you think about how to address the information gaps, it is important to think about what research would be "nice to do" versus what is a "must do" to answer the evaluation questions. The amount of information you can gather about your program is potentially limitless; however, evaluations are restricted by the questions that can be realistically answered, and the data that can be collected with quality and with finite resources.

For example, a small program that collects only process monitoring information may focus on questions about the quality of its program (PHQA step 5: Are we doing it right?) while a large national program with more resources that has recently scaled up an intervention might want to conduct a survey to determine not only the scale of the problem (PHQA step 1: Know your health problem) but the effectiveness of its efforts (PHQA step 8: Impact M&E).

The Namibia Experience

MEASURE Evaluation mapped ongoing and planned M&E activities to the PHQA model. The mapping was reviewed by the KP TWG, and a discussion was facilitated around three key questions:

1. Which of the public health questions do we need to answer when we next revise the National Strategic Plan?

2. Which of the priority public health questions do we already have data collection planned to address?

3. Which of the public health questions do we need to answer but do not currently have data collection to address?

In doing this, it was noted that while there was expected data for the PHQA model's steps 1, 2, and 3 and for steps for 7 and 8, the data were not complete. There were also very few process indicators to address steps 4, 5, and 6. A discussion around these gaps led to the prioritization of two questions: a) what is the size of the KP and the HIV epidemic in this population? and b) how are programs addressing this population?

The Dominican Republic Experience

In reviewing the PIP and PHQA model, stakeholders found that by harmonizing existing frameworks, they could understand what data were already available to answer priority questions and what data were missing. Of 37 priority indicators identified to measure results, data for 27 indicators were already being collected. Ten indicators were identified as needing to be better standardized and integrated into existing data collection activities.

Choose an Appropriate Evaluation Design

Once you have chosen the evaluation question(s) with your stakeholders, you should select the appropriate evaluation design to answer the question(s). You may have already identified possible data sources to answer some of your evaluation question(s). For example, the national HIV response in many countries tends to have strong program and outcome monitoring and reporting systems, including surveillance activities.
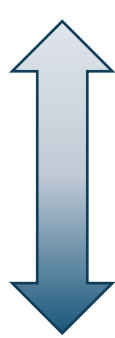
Gaps in information may remain require additional data collection. It is best for research question(s) to inform the appropriate study design and data collection methods by:

- Keeping in mind the PIP, PHQA model, and the gaps that exist
- Determining if/what type of baseline can be established (e.g., at what point in program implementation is the evaluation being planned)
- Identifying sources of evidence and responsible parties for the collection of quality data
- Thinking about what will constitute credible evidence for users
- Confirming that the method fits the question(s)

Decisions also need to be made regarding resource allocation and to adjust and improve programs. Although information should be routinely used during all phases of the program cycle, quality information is especially needed when planning a program, at midterm to make corrections, and at end term to know whether goals were achieved. Before you plan an evaluation, your stakeholders need to answer two questions: What do you want to measure (provision of services, use, coverage, or impact)? How sure do you want to be? (inferences of adequacy, plausibility, or probability)? (Habicht, Victoria & Vaughan, 1999).

The appropriate study design will yield a certain "level of evidence," which influences how certain one is about the results. Different study designs have different strengths. A discussion of the distinction among adequacy, plausibility, and probability designs follows (Table 1). Complexity-aware evaluation methods may be particularly important where social and political context (such as stigma or criminality) may influence the successful implementation of the program (MEASURE Evaluation, 2017).

Table 1. Levels of evaluation evidence

Type of Evidence	Type of Statement	Compared to what?	Strength of Evidence
Adequacy	The expected change occurred	• No control/comparison group • Pre-defined criteria/value reached or not	Weakest
Plausibility	Program seemed to have an effect based on a step-by-step ruling out of alternative explanations	• A non-random control or comparison group/area	
Probability	The program caused the change with only a small probability that the difference between intervention and comparison group was due to confounding, bias, or chance	• A control or comparison group/area selected by randomization	Strongest

Source: Habicht, Victoria, and Vaughan,1999

Adequacy evaluations assess whether the program activities met their objective(s), and whether indicators of behavior or health changed among beneficiaries or in the general population. Because there is no control or comparison group, adequacy evaluations are limited to describing whether the expected changes took place. These evaluations are often referred to as observational studies. When measuring coverage or impact, it may be difficult to infer that any observed improvements were due to the program. The observed improvements may have been caused by outside influences, and changes may have taken place regardless of the program's activities.

Plausibility evaluation designs determine whether a program has attained expected goals, and identify changes as potential effects of program activities rather than external or confounding sources. Plausibility evaluations attempt to control for the influence of external factors by employing control groups. These designs are appropriate when random assignment into intervention and control groups is not feasible. Plausibility designs may rely on comparison groups that are non-random or use other types of controls, such as an historical control group or control groups constructed based on program exposure (dose-response relationship). Ideally, a plausibility assessment will incorporate baseline and post-intervention data points to explicitly show improvements in target indicators. By measuring identical indicators in controls, evaluations can better link outcomes to program activities by eliminating other external and confounding factors. The non-randomized nature of control groups allows for certain selection bias confounders that are not accounted for in the analysis. However, plausibility may be "good enough," as the evidence supports a plausible link between program operations and outcomes.

Probability evaluation designs, like plausibility designs, seek to determine the success of a program's activities and outcomes. However, probability assessments use the most robust study design—randomized control trials (RCTs)—to determine the true effect of the program's activities on the indicators of interest. This type of assessment is the most expensive and time-consuming, so it should ideally be used only when evaluators and stakeholders have found it necessary for funding or research purposes. Depending on the evaluation and associated project, it may be impossible or unethical to conduct a true RCT. Moreover, this design is not fully feasible if the evaluation is not

discussed in the initial phases of program planning because randomized control is required, and is difficult to conceive mid-intervention. Evaluators should be involved early in program design and implementation to ensure that program activities are rolled out in a way that meets requirements for randomization. Typically, programs are rolled out to meet the greatest needs first, and by definition, this is not random.

Beyond the level of evidence, there are additional factors that influence evaluation design, which are intertwined with the level of evidence attainable (Skiles, Hattori & Curtis, 2014):

- Identification of beneficiaries: How complete and reliable is your sampling frame? How dynamic is your target population?
- Comparison group: Can a comparison group be identified/available? Are members of the comparison group exposed to another intervention that affects the same outcome (contamination)? Has the intervention had an impact on individuals not in the intervention group, with the potential to bias estimates of program impact (spillover)?
- Scale: Are you looking at a few instances, or a small group, or an entire population? Do you want representative information that can be generalized?
- Resources available and cost: RCTs are costly and may require resources your program does not have. You may need to balance rigor with reality.
- Timeline: When do you need results? When did your program start? Is there a baseline to which you can compare to reach a plausibility design? Do you have the flexibility to follow a program longitudinally through a probability design, or do you need "quick and dirty" information now through a cross-sectional adequacy design?
- Ethics: Is it ethically sound to randomize groups for a probability design, thereby excluding some from the program?

