

**Operational Guidance:** 

# SUMMATIVE PROGRAM EVALUATION

(Annex to Basic Toolkit: Adaptive Learning in the Programming Cycle)

MOMENTUM Country and Global Leadership





MOMENTUM works alongside governments, local and international private and civil society organizations, and other stakeholders to accelerate improvements in maternal, newborn, and child health services. Building on existing evidence and experience implementing global health programs and interventions, we help foster new ideas, partnerships, and approaches and strengthen the resiliency of health systems.

This operational guidance is made possible by the generous support of the American people through the U.S. Agency for International Development (USAID) under the terms of the Cooperative Agreement #7200AA20CA00002, led by Jhpiego and partners. The contents are the responsibility of MOMENTUM Country and Global Leadership and do not necessarily reflect the views of USAID or the United States Government.

Cover photo: Kate Holt/MCSP

#### Suggested Citation

Tappis H and Bryce E. 2022. *Operational Guidance: Summative Program Evaluation (Annex to Basic Toolkit: Adaptive Learning in the Programming Cycle)*. Washington DC, USAID MOMENTUM. Country and Global Leadership.

# **TABLE OF CONTENTS**

Acknowledgements
Abbreviations
Introduction4
Preparing for a Summative Program Evaluation5
Defining the evaluation purpose and scope5
Selecting an evaluation approach5
Developing an evaluation plan6
Description of Selected Evaluation Approaches7
Theory-based evaluation approaches7
Effectiveness/impact evaluation approaches10
Practical Considerations14
Theory-based evaluation approaches14
Potential challenges and considerations17
Effectiveness/impact evaluation approaches18
Modifying an Evaluation Approach22
Additional Resources
Appendix A. Evaluation Concept Note Template
Part 1: Concept note
Part 2: Additional information needed for non-human subjects research/non-research/human subjects research determination

# ACKNOWLEDGEMENTS

MOMENTUM Country and Global Leadership is part of a suite of innovative awards funded by the U.S. Agency for International Development (USAID) to holistically improve voluntary family planning (FP) and maternal and child health (MCH) in partner countries around the world. The project focuses on technical and capacity development assistance to ministries of health and other country partners to improve outcomes.

We would like to acknowledge the valuable feedback from Elana Dhuse, Young-Mi Kim, Elizabeth Oliveras, Jim Ricca, Aleefia Somji, Soumya Alva, and Lara Vaz, which greatly improved the final version of this document.

# **ABBREVIATIONS**

C-I-A-M-O	Context-intervention-actor-mechanism-outcome
C-I-M-O	Context-intervention-mechanism-outcome
C-M-O	Context-mechanism-outcome (for realist evaluation)
cRCT	Cluster randomized controlled trial
DiD	Difference in difference
ITS	Interrupted time series
MEL	Monitoring, evaluation, and learning
МОН	Ministry of Health
PBF	Performance-based financing
PII	Personal identifying information
QUIP	Qualitative Impact Assessment Protocol
RCT	Randomized controlled trial
SDT	Self-determination theory
SW-cRCT	Stepped-wedge cluster randomized controlled trial
USAID	United States Agency for International Development

## **INTRODUCTION**

Projects last for a specified period of time, but implementers and donors often expect that such a timebound investment will yield sustained results beyond the project time limits and/or geographic area. With this in mind, knowing **how** results were achieved and/or **why** a strategy was less successful in some contexts than in others can be even more important than project performance itself. Projects that have a long enough timeline to implement a learning agenda and/or have sufficient time to accumulate significant operational evidence about how a programmatic approach is functioning should strongly consider including a summative program evaluation activity in their plans. Unlike formative assessments conducted to inform the design of

an intervention or process evaluations conducted to assess how implementation is progressing and inform mid-course corrections, summative program evaluations involve overall assessments of an intervention or program, for example, if objectives were met or how and why results varied across implementation sites. A summative program evaluation pulls together programmatic evidence for use by program implementers and other stakeholders. Within MOMENTUM Country and Global Leadership, we plan to conduct internal summative program evaluations<sup>1</sup> of country buy-ins with substantial technical scopes and implementation time periods of at least two sequential years. Evaluation plans will be co-created with the donor, Ministry of Health (MOH), and other local stakeholders.