It is beyond the scope of this workbook to discuss in detail the complexities of all possible evaluation designs, methods, or data sources. Additional resources on several specific study designs are provided in Appendix B. It is important to remember that not all methods fit all evaluation question(s). One thing is certain, quality assurance procedures should be put in place so that data are collected in a reliable way, coded, and entered correctly. The evaluation plan should include a detailed analysis plan. Planning for analysis reduces missed opportunities to collect data that can be turned into meaningful, useful, and accessible information. It is equally important to avoid becoming "data rich but information poor" by focusing all efforts on collecting data, but not taking the time to prepare for analysis, interpretation, and conclusions.

At this point, you have:

✓ Identified a target program
✓ Organized program stakeholders
✓ Collected background program information
✓ Composed a Program Impact Pathway
✓ Mapped ongoing and planned research activities to the PHQA model
✓ Identified information gaps
✓ Determined your evaluation needs
✓ Prioritized evaluation questions and discussed design issues
✓ Linked available indicators and/or measures to your evaluation questions

The Ghana Experience

At the first in-person meeting, stakeholders discussed possible research questions based on ongoing and planned research activities and the program's evaluation needs. Breakout groups used information from the PIP gap analysis and the PHQA model to list and prioritize research questions. Groups then reconvened for a plenary discussion to refine/prioritize the research questions, and review potential study designs and their integration into the existing M&E systems. MEASURE Evaluation facilitators synthesized this information to determine the gaps in information and specific decisions needed.

At the second in-person meeting, the facilitators provided overviews for participants on the types of research questions and the strength of evidence produced by study designs. The point was to demonstrate how the selection of a research question should be based on *what is necessary to know and for what purpose,* not based on what is "nice to know." Presentations provided Ghana-specific examples of what the existing M&E system was likely to yield in terms of evidence. Combining sources of data would demonstrate that changes in HIV prevalence or outcomes occurred at the same time program reach was intensifying. It would not, however, be possible to say that the program resulted (or "caused") these changes. Participants were then asked if this was "good enough" and "what evidence would represent success of the MARP program in 2015?"

Participants revealed that it was equally important to understand how the program was implemented and how HIV prevalence had changed over time. Primary and secondary research questions were defined and prioritized by participants as: 1) Are changes in outcomes due to the implementation of services and program components? 2) Are there changes in behavioral outcomes and HIV prevalence and incidence over time? 3) To what extent are planned MSM and FSW program activities realized/implemented and with improved quality?

MEASURE Evaluation recommended a plausibility evaluation design to answer these research questions. Random assignment into intervention and control groups, and the identification of a traditional control group were not feasible because the program was implemented in areas with the greatest need. The evaluation plan used a post-test only, non-equivalent design using data from the IBBSS, supplemented by other data to rule out alternative explanations.

A plausibility design was appropriate because the program includes interventions of known effectiveness; it systematically addresses alternative explanations for observed trends in behavior; and is feasible to implement, even among hard-to-reach populations. A plausibility evaluation was "good enough" to inform strategic plans and produce guidelines to standardize program implementation.

ACTION: *Use worksheet 4.1: Research Question(s), Data Sources, and Analysis Worksheet to link available data sources, indicators, and/or measures to the evaluation questions.*

WORKSHEET 4.1. RESEARCH QUESTION(S), DATA SOURCES, AND ANALYSIS

You should solicit evaluation questions from the stakeholders by asking "what evidence will represent success of the program?" The objective of this exercise is for stakeholders to identify what answers they want to have about their program. The questions should then be considered through the lenses of the PIP and PHQA model to ensure that you are asking questions in line with the logical flow of your program, and that are weighed against the state of development of a program, the available data sources, and the strength of evidence produced by different study designs. Use the following table to organize this information.

Research question	Location in the PHQA model	Type of evaluation needed (process/ outcome/ impact)	Potential data source(s) (existing/new)	Strength of evidence

Once you have organized the information, your stakeholders can discuss which questions are most important for the program to answer (e.g., what is "must know" versus "nice to know"). Consider which questions provide results that are useful for program improvement or scale up, which can be answered with available data, and/or which are within a program's available resources to answer. While no chart, grid, or exercise can fully answer the question of how to focus your evaluation, the information above can help facilitate informed discussions, and avoid evaluation activities that are misaligned with the program needs and available resources.

With these considerations in mind, your stakeholders should prioritize the proposed research questions and collectively select one to three research questions for the evaluation plan:

Primary Research Question: _____

Secondary Research Question (Optional): _____

Secondary Research Question (Optional): _____

Once you have identified your main research questions, it is important to ensure that you have (or will have) the appropriate data to answer the questions and to understand what analytical methods will be needed for interpreting the data.

Research question	Data source(s)	Timeline for collection	Analysis method(s)
Primary:			
Secondary:			
Secondary:			

STEP 5. DRAFT THE EVALUATION PLAN

An evaluation plan defines the steps needed to assess a program by highlighting program goals, clarifying measurable program objectives, and linking program activities to intended outcomes. The research and M&E activities identified in the previous sections will influence the time, resources, expertise needed, and other inputs required to implement your evaluation plan. You should begin to put together the pieces of the previous sections into a reviewable document.

Roles and Responsibilities of Stakeholders

Writing the evaluation plan will not ensure that the evaluation is implemented. A critical element is to identify the roles and responsibilities of donors, governments, program staff, evaluation staff, and other stakeholders.

This extends to data sources and collection. Partners are perhaps willing to share data, but who will be collecting, reviewing, and archiving the data? When assigning roles and responsibilities, it is important to note any applicable oversight or procedures to which stakeholders should adhere. This may include nondisclosure agreements, data sharing agreements, archiving, data access procedures, etc.

Discussion of these issues should occur throughout the evaluation planning process. The information may change throughout the planning and implementation phases of the evaluation and should be updated accordingly.

ACTION: *Use worksheet 5.1: Roles and Responsibilities Exercise to determine the activities and inputs associated with planned activities and who will carry them out.*

Budget

Discussion of the budget and resources has likely occurred throughout the evaluation planning process. Based on experience, basic M&E activities should account for 5 to 10 percent of the total program budget. When rigorous special studies are planned, 15 to 25 percent of the total program budget may be needed. The evaluation questions and the associated evaluation methods and analyses selected have a direct relationship on the resources required (both human and financial).

It may be difficult to estimate the exact costs of the evaluation activities. Salary costs make up a large part of evaluation budgets, which may not be possible to estimate without knowing exactly who will be involved. Travel costs are also significant, and are also dependent on where evaluation partners are located.

ACTION: *Use worksheet 5.2: Budget Template to organize needed resources, costs, and responsible parties.*

Timeline

Evaluations are sometime criticized for not producing results in a timely manner. Without an idea of when actions should occur, it is possible that activities can drag on for weeks, months, or years. By engaging stakeholders in identifying questions relevant to the users of evaluation findings, a timeline for information needs can be built in and inform the evaluation method selected.

You should develop a timeline for the evaluation planning process that considers information needs and fiscal timing of stakeholders (do stakeholders work on a January to December fiscal year, or October to September?). The plan should include a detailed timeline for resource mobilization, planning implementation, and data collection and analysis. Deciding what data will be collected and when is an important part of evaluation planning. Data collection and analysis may be dictated by the program itself, if baseline and midline data are expected. Stakeholders should be consulted in this process to ensure that the timeline is realistic and achievable. This information may change throughout evaluation planning and implementation, and should be updated accordingly.

ACTION: *Use worksheet 5.3: Timeline for Achieving Results to set realistic time points for conducting evaluation activities*

Because resources, opportunities, priorities, and activities in programs change, the evaluation plan should be considered a "living document," and revised on an ongoing basis to reflect changes over time. Your plan should be adapted to your specific evaluation needs and context, but remain flexible enough to account for changes in capacity or resources.

The Namibia Experience

MEASURE Evaluation drafted an evaluation plan/strategic information plan that will be used by the KP TWG as a living document to guide collective efforts for reporting against the current NSF and to plan for NSF targets.

The Dominican Republic Experience

The USAID mission requested a presentation with the elements of an M&E plan to integrate the information gained from the process into its existing M&E mechanisms. This presentation was used to plan additional data collection activities and standardize indicators among the mission's implementing partners.

> At this point, you have:
>
> ✓ Identified a target program
> ✓ Organized program stakeholders
> ✓ Collected background program information
> ✓ Composed a Program Impact Pathway
> ✓ Mapped ongoing and planned research activities to the PHQA model
> ✓ Identified information gaps
> ✓ Determined your evaluation needs
> ✓ Prioritized evaluation questions and discussed design issues
> ✓ Linked available indicators and/or measures to evaluation questions
> ✓ Defined roles and responsibilities of stakeholders
> ✓ Created a timeline and budget for the achievement of results

WORKSHEET 5.1. ROLES AND RESPONSIBILITIES

At this point you have identified which data sources are needed to answer your main research questions. To facilitate implementation of your evaluation plan, you should identify the inputs (particularly those that affect your budget) that are needed to carry out your evaluation activities. It is also important to identify and record who is responsible for providing these inputs, and any oversight, protocols, or procedures that should be adhered to. This information will facilitate the budgeting process.

Some important activities to consider are:

- Protocol development
- Pilot testing
- Survey development
- International Review Board applications
- Translations
- Data collection
- Data entry
- Data analysis
- Data storage
- Dissemination meetings/workshops
- Reports/manuscript preparation
- Publication costs

It may be helpful to organize the information by data source, because each source has a unique set of activities and inputs.

Activity	Inputs	Responsible	Notes
Data Source #1:			
Data Source #2:			
Data Source #3:			

WORKSHEET 5.2. BUDGET TEMPLATE

A budget template is provided; it can be modified to fit your needs. You may not need all line items presented here. The actual costs of evaluation are directly tied to the activities you have planned. Based on experience, basic M&E costs should account for 5–10 percent of the total program budget. If rigorous special studies are planned, 15–25 percent of the program budget may be needed.

Salary								
Personnel (Name)	Role/ position	Annual/ monthly salary or daily rate (C)	Year 1 time allocation in months, weeks, days (D)	Year 2 time allocation in months, weeks, days (E)	Year 3 time allocation in months, weeks, days (F)	Fringe benefits (G)	Health insurance (H)	Salary subtotal‡ Formula: [C*D + (C*1.03)*E + (C*1.07123) *F)] + G + H
	Principal Investigator (PI)/ Research Lead							
	Co-PI(s)							
	Research Assistant							
	Other Researcher(s)							
	Statistician							
	Geospatial Specialist							
	Data Manager							
	Clinical Researcher							
	Translator							
	Data Collectors (multiple)							

Salary (cont.)

Personnel (Name)	Role/ Position	Annual/ monthly salary or daily rate (C)	Year 1 time allocation in months, weeks, days (D)	Year 2 time allocation in months, weeks, days (E)	Year 3 time allocation in months, weeks, days (F)	Fringe benefits (G)	Health insurance (H)	Salary subtotal‡ Formula: [C*D + (C*1.03)*E + (C*1.07123) *F)] + G + H
	Study Implementation Manager/ Data Collection & Entry Supervisor							
	Data Entry Staff (for double data entry)							
	Data Analyst							
	Information Technology Support							
	Support Staff/ Administrative							
TOTAL								= SUM ABOVE (W)

‡The 1.035 and 1.07123 is cost of living adjustment (COLA) 3% to 4% additional salary after the first year.

International Travel

	$ Amount (B)	Quantity (C)	Travel subtotal					
Airfare		# trips	=B*C					
Lodging		# nights	=B*C					

International Travel (cont.)									
	$ Amount (B)	**Quantity (C)**	**Travel subtotal**						
Meals and incidental expenses		# days	=B*C						
Visa		# visas	=B*C						
Ground transportation			=B						
Miscellaneous			=B						
TOTAL			= SUM ABOVE (X)						
Other									
	$ Amount (B)	**Quantity (C)**	**Quantity (D)**	**Other subtotal**					
Domestic transportation for data collection (air/car rental + fuel/public transport)		# flights/ cars/trips	# days	=B* C* D					
Printing		Pages		=B*C					
Publication costs		Pages		=B*C					
Communications (phone, fax, internet)									

Other (cont.)

	$ Amount (B)	Quantity (C)	Quantity (D)	Other subtotal				
Computers		# Computers		=B*C				
Software		# Programs		=B*C				
Copies		Pages		=B*C				
Cost to present results at meetings (travel, conference, registration, per diem)								
TOTAL				= SUM ABOVE (Y)				

Meetings (Local Dissemination, etc.)