Summative program evaluations should be designed with a level of rigor appropriate to the intervention(s), and consider the available resources, operational timeframe, and geographic scope of activities. The evaluation approaches described here Including an analysis in the MEL plan of the need for evaluations during the project (*tied to some threshold or key decision*) and at the end of the project (*either for decisions or to capture learning*) lays the foundation for allocating sufficient evaluation resources and planning in a way that allows for the use of the best methods for quality evaluation.

United States Agency for International Development (<u>USAID</u>) <u>Project Starter</u>, <u>Performance Evaluation Designs</u>

will be used less frequently than the monitoring approaches described in the companion <u>MOMENTUM Basic</u> <u>Toolkit for Adaptive Learning</u>. The evaluation approaches would likely be used toward the end of planned activities and would be intended to generate consensus interpretations of results and recommendations for action by stakeholders in future projects or initiatives.

The purpose of this document is to provide guidance on how to select an approach or approaches for a summative program evaluation. This guidance outlines how to prepare for a summative program evaluation, introduces two categories of approaches and examples of approaches under each, outlines practical considerations for implementation and concludes with guidance on modifying the approaches introduced. The intended users of this guidance are program managers, technical advisors, and monitoring, evaluation, and learning (MEL) staff responsible for considering, developing, contributing to, and executing such evaluations.

<sup>&</sup>lt;sup>1</sup> Per USAID terminology, an external evaluation is commissioned by USAID and must be managed by the Operating Unit's Program Office, while an internal evaluation is one commissioned and managed by an implementing partner.

This document is arranged in the following sections:

- Preparing for a Summative Program Evaluation
- Description of Selected Evaluation Approaches:
  - Theory-based evaluation approachesTheory-based evaluation approaches
  - Effectiveness/impact evaluation approaches
- Practical Considerations:
  - Theory-based evaluation approaches
  - Effectiveness/impact evaluation approaches
- Modifying an Evaluation Approach
- Additional Resources

# PREPARING FOR A SUMMATIVE PROGRAM EVALUATION

### DEFINING THE EVALUATION PURPOSE AND SCOPE

A summative program evaluation should only be undertaken if it has a clearly identified aim, and when it is likely to produce meaningful findings based on appropriate methodology and valid and reliable data (to be explained further below). The availability of resources and timing of decisions that will need to be made about the program or policy under investigation should be taken into consideration. The process of defining the evaluation purpose and scope (whether it should focus on an entire program/initiative, a specific program objective, or a specific intervention strategy) should engage multiple stakeholders. The evaluation purpose should be to answer a clearly defined overarching research question.

Illustrative summative program evaluation questions:

- How effective is the program, strategy, or intervention?
- What is the program's contribution to changes observed?
- Does a strategy or intervention work, for whom, in what contexts, and how?

### SELECTING AN EVALUATION APPROACH

Ideally, a summative program evaluation would answer all of the questions above. However, such complex evaluations are rarely feasible outside of well-resourced research investments. Although there may be some overlap, each evaluation approach requires different study designs, skill sets, and levels of resources.

To determine whether you should conduct a summative program evaluation, consider the following criteria:

- Is there stakeholder demand for an evaluation? What are the questions they would like to see answered? How will evaluation findings be used?
- Will an evaluation help influence decisions about *if* an intervention should be continued or scaled up? Will the intervention have been implemented for long enough, or with enough fidelity, to observe effects?

- Will an evaluation help inform decisions about *how* an intervention should be resourced or implemented? Will the intervention have been implemented for long enough to observe effects and understand factors affecting implementation?
- Is there enough time to plan and conduct an evaluation that answers the questions of interest? Are there sufficient resources (e.g., funding, data sources) to conduct a high-quality evaluation?

Selecting a summative program evaluation approach is likely to be an iterative process, involving consultation with USAID Missions, MOHs, program partners, technical experts, and other stakeholders. These conversations should begin early in the project design process, so that work plans and MEL plans can be developed accordingly.