	$ Amount (B)	Quantity (C)	Meeting subtotal					
Transportation		# People	=B*C					
Venue rental		# days	=B*C					
Food and drinks		# days	=B*C					
Hotel rooms		# nights	=B*C					
Materials (flip charts, pens)								
TOTAL			= SUM ABOVE (Z)					

Total								
TOTAL COSTS			= W+X+Y+Z					
Indirect costs (organizational)								
"Indirect costs" range widely, usually 10-50% of total project costs								

WORKSHEET 5.3. TIMELINE FOR ACHIEVING RESULTS

Stakeholders may want to know when the activities you have identified in worksheet 5.1 are expected to be completed. A Gantt chart may be used to present the amount of time needed to complete activities, and depict when recurring activities will occur. Depending on your needs, time intervals can be in weeks, months, quarters, or years.

Activity	Completed or ongoing	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Data Source #1:							
Data Source #2:							
Data Source #3:							

While Gantt charts can present a lot of useful information in a small area, you may find it helpful to depict this information more visually. A timeline can convey information sequentially, allowing users to quickly see the flow of the plan and when things need to be done.

You may want to use color coding to represent activities that are completed or are underway (black), planned and funded (green) and/or planned and unfunded (red).

TIME 1   TIME 2   TIME 3   TIME 4

STEP 6. ENSURE USE FOR EVIDENCE

Existing M&E systems typically focus on data collection and reporting to higher levels, while little attention is paid to how the data can be used locally for program improvements. This workbook does not go into specifics about data-driven decision making. Resources are provided elsewhere that can be used as guidance (Appendix B). The use of the evaluation plan and information it will produce should be anticipated, cultivated, and explicitly written into the plan. Planning for the use of evaluation plan is directly tied to the identified purpose(s) of the evaluation and program priorities.

Develop a Dissemination Plan

Because stakeholders have been engaged throughout the process, there is a greater likelihood that the evaluation plan will be carried out, but it does not guarantee its use. Stakeholders are often *representatives* of interest groups, governments, nongovernmental organizations, etc., and may not include *all* end users of the evaluation plan. It is important to identify as many end users as possible, and to develop a dissemination and communication plan to ensure use at the local level. This increases the likelihood that evaluation results will be used because the evaluation will answer questions relevant to users and timeline for needed information can be built in and inform the evaluation design.

Effective dissemination relies on the end users' understanding of the goals and purpose of the evaluation plan. Messages should be clear, simple, and action-oriented. The style, content, and channels of communication (e.g., publications and reports; electronic communications; meetings and conferences; person-to-person communication; formal collaborations or information networks) should be tailored to each audience.

Once you have identified the audience(s) and activities/tools you will use to reach each audience, your stakeholders should agree on the timing (what will occur first and when) and responsible parties for carrying out dissemination activities.

Evaluating the success of your dissemination efforts is an iterative process. Once you have begun to disseminate your plan, consider how you might evaluate the effect that your dissemination strategies have on getting your plan to end users. Dissemination is not a one-time activity; rather, it is a long-term relationship with your users that will provide ongoing feedback to help you improve your message. Schedule meetings to report back and ensure that commitments are being met.

ACTION: *Use worksheet 6.1: Developing a Dissemination Plan to identify end users, channels for communication, and responsible parties to carry out dissemination activities.*

Pull It All Together

At this point, you should synthesize the information gathered thus far and create a first draft of the evaluation plan. The basic elements of an evaluation plan are these:

- Title Page
- Program Background and Evaluation Rationale
- Evaluation Question Overview
- Evaluation Design
- Methods
- Data Analysis
- Use, Dissemination, and Sharing Plan
- Costs
- Roles and Responsibilities
- Timeline

Your evaluation plan should reflect the needs and context of your program. You may also want to include information on ethics, data sharing, and authorship.

Adopt and Endorse

The evaluation plan should then be reviewed and revised with input from the stakeholder group. Once all comments and suggestions have been addressed, and a consensus has been reached, the stakeholder group should formally finalize and adopt the plan for use. The stakeholders' endorsement should be indicated in the plan through branding on the front page and an acknowledgement section.

The Ghana Experience

Between the second and third in-person meetings, the Evaluation Plan for the Ghana National Strategy for Key Populations was drafted by MEASURE Evaluation. The objective of the final two-day meeting was to obtain TWG member input to finalize the draft evaluation plan. Roles and responsibilities for action items, and for carrying out proposed data collection activities, including identifying sources of funding to address data gaps and responsible parties to follow up on action points, were agreed on.

Information obtained during the third meeting was used to revise and finalize the evaluation plan, which was submitted electronically to the TWG and presented at the Ghana National HIV and AIDS Research Conference. The evaluation plan was posted online, and printed copies were made and disseminated to stakeholders in Ghana.

The Namibia Experience

The Namibian KP TWG identified data needs tied to the next strategic plan and possible methods for filling those gaps. While the evaluation plan was designed as a living document, the end "use" will be to have the necessary data available to inform national strategic plans.

At this point, you have:

✓ Identified a target program
✓ Organized program stakeholders
✓ Collected background program information
✓ Composed a Program Impact Pathway
✓ Mapped ongoing and planned research activities to the PHQA model
✓ Identified information gaps
✓ Determined your evaluation needs
✓ Prioritized evaluation questions and discussed design issues
✓ Linked available indicators and/or measures to evaluation questions
✓ Defined roles and responsibilities for stakeholders
✓ Created a timeline and budget for achieving results
✓ Created an evaluation plan dissemination and use plan
✓ Finalized the evaluation plan and received stakeholder endorsement and adoption

WORKSHEET 6.1. DEVELOPING A DISSEMINATION PLAN

It is important to identify as many end users as possible, and to develop a dissemination and communication plan to ensure use of the evaluation plan at the local level. Once you have identified the audience(s) and activities/tools you will use to reach each audience, your stakeholders should agree on the timing of activities (what will occur first and when) and responsible parties for carrying out dissemination activities.

Audience	What information do they want?	How do they want this information? (format/tool)	When do they want this information?	Who is responsible for sharing this information?

CONCLUSION

Evaluations are conducted to obtain evidence that can inform judgements about a program's performance, to improve the effectiveness of programming, for program accountability and transparency, and to inform decisions about policies and programming, including scale up. As health programs become more tailored and targeted, driven by country needs and engagements, evaluations need to be aligned with a country's own information needs, timelines, and priorities.

Using the UNAIDS *Strategic Guidance for Evaluating HIV Prevention Programmes* (UNAIDS, 2010b) and best practices identified in evaluation policies of international organizations, such as PEPFAR and USAID (PEPFAR, 2015; USAID, 2011), it is possible to plan an evaluation programs at any level that is responsive to national and international priorities and is part of a comprehensive M&E system. This workbook operationalizes this guidance into actionable steps, and provides companion worksheets to facilitate the process. The development of an evaluation plan in cooperation with a group of stakeholders fosters collaboration, a sense of shared purpose, transparency, and ensures that stakeholders are on the same page about the purpose, use, and users of evaluation results. Application of this workbook will yield a complete evaluation plan that is wholly owned by the stakeholders who participate in the planning process.

REFERENCES

Better Evaluation. (n.d.). Evaluation plan. Retrieved from
http://www.betterevaluation.org/en/evaluation-options/evaluation_planning_templates.

Creese, A., Floyd, K., Alban, A., & Guinness, L. (2002). Cost-effectiveness of HIV/AIDS
interventions in Africa: a systematic review of the evidence. *The Lancet*, 359(9318), 1635–1643.
Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12020523.

Frankel, N. & Gage, A. (2007, rev. 2016). M&E fundamentals: A self-guided minicourse. Chapel
Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from
https://www.measureevaluation.org/resources/publications/ms-07-20-en.

Ghana AIDS Commission (GAC). (2013). Evaluation plan for the Ghana national strategy for key
populations. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina.
Retrieved from https://www.measureevaluation.org/resources/publications/sr-13-75.

Ghana AIDS Commission (GAC). (2010). Ghana national strategic plan for HIV and AIDS 2011-
2015. Accra, Ghana: GAC.

Ghana AIDS Commission (GAC). (2011a). Integrated biological and behavioural surveillance
survey (IBBSS) of female sex workers and behavioural surveillance of clients of sex workers in
Ghana. Accra, Ghana: GAC.

Ghana AIDS Commission (GAC). (2011b). MARP operational plan framework 2011–2015. Accra,
Ghana: GAC.

Ghana AIDS Commission (GAC). (2011c). National strategic plan for most at risk populations
2011–2015. Accra, Ghana: FHI 360. Retrieved from http://pdf.usaid.gov/pdf_docs/pnaed197.pdf.

Habicht, J. P., Victora, C. G., & Vaughan, J. P. (1999). Evaluation designs for adequacy, plausibility
and probability of public health programme performance and impact. *International Journal of
Epidemiology*, 28, 10–18. Retrieved from
https://www.measureevaluation.org/resources/training/capacity-building-resources/m-e-of-hiv-
aids-programs-in-india-english/session-5-gis-and-research-evaluation/research-
evaluation/Int.%20J.%20Epidemiol.-1999-Habicht-10-8.pdf.

Handley, K., Boerma, T., Victora, C., & Evans, T. G. (2015). An inflection point for country health
data. *The Lancet Global Health*, 3(8), e437–e438. Retrieved from
http://www.thelancet.com/journals/langlo/article/PIIS2214-109X(15)00067-4/fulltext.

Hart, L., Iskarpatyoti, B. S., Mandal, M., & Thomas, J. C. (2016). *Evaluating structural interventions –
guidance for HIV prevention programs* (tr-16-139). Chapel Hill, NC, USA: MEASURE Evaluation,
University of North Carolina. Retrieved from
https://www.measureevaluation.org/resources/publications/tr-16-139.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2010a). Combination HIV
prevention: tailoring and coordinating biomedical, behavioural and structural strategies to reduce
new HIV infections. Geneva, Switzerland: UNAIDS. Retrieved from
http://www.unaids.org/sites/default/files/media_asset/JC2007_Combination_Prevention_paper_
en_0.pdf.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2015). Fast-tracking combination prevention: towards reducing new HIV infections to fewer than 500,000 by 2020. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS (UNAIDS). Retrieved from http://www.unaids.org/sites/default/files/media_asset/20151019_JC2766_Fast-tracking_combination_prevention.pdf.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2014a). Fast-track: Ending the AIDS epidemic by 2030. Geneva, Switzerland: UNAIDS. Retrieved from http://www.unaids.org/sites/default/files/media_asset/JC2686_WAD2014report_en.pdf.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2014b). 90-90-90: An ambitious treatment target to help end the AIDS epidemic. Geneva, Switzerland: UNAIDS. Retrieved from http://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2007). Practical guidelines for intensifying HIV prevention: Towards universal access. Geneva, Switzerland: UNAIDS. Retrieved from http://data.unaids.org/pub/manual/2007/20070306_prevention_guidelines_towards_universal_access_en.pdf.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2010b). Strategic guidance for evaluating HIV prevention programmes. Geneva, Switzerland: UNAIDS. Retrieved from http://www.unaids.org/sites/default/files/sub_landing/files/12_7_MERG_Guidance_Evaluating%20HIV_PreventionProgrammes.pdf.

MEASURE Evaluation. (2011). Tools for data demand and use in the health sector: Stakeholder engagement tool (ms-11-46e). Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/ms-11-46.

MEASURE Evaluation. (2016). Types of evaluations conducted by MEASURE Evaluation (fs-15-158). Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/fs-15-158.

MEASURE Evaluation. (2017). Complexity-aware methods (fs-17-217). Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/fs-17-217

Reynolds, H. W., Atuahene, K., Sutherland, E., Amenyah, R., Kwao, I. D., & Larbi, E. T. (2014). Development of a nationally coordinated evaluation plan for the Ghana national strategy for key populations. *Journal of AIDS & Clinical Research*, 5(12), 389. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4479146/.

Sachs, J. D. (2012). From millennium development goals to sustainable development goals. *The Lancet*, 379(9832), 2206-2211. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22682467.

Segone, M. (2009). Enhancing evidence-based policy-making through country-led monitoring and evaluation systems. In Segone, M. (Ed.), *Country-led monitoring and evaluation systems* (pp. 17–31). Geneva, Switzerland: United Nations Children's Fund (UNICEF). Retrieved from https://evalpartners.org/sites/default/files/images/Country_ledMEsystems.pdf.

Skiles, M. P., Hattori, A., & Curtis, S. L. (2014). Impact evaluations of large-scale public health interventions: experiences from the field. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/wp-14-157.

United States Agency for International Development (USAID). (2017, August 22). Data-driven decision making for HIV/AIDS. Retrieved from https://www.usaid.gov/what-we-do/global-health/hiv-and-aids/technical-areas/strategic-information.