### **DEVELOPING AN EVALUATION PLAN**

As noted above, planning for a summative program evaluation should ideally start during the project or activity design stage, and be co-created alongside the intervention strategy and other components of the MEL plan. This approach will allow teams to maximize use of data collected over the life of the project, determine the extent to which evaluation questions can be answered through these data sources, and budget for any additional data collection, analysis, and synthesis at the midterm or end of the project.

Preparing a concept note that outlines the proposed evaluation questions, scope, methods, and timeline for data collection and analysis can help facilitate stakeholder engagement in evaluation design as well as internal planning. If an implementing partner intends to conduct an internal evaluation, this should be explained in the project MEL plan. The *Evaluation Plan* section of the MEL plan should include information on the evaluation purpose, evaluation type (performance or impact evaluation), possible evaluation questions, estimated budget, planned start date, and estimated completion date.

USAID How-To Note: Activity MEL plan

See Appendix A. Evaluation Concept Note Templatefor an adaptable evaluation concept note template.

Evaluations should be formally planned and managed as a discrete activity, with management arrangements, roles, and responsibilities clearly described from the beginning of the process and appropriate time and resources allocated. Consider:

- What kinds of expertise, either within the project or through partnerships, is required to execute the selected evaluation approach with the complexity and rigor needed to answer these questions?
- What time and resource requirements are needed for evaluation design, planning, data collection, analysis, and dissemination?

Evaluation plans and budgets should be reviewed and refined as the projects evolve, and MEL plans updated accordingly. As you develop and refine evaluation questions, be realistic about the expertise and time required to obtain answers that will inform future program planning or policy decisions.

Any evaluation that collects, accesses, or uses data from or about human beings must be classified as either human subjects research, not human subjects research, or not research by appropriate authorities (refer to organization- or project-specific policies to identify your Ethical Review Committee or Institutional Review Board of record). If it is determined to be human subjects research, a detailed study protocol must be developed, reviewed, and approved by an in-country institutional review board before starting any data collection. Detailed protocol development should begin as soon as the purpose and scope of the evaluation are defined and be completed as soon as there is sufficient information to do so.

### **DESCRIPTION OF SELECTED EVALUATION APPROACHES**

There are many ways of classifying evaluation approaches.<sup>2</sup> For the purposes of this guidance, we have grouped approaches into two categories: **theory-based evaluation** and **effectiveness/impact evaluation**. These are not mutually exclusive. Examples of where these approaches can overlap, or be implemented concurrently, are presented in the *Modifying an Evaluation Approach* section.

### **THEORY-BASED EVALUATION APPROACHES**

Theory-based evaluation approaches start by clarifying the program theory—or how program activities are believed to cause or contribute to outcomes and impacts. The evaluation is designed to test the extent to which the theory holds true. The task of the evaluation team is to produce a plausible case, with evidence, to support or refute the theory.

### CONTRIBUTION ANALYSIS

# WHAT WAS THE PROGRAM'S CONTRIBUTION TO OBSERVED CHANGES?

Theory-based evaluations seek to explain why programs work and are used when the aim is to learn so as to improve success or replicate programs elsewhere; theory-based evaluations use an explicit theory of change to draw conclusions about whether and how an intervention contributed to observed results.

Contribution analysis is a methodology developed by John Mayne in the early 2000s to identify the contribution a development intervention made to a change or set of changes. It is designed to be used alongside theories of change that explicitly describe how change is, or was, supposed to happen. The aim is to produce a plausible, evidence-based narrative that a reasonable person would be likely to agree with.

<sup>&</sup>lt;sup>2</sup> USAID's Evaluation Policy differentiates between **performance evaluations** and **impact evaluations**. It notes that there is no commonly accepted taxonomy of performance evaluations on which all evaluators would agree, but for USAID these may address descriptive, normative, or cause-and-effect related questions, but generally do not involve a counterfactual comparison group. Impact evaluations, in contrast, are defined as "an evaluation that involves a structured test of one or more hypotheses underlying a program or project intervention that involves a comparison between those affected by the intervention and a comparable group or area that was not affected by the intervention."