United States Agency for International Development (USAID). (2011). Evaluation: Learning from experience, USAID evaluation policy. Washington, DC, USA: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/Pdacq800.pdf.

United States Agency for International Development (USAID). (2014b). PEPFAR evaluation standards of practice. Washington, DC, USA: USAID. Retrieved from https://www.pepfar.gov/documents/organization/221324.pdf.

United States Agency for International Development (USAID). (2017). Strategy for accelerating HIV/AIDS epidemic control (2017–2020). Washington, DC, USA: USAID. Retrieved from https://www.pepfar.gov/documents/organization/274400.pdf.

United States Agency for International Development (USAID). (2016). Sustainable HIV epidemic control: PEPFAR position paper. Washington, DC, USA: USAID. Retrieved from https://www.pepfar.gov/documents/organization/264884.pdf.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2014a). 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.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2015). Evaluation standards of practice, version 2.0. Washington, DC, USA: PEPFAR. Retrieved from https://www.pepfar.gov/documents/organization/247074.pdf.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2014b). PEPFAR evaluation standards of practice. Washington, DC, USA: PEPFAR. Retrieved from https://www.pepfar.gov/documents/organization/221324.pdf.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2017). Strategy for accelerating HIV/AIDS epidemic control (2017–2020). Washington, DC, USA: PEPFAR. Retrieved from https://www.pepfar.gov/documents/organization/274400.pdf.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2016). Sustainable HIV epidemic control: PEPFAR position paper. Washington, DC, USA: PEPFAR. Retrieved from https://www.pepfar.gov/documents/organization/264884.pdf.

United States President's Emergency Plan for AIDS Relief (PEPFAR). (2011). Technical guidance on combination HIV prevention. Washington, DC, USA: PEPFAR. Retrieved from https://www.pepfar.gov/documents/organization/164010.pdf.

Vermund, S. H. & Hayes, R. J. (2013). Combination prevention: New hope for stopping the epidemic. *Current HIV/AIDS Reports*, 10(2), 169–186. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23456730.

Walker, D. (2003). Cost and cost-effectiveness of HIV/AIDS prevention strategies in developing countries: Is there an evidence base? *Health Policy and Planning*, 18(1), 4–17. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12582104.

World Health Organization (WHO). (2016). World health statistics 2016: monitoring health for the SDGs, sustainable development goals. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/gho/publications/world_health_statistics/2016/EN_WHS2016_TOC.pdf?ua=1

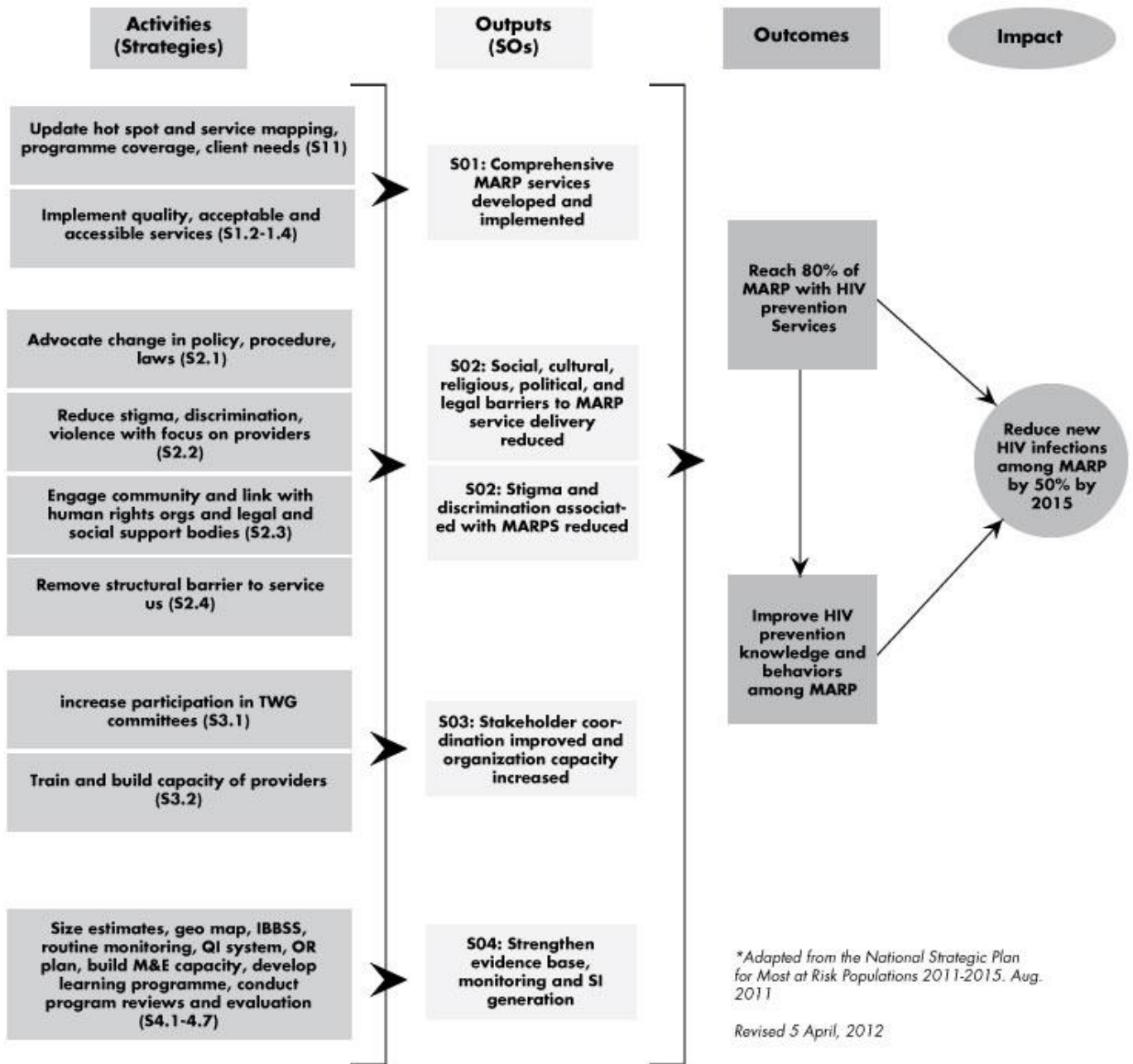# APPENDIX A. SAMPLE EVALUATION PLAN COMPONENTS

A1. Sample PIP

Ghana MARP Program Impact Pathway (Draft)

Activities	Outputs	Outcomes (short-term)	Outcomes (long-term)	Impact
1.1. Update existing and generate new hot spot and services mapping, coverage, and client need information 1.2-1.4 Implement a package of high quality, acceptable, and accessible HIV prevention services (1.2); treatment and care services (1.3); mental health and psychosocial support services (1.4) for each MARP subgroup. 2.1 Advocate for changes in HIV policies, procedures, and laws that may impede the HIV response among MARPs 2.2 Reduce stigma, discrimination, and violence experienced by MARPs (focus on service providers) 2.3. Engage with the broader community to establish linkages and coordination with human rights organizations and community legal and social support bodies 2.4 Remove structural barriers to the use of services and programs by MARPs 3.1 Increase the level of participation and representation of implementers and MARP representatives in TWG subcommittees at the decentralized level 3.2 Support training and capacity building of MARP service providers SO4: Strengthen evidence base and MARP monitoring systems and promote generation of SI to improve MARP programs	Number reached by interventions (prevention; treatment, care & support; psycho-social support) Number referred for services Total number of referrals Number of condoms and lubricants distributed Number of condom service outlets Number of drop in centers Number of peer groups formed Number of heath care workers trained as FSW and MSM friendly Number of peer educators recruited and trained	**Knowledge**: Correctly identify ways of preventing the sexual transmission of HIV and who reject major misconceptions about HIV transmission **Behavior**: Increased condom use with most recent client (FSW) Increased condom use with every client last month (FSW) Increased use of condom with non-paying partner (FSW) Reduce the number of male partners in last six months (MSM) Condom used at last insertive/receptive anal sex with male (MSM) Consistent condom use during anal sex with male in last three months (MSM) Increase HIV test and know results Increased use of care services & antiretrovirals	Provide evidence based prevention, treatment, care and support services to 80% of all identified MARP groups by 2015	Reduction in new HIV infections by 50% by 2015 (% MARP who are HIV+)

A2. Sample PIP Gap Analysis

[4]From the National Strategic Plan for Most at Risk Populations 2011-2015 (August, 2011), pages 35-7

Ghana MARP Illustrated Program Impact Pathway

Activities (Strategies)

Update hot spot and service mapping, programme coverage, client needs (S11)

Implement quality, acceptable and accessible services (S1.2-1.4)

Advocate change in policy, procedure, laws (S2.1)

Reduce stigma, discrimination, violence with focus on providers (S2.2)

Engage community and link with human rights orgs and legal and social support bodies (S2.3)

Remove structural barrier to service us (S2.4)

increase participation in TWG committees (S3.1)

Train and build capacity of providers (S3.2)

Size estimates, geo map, IBBSS, routine monitoring, QI system, OR plan, build M&E capacity, develop learning programme, conduct program reviews and evaluation (S4.1-4.7)

Outputs (SOs)

SO1: Comprehensive MARP services developed and implemented

SO2: Social, cultural, religious, political, and legal barriers to MARP service delivery reduced

SO2: Stigma and discrimination associated with MARPS reduced

SO3: Stakeholder coordination improved and organization capacity increased

SO4: Strengthen evidence base, monitoring and SI generation

Outcomes

Reach 80% of MARP with HIV prevention Services

Improve HIV prevention knowledge and behaviors among MARP

Impact

Reduce new HIV infections among MARP by 50% by 2015

*Adapted from the National Strategic Plan for Most at Risk Populations 2011-2015. Aug. 2011

Revised 5 April, 2012

A3. Sample HIV Program Mapped to the PHQA
Ongoing M&E Activities in Ghana Mapped to the PHQA Model

PHQA step	Ongoing M&E activity	Anticipated results	Questions, gaps, additional information needs
1. Know your epidemic: What is the size and nature of the problem?	IBBSS with MSM and FSW (2011)	• Size estimates • Denominator for coverage estimates • Populations defined • National estimate of HIV prevalence • Behavioral data	• Will size estimates yield sub-national estimates? • Will the IBBSS methods used be replicable over time? • Will study methods yield HIV incidence? • Population size estimates could be presented on a national map showing distribution and numbers of MSM and FSW.
2. Determinants: What are the contributing factors?	• IBBSS with MSM and FSW (2011) • Project SEARCH study: KAP, risk behaviors, HIV needs of young (18-20 years) FSW (2011-12) • Project SEARCH study: Transactional sex among female post-secondary education students in Kumasi (2011-12)	• Measures of direct determinants: exposure, infectiousness, biologic susceptibility • Understanding of the social determinants	• Identify what other formative research and needs assessments have been conducted (to inform programs) and with what methods and level of rigor. • Decide what formative or qualitative studies are needed to help interpret or fill in IBBSS results, such as providing a deeper understanding about psychosocial, economic, and other contextual determinants, facilitators, and barriers to health-seeking behaviors and their relation to HIV risk perceptions. • HIV and behavioral data from IBBSS could be combined in models to predict trends about epidemic (and estimate changes under different assumptions about behavior change and health care use and treatment).
3. Know your response: Identify which interventions can work	MARP Strategy 2011-2015 (2011) and Operational Plan	• Guidelines and Operational Plan • Defined comprehensive programs informed by international evidence and guidelines • Defined minimum packages of services with which each MARP should be reached	• Service availability mapping (combine with map of populations size estimates). • Complete the logic model: Do the planned interventions appear to be complete in terms of theoretically leading to improved outcomes? Are the key activities aligned with routinely collected program output measures? Are there gaps? • Targets for impact, outcome, and coverage indicators • Output indicators and standardized forms for data collection, reporting, and aggregating • Priority operations research questions