Through gathering evidence, evaluators develop a plausible contribution narrative. A narrative is considered to be plausible when these four different conditions are met:

- 1. The development intervention is based on a sound theory of change, accompanied by agreed upon and plausible assumptions, that explains how the intervention sought to bring about any desired changes.
- 2. The activities of the development intervention were implemented properly (implemented as planned, with a reasonable level of consistency across sites).
- 3. There is adequate evidence showing that change occurred at each level of the theory of change.
- 4. The relative contribution of external factors or other development interventions can be dismissed or demonstrated.

The data sources should be informed by the theory of change in order to collect information to examine the conditions described above. Sources of evidence may include data generated through performance monitoring, context monitoring, or adaptive learning activities. It may also require additional collection of knowledge and insights from project staff and stakeholders.

### RESOURCES

- <u>MOMENTUM Knowledge Accelerator. Guide to complexity-aware monitoring approaches for</u> <u>MOMENTUM projects</u>
- Contribution Analysis: An approach to exploring cause and effect, ILAC methodological brief

### **EXAMPLES OF CONTRIBUTION ANALYSIS**

- <u>Contribution of the Alive & Thrive-UNICEF advocacy efforts to improve infant and young child feeding</u> policies in Southeast Asia
- Translating the International Code of Marketing of Breast-milk Substitutes into national measures in nine countries
- Using Contribution Analysis to Evaluate Competency-Based Medical Education Programs: It's All About Rigor in Thinking
- <u>Maternal and Child Survival Program Burma's impact on strengthening the health workforce for a better</u> tomorrow: Results from a contribution analysis
- An Analysis of Maternal and Child Survival Program Contributions to Expanding Access to and Uptake of Quality Family Planning Services in Five States of India
- <u>Strengthening Human Resources for Health Capacity in the Republic of Rwanda under the President's</u> Emergency Plan for AIDS Relief Evaluation Protocol: Summary Excerpt

### **REALIST EVALUATION**

#### WHAT WORKS, FOR WHOM, TO WHAT EXTENT, IN WHAT CONTEXTS, AND HOW?

The realist evaluation approach was first developed by Pawson and Tilley (1997) to improve understanding of how and why different projects and programs work in different contexts. Underlying all realist evaluations is the assumption that projects and programs work under certain conditions and are heavily influenced by the way that different stakeholders respond to the project or program. The task of the realist evaluation team is to systematically examine how context affects the way people respond to resources provided by programs in a complex environment, and how this influences program implementation and outcomes. This is done by developing, testing, and refining program theory in the form of "context-mechanism-outcome" (C-M-O)

configurations. These configurations represent hypotheses on how contexts interact with mechanisms (under which contexts do certain mechanisms occur or not) to generate outcomes. In some cases, authors add explanatory factors to the C-M-O configuration, such as "actors" or "intervention context," creating "C-I-M-O" or "C-I-A-M-O" configurations (see Table 1).

Components of realist program theory	C-M-O configuration examples
The <b>context</b> in which a program operates—including the socioeconomic and political environment, organizational context, local history, and culture—can "spark" mechanisms, which generate outcomes. <b>Mechanisms</b> are the causal forces, processes, or interactions that generate change within an intervention—including the choices, reasoning, and decisions that people make as a result of the resources provided by the program. [Mechanisms are not program activities/interventions. Interventions create opportunities upon which people may choose to act. Mechanisms are the forces or powers that explain how and why an intervention leads to change.] <b>Outcomes</b> are the intended and unintended short-, medium-, and long-term changes resulting from an intervention.	<ul> <li>C-M-O: In contexts where people are aware of their rights to provide feedback and perceive that they will be heard, not penalized, and if the means to provide feedback are accessible [C], it will trigger a sense of trust and confidence in the system [M], translating into exercise of agency for greater use of feedback chains [O].</li> <li>C-I-M-O: In two districts, which have a pool of malnourished children and identified causes of malnutrition that include low socioeconomic status and a lack of hygiene knowledge [C]; where the nutritional assistants are friendly, smiling, understanding, and congratulate/applaud the mothers [I], the mothers feel well received and not discriminated against and that they are considered equal [M], which in turn results in high program attendance [O].</li> <li>C-I-A-M-O: Health care delivery is conducted by health extension workers at health posts and at home [C]; where now Expanded Program on Immunization and family planning services are offered at the 45-day postnatal check [I], health extension workers [A] perceive a reduced work burden due to Expanded Program on Immunization/family planning service integration [M] and therefore provide integrated services [O].</li> </ul>

### TABLE 1. REALIST PROGRAM THEORY EXAMPLES

A wide range of quantitative and qualitative data collection and analysis methods can be used in realist evaluations, with decisions about sampling and data collection methods driven by consideration of who researchers need to talk to in order to test their program theory, and what the most appropriate tools and methods to do this are. The unifying features of realist evaluations are that they are based on realist principles and always progress through three iterative stages: developing theory, testing theory, and refining theory articulated as C-M-O configurations. For examples of realist evaluation in practice, please see the links below.

### RESOURCES

- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.
- <u>Realist Impact Evaluation: An Introduction. Overseas Development Institute</u>
- Realist Evaluation. BetterEvaluation
- <u>Reality Bites: Making Realist Evaluation Useful in the Real World. Center for Development Impact</u> <u>Practice Paper</u>

#### **EXAMPLES OF REALIST EVALUATION**

- <u>Realist evaluation to improve health systems responsiveness to neglected health needs of vulnerable</u> <u>groups in Ghana and Vietnam: Study protocol</u>
- Applying the Realist Evaluation Approach to the Complex Process of Policy Implementation-The Case of the User Fee Exemption Policy for Cesarean Section in Benin
- How do patient feedback systems work in low-income and middle-income countries? Insights from a realist evaluation in Bangladesh
- What mechanisms drive uptake of family planning when integrated with childhood immunisation in Ethiopia? A realist evaluation
- <u>Harnessing the health systems strengthening potential of quality improvement using realist evaluation: an</u> <u>example from southern Tanzania</u>
- Role of trust in sustaining provision and uptake of maternal and child healthcare: evidence from a national program in Nigeria
- Tracing theories in realist evaluations of large-scale health programs in low- and middle-income countries: experience from Nigeria

# EFFECTIVENESS/IMPACT EVALUATION APPROACHES

Effectiveness (or impact) evaluations ask whether or not, and to what extent, observed changes are due to the intervention being evaluated rather than to other factors, including other programs and/or policies. An organization cannot claim that changes observed in a country or program area are project impacts unless there is a demonstrated link between these changes and the project intervention(s). Establishing this connection is called "causal attribution."

There are two broad approaches to assessing causal attribution one is non-experimental and the other is experimental (or quasiexperimental). We will deal with each separately.

#### **Effectiveness/impact**

evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.

<u>Note</u>: Although it is not uncommon to have qualitative or mixed methods evaluations that analyze program outcomes or causal pathways, the term "impact evaluation" is typically reserved for experimental or quasi-experimental evaluations with a counterfactual comparison/control group (described in the next section).

### QUALITATIVE COMPLEXITY-AWARE EVALUATION METHODS (NON-EXPERIMENTAL)

# IS THE EVIDENCE CONSISTENT WITH WHAT WOULD BE EXPECTED IF THE INTERVENTION WAS PRODUCING THE OBSERVED OUTCOMES? ARE THERE OTHER FACTORS THAT COULD PROVIDE AN ALTERNATIVE EXPLANATION?

Qualitative outcome evaluations attempt to establish causal attribution by identifying what evidence would be considered consistent with a causal relationship, gathering and analyzing data from different sources to determine whether the evidence matches this, and ruling out alternative explanations for observed outcomes. Qualitative outcome evaluation methods do not require baseline data collection or a control group and can be easily integrated into nearly any program work plan and MEL plan with a program strategy and duration that it is reasonable to expect outcomes to be observed during the life of the award. There is a wide variety of qualitative and complexity-aware monitoring and evaluation approaches that can be used for summative program evaluations, many of which build on or can be easily integrated with techniques used for performance monitoring and adaptive management and learning.

Qualitative complexity-aware evaluations are often written up as **case studies.** However, "case study" is a broad term that covers many types of activities, with different purposes and characteristics.

For guidance on case study approaches commonly used in formative assessments, process evaluations, and summative program evaluation, see <u>Better</u> <u>Evaluation</u>.

One example of qualitative complexity-aware study designs is

Outcome Harvesting, whereby evaluators search for outcomes (intended or unintended, positive or negative) and work backwards to build a narrative based on the intervention's contribution. Another example is Qualitative Impact Assessment Protocol (QUIP), which builds on contribution analysis (outlined above) to assess an intervention's impact by building a narrative of the main drivers and their contribution from the viewpoint of the intervention's beneficiaries. More details on these methods and other examples of qualitative complexity-aware monitoring approaches can be found at the links below.

### RESOURCES

- MOMENTUM Knowledge Accelerator. Guide to complexity-aware monitoring approaches for MOMENTUM projects
- Outcome Harvesting. BetterEvaluation
- OutcomeHarvesting.net
- Qualitative Impact Assessment Protocol (QUIP). Better Evaluation
- <u>Attributing Development Impact: The Qualitative Impact Protocol Case Book</u>

### EXAMPLES OF QUALITATIVE AND COMPLEXITY-AWARE APPROACHES

- Outcome Harvesting Evaluation of Social and Behavior Change Communication Capacity Strengthening Activities in Liberia
- Evaluating Capacity Strengthening for Social and Behaviour Change Communication through Outcome Harvesting
- Improving Results of District Health Plans in Niger. HNP Knowledge Brief
- Breaking Down Barriers to HIV Care for Gay and Bisexual Men and Transgender Women: The Advocacy and Other Community Tactics (ACT) Project

### EXPERIMENTAL AND QUASI-EXPERIMENTAL EVALUATION METHODS

### IS THE INTERVENTION PRODUCING THE DESIRED OUTCOME(S)?

Teams considering undertaking experimental or quasi-experimental evaluations should be experienced researchers and/or consider partnering with academic institutions that have experience with these approaches.

**Experimental designs** are meant to establish causal attribution by developing an estimate of what would have happened in the absence of a program or intervention (a counterfactual scenario) and comparing this to observed effects of the program. This is done by randomly selecting individuals (randomized control trials) or groups of individuals (cluster randomized control trials) to receive either an intervention or control treatment. In large enough samples, this randomization process ensures a balance in observed and unobserved characteristics between treatment and control groups. This approach minimizes the risk that changes observed are due to differences between participants in each group and not the program or intervention itself. Experimental designs are considered the

### **Establishing causal attribution**

**Experimental designs** are defined by construction of a control group through random assignment.

Quasi-experimental designs construct a comparison group through matching, propensity scores, regression discontinuity, instrumental variables, or other means.

gold standard for causal attribution in clinical research (i.e., testing the safety and efficacy of pharmaceutical products or medical devices), but may not be appropriate, practical, or ethical when evaluating the impact of complex health service delivery or system strengthening interventions. This overview of experimental designs and practical guidance on experimental evaluation methods later in this document are included to serve as a comparison for more pragmatic quasi-experimental approaches. They illustrate why the most rigorous biomedical evaluation approaches may not be the best fit for summative program evaluations.

Quasi-experimental designs do not establish as strong a causal link as experimental designs. They also establish causal attribution by comparing observed effects of a program or intervention to a counterfactual scenario, but do not involve random assignment of participants to intervention and comparison groups. Instead, quasi-experimental studies attempt to mimic an experimental design by ensuring treatment and control groups are equivalent in terms of observed characteristics using statistical or judgment-based matching techniques.<sup>3</sup> If successful, then it can be assumed that comparison group outcomes are representative of those that the treated group would have experienced if nothing had changed (i.e., the counterfactual). Effective matching reduces the risk of bias, which in turn produces more valid and reliable results for the summative program evaluation. That said, it is rarely possible to eliminate the risk that intervention and comparison groups differ in terms of unobserved/unmeasured variables, which means that stakeholders must be sensitive to the possibility that the differences in outcomes were caused by factors other than the intervention. There are many quasi-experimental analysis methods. Two of the most common methods, difference in difference (DiD) and interrupted time series (ITS) analysis, will be presented later in this document. DiD measures the relative change in outcomes between intervention and comparison groups at two or more time points, while ITS measures changes in outcome levels and trends over several waves of observation before and after the introduction of an intervention.

<sup>&</sup>lt;sup>3</sup> https://www.betterevaluation.org/en/plan/approach/case\_study.

Due to the resources required in collecting information from non-intervention sites/participants, quasi-experimental designs may be more appropriate for summative evaluation of a specific program strategy or intervention approach rather than summative evaluation of the overall program (e.g., using a quasi-experimental design to evaluate effectiveness of integrating gender-based violence first-line response in family planning services at public facilities, but not for evaluation of a multi-faceted program to strengthen local partner capacity for preventing violence against women and child/early forced marriage).

When evaluating only one strategy within a multi-faceted program, evaluators must be able to clearly outline and measure the specific contribution of this strategy, apart from other strategies that may also have been applied. Ideally, this involves a detailed theory of change specific to the strategy being

#### RESOURCES

- USAID Technical Note: Impact Evaluation
- Impact Evaluation in Practice: Second Edition
- <u>Randomized Controlled Trials (RCTs)</u>, <u>Methodological Briefs: Impact Evaluation 7</u>, UNICEF Office of <u>Research</u>
- Quasi-experimental Design and Methods, *Methodological Briefs: Impact Evaluation 8*, UNICEF Office of <u>Research</u>

### EXAMPLES OF EXPERIMENTAL AND QUASI-EXPERIMENTAL EVALUATIONS

### Experimental study designs

- Immediate "Kangaroo Mother Care" and Survival of Infants with Low Birth Weight (RCT)
- <u>A cluster randomized trial of delivery of intermittent preventive treatment of malaria in pregnancy at the</u> <u>community level in Burkina Faso</u> (cRCT)
- An integrated approach to improve maternal and perinatal outcomes in rural Guatemala: A steppedwedge cluster randomized trial (SW-RCT)

### Quasi-experimental study designs

- Impact of the Ananya program on reproductive, maternal, newborn and child health and nutrition in Bihar, India: early results from a quasi-experimental study (DiD)
- Addressing broader reproductive health needs of female sex workers through integrated family planning/ HIV prevention services: A non-randomized trial of a health-services intervention designed to improve uptake of family planning services in Kenya (DiD)
- No effects of pilot performance-based intervention implementation and withdrawal on the coverage of maternal and child health services in the Koulikoro region, Mali: an interrupted time series analysis (ITS)
- Effects of implementing free maternity service policy in Kenya: an interrupted time series analysis (ITS)

# PRACTICAL CONSIDERATIONS THEORY-BASED EVALUATION APPROACHES

### **CONTRIBUTION ANALYSIS**

### WHAT IT IS (AND ISN'T)

Contribution analysis is a theory-based evaluation approach that can be helpful in understanding if and how an intervention contributed to observed outcomes. It is most commonly used in projects focusing on capacity strengthening and policy influencing, where there are often many different contributors to change.

Contribution analysis is not a method for quantifying effectiveness or contributions in the same way that experimental and quasi-experimental impact evaluations can. It does not require baseline measurements or establishment of control or comparison groups.

### WHEN IT SHOULD (AND SHOULDN'T) BE USED

When to use contribution analysis	When NOT to use contribution analysis		
• When interventions are based on a relatively clear/established theory of change.	• When the causal framework is weak or loosely defined.		
• When there is little variation in how a program is implemented across sites.	• When project implementation cannot be well documented.		
• When there is a desire to articulate the project's contribution to observed outcomes (and sole attribution is difficult).	• When the program is of relatively short duration or it is unlikely that implementation was of sufficient intensity or duration for notable change to occur.		

### HOW IT WORKS

Contribution analysis is based on a defined series of steps, outlined in the <u>MOMENTUM Knowledge</u> Accelerator Guide to Complexity-Aware Monitoring Approaches for MOMENTUM projects.

**STEP 1.** Set out the question(s) to be addressed (e.g., "what role did the intervention play in bringing about the outcome?").

**STEP 2.** Develop (or review and refine) a causal framework, including risks, assumptions, and evidence gaps. Identify factors outside of the project that influence change. Consider alternative, rival explanations of how change may have come about.

**STEP 3.** Gather qualitative and/or quantitative evidence to assess the causal framework, the outcome, and potential contributing factors. This may include conducting desk reviews, synthesis of existing data sources and project documentation, or focused primary data collection.

**STEP 4.** Develop a contribution narrative describing the intervention, how it led to the outcome, and other contributing factors. Assess the strength of the narrative and supporting evidence. If the narrative is weak in some areas, determine whether it is possible to gather further evidence to increase confidence in the findings, or if uncertainties must simply be acknowledged.

### **STEP 5.** Gather further evidence to strengthen the narrative.

**STEP 6.** Revise and strengthen the contribution narrative.

### POTENTIAL CHALLENGES AND CONSIDERATIONS

- Contribution analysis can be time and resource intensive. The technique is intended to be carried out through an iterative process, meaning that evidence is repeatedly collected and analyzed, and narratives are gradually refined.
- Identifying and testing various contributing factors requires analytical expertise.
- If designed as part of a comprehensive MEL plan, many performance monitoring and adaptive learning activities can be leveraged for contribution analysis.

### **REALIST EVALUATION**

### WHAT IT IS (AND ISN'T)

Realist evaluation is a theory-based evaluation approach that examines how and why interventions work or do not work, for whom, and under what circumstances. It is most commonly conducted after decisions have been made to scale up or expand programs to new areas, particularly where needs for adaptation are anticipated.

Realist evaluation is not for research and evaluation novices. Realist evaluation is based on the notion that programs are "complex interventions introduced into complex social systems" and is best implemented by, or in collaboration with, experienced social scientists.

#### WHEN IT SHOULD (AND SHOULDN'T) BE USED

When to use realist evaluation	When NOT to use realist evaluation
<ul> <li>For evaluating new initiatives or programs that appear to work, but how, in what context(s), and among which sub-populations is not fully understood.</li> </ul>	<ul> <li>When, how, why, and where programs work is already well understood.</li> <li>When there is no context or outcome data available.</li> </ul>
<ul> <li>For evaluating programs that will be rolled out/scaled up, to understand how to tailor the intervention to new contexts.</li> </ul>	<ul> <li>When selected or potential evaluation team members lack experience with theory-based analysis.</li> </ul>
• For evaluating programs that have previously demonstrated mixed patterns of outcomes, to understand how and why the differences occur.	<ul> <li>When the program is of relatively short duration or project staff are unable to agree on an initial program theory.</li> </ul>

#### HOW IT WORKS

The research questions for a realist evaluation need to reflect the principles of realist evaluation, as well as the project-specific purpose for the evaluation. Some questions will be descriptive (saying what has happened and to what extent) and others will be explanatory (how or why something happened).

### **DEVELOPING THEORY**

### STEP 1: Understand the program and its theory of change

The first step in realist evaluation is developing an initial working program theory. The program-specific theory of change (in the case of MOMENTUM Country and Global Leadership, this should have been adapted from the global theory of change) is a good starting point. However, as a high-level overview of how the program is supposed to work, it may not be detailed or focused enough to serve as a strong initial program theory for a realist evaluation.

Initial program theories typically go beyond the theory of change, as presented in a program description or work plan, to unpack assumptions and explain causal links (hypothesize what goes on underneath arrows that link outputs, outcomes, and impacts). Developing initial program theories may require literature review, particularly review of established theories in academic and applied research literature, and iterative discussions with the program design team and local stakeholders.

### STEP 2: Articulate C-M-O hypotheses

C-M-O hypotheses are more granular explanations about what works for whom, in what context, related to specific interventions and settings. These can be developed at the start of a project, or in early/mid stages of implementation, once it is clearer what the intervention entails and how it works. C-M-O hypotheses are often organized in a table as an "equation" to clearly demonstrate the relationships between the different components. These hypotheses can provide explanations at multiple levels (individuals, organizations/communities, regions) of a program, and within different sub-groups. Within a given level, the researcher may hypothesize multiple context, mechanism, and outcome configurations.

For example, the following is one of multiple C-M-O hypotheses from the protocol for an evaluation of community health committees' influence on community capacity-building in Uganda:<sup>4</sup>

"Individuals within the community health committee are likely to provide supportive and consistent engagement for activities if they have strong motivation, a desire for volunteering for their community, and are committed to the group and its objectives