PHQA step	Ongoing M&E activity	Anticipated results	Questions, gaps, additional information needs
4. Input monitoring: What interventions and resources are needed?	• GOALS exercise (resource analysis) (2011) • MARP strategy 2011-2015 (2011)	Key activities and resources needed are defined	Present outcome of the GOALS exercise. What is the funding gap between what is planned in the completed logic model and what can be funded through current obligations (PEPFAR, Global Fund, etc.)? What is the implication for planned activities and targets?
5. Quality monitoring: What activities are we doing? Are we doing them right?	• MARP Strategy 2011-2015 (2011) • MARP Operational Plan (2011-draft) • Program process monitoring (by implementing partners)	• Key activities defined • Routine program monitoring indicators and data collection, analysis, reporting, and use systems • Quality standards and tools	• Are there doubts about data quality? Are these data analyzed and used to improve programs? Do program monitoring data need to be harmonized? • Conduct service quality assessments and client satisfaction surveys/assessments.
6. Monitoring outputs and coverage: Are we implementing the program as planned?	• Routine program monitoring • Methods to avoid double counting	• Routine program data and aggregation on regular basis • Combine with population size estimates for coverage • Define "reached" by program	• Assess coverage indicators and trends. • Operational definition of person "reached" by program. • Conduct process evaluation: Are services available in the right place and are we reaching the target population (geographic and individual coverage)? Are services acceptable to clients? Are we implementing our services as planned? What is the capacity of programs to provide services? Are programs linking with other services? Are current program activities of sufficient quality, coverage, and uptake to reach 80% of MARP?
7-8. Outcome and impact monitoring and evaluation: Are interventions making a difference?	IBBSS with MSM and FSW (planned)	Trends in outcomes (knowledge, attitudes, practices, and behavior) and impact (sexually transmitted infection and HIV prevalence and incidence) among target populations	• Are outcomes positively changing as desired: behaviors, HIV incidence, etc.? • Are changes observed in outcomes likely the results of the program? (Outcome evaluation; study design needed) What program components are contributing the most to outcomes? Are they cost-effective? • What is the optimal mix of services? Which combination of services best affects changes in outcomes? • Is the program having an effect on HIV in the general population?

A4. Sample Timeline

Activity	Completed or on-going	Oct-Dec 2012	Jan-Mar 2013	Apr-June 2013	July-Sept 2013	Oct-Dec 2013	Jan-Mar 2014	Apr-June 2014	July-Sept 2014	Oct-Dec 2014	Jan-Mar 2015
2011 IBBSS	X										
Initial costing study	X										
Process Evaluation 2012	X										
Performance (quality) evaluation 2013	X										
Output monitoring indicators report monthly to GAC	X	X	X	X	X	X	X	X			
Contextual information collected semi-annually				X		X		X			
Resource mobilization for IBBSS, both MSM and FSW						X					
IBBSS planning							X				
IBBSS implementation								X			
IBBSS data analysis									X		
Resource mobilization for quality and cost assessment					X						
Planning for quality and cost assessment						X	X				
Data collection for quality and cost assessment								X			
Data analysis for quality and cost assessment									X		
Plausibility analysis and report writing									X	X	

APPENDIX B. ADDITIONAL RESOURCES

Evaluation Design

- Reynolds, H. W. & Guest, G. S. (2014). Designing research. In Guest, G.S. & Namey, E.E. (Eds.) *Public health research methods* (pp. 33-68). Thousand Oaks, CA, USA: SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/public-health-research-methods/book237897.
- Centers for Disease Control and Prevention (CDC), National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Division of Nutrition, Physical Activity, and Obesity. (2011). *Developing an effective evaluation plan.* Atlanta, GA, USA: CDC. Retrieved from https://www.cdc.gov/obesity/downloads/cdc-evaluation-workbook-508.pdf.
- USAID. (n.d.). USAID Evaluation Toolkit. Retrieved from https://usaidlearninglab.org/evaluation.
- Tayabas, L. M. T., León, T. C., & Espino, J. M. (2014). Qualitative evaluation: A critical and interpretative complementary approach to improve health programs and services. *International Journal of Qualitative Studies on Health and Well-Being*, 9(1), 24417. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142224/.
- Victora, C. G., Habicht, J.-P., & Bryce, J. Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health,* 94 (3), 400-405. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448265/.
- Frankel, N. & Gage, A. (2007, rev. 2016). M&E fundamentals: A self-guided minicourse. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/ms-07-20-en.

Sampling and Data Collection

- MEASURE Evaluation. (2015). Priorities for local AIDS control efforts (PLACE). Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/tools/hiv-aids/place.
- UCSF Global Health Sciences. (n.d.). Integrated HIV bio-behavioral surveillance toolbox. Retrieved from http://globalhealthsciences.ucsf.edu/resources/integrated-hiv-bio-behavioral-surveillance-toolbox.
- The DHS Program. (n.d.). Retrieved from https://dhsprogram.com/.
- Weir, S., et al. (2012). *Operational guidelines for monitoring and evaluation of HIV programmes for sex workers, men who have sex with men, and transgender people* (Volume 1 & Volume 2). Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/tools/hiv-aids/operational-guidelines-for-monitoring-and-evaluation-of-hiv-programmes-for-sex-workers-men-who-have-sex-with-men-and-transgender-people.
- Sorensen, H. T., Sabroe, S., & Olsen, J. (1996). A framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology*, 25(2), 435–442. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9119571.

- Rutherford, G.W., McFarland, W., Spindler, H., White, K., Patel, S.V., Aberle-Grasse, J., … Stoneburner, R.L. (2010). Public health triangulation: approach and application to synthesizing data to understand national and local HIV epidemics. *BMC Public Health,* 10:447. Retrieved from https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-10-447.
- MEASURE Evaluation. (2015). Evaluation FAQ: What sample size do I need for an impact evaluation? Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/fs-15-157.
- Lance, P. & Hattori, A. (2016). Sampling and evaluation—a guide to sampling for program impact evaluation. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/ms-16-112.

Data Analysis

- MEASURE Evaluation. (2011). Introduction to basic data analysis and interpretation for health programs: a training tool kit. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from http://www.cpc.unc.edu/measure/resources/training/materials/basic-data-analysis-for-health-programs.
- Ulin, P. R., Robinson, E. T., and Tolley, E. E. (2005). Qualitative methods: A field guide for applied research in sexual and reproductive health. San Francisco, CA, USA: Jossey-Bass. https://www.amazon.com/Qualitative-Methods-Public-Health-Research/dp/0787976342.
- Lance, P., Guilkey, D., Hattori, A., Angeles, G. (2014). How do we know if a program made a difference? A guide to statistical methods for program impact evaluation. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/ms-14-87-en.

Data Use

- MEASURE Evaluation. (2013). Data demand and use toolkit. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from http://www.cpc.unc.edu/measure/resources/tools/data-demand-use/data-demand-and-use-strategies-and-tools.html
MEASURE Evaluation. (2014). Complementing routine data with qualitative data to understand the why behind program data. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/training/capacity-building-resources/ddu.
MEASURE Evaluation. (2013). Building leadership for data demand and use: A facilitator's guide. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from https://www.measureevaluation.org/resources/publications/ms-13-59.
- MEASURE Evaluation. (n.d.). Data quality assurance tools. https://www.measureevaluation.org/resources/tools/health-information-systems/data-quality-assurance-tools/data-quality-assurance-tools.
- Global Health eLearning Center. (2010). Data quality. Retrieved from https://www.globalhealthlearning.org/course/data-quality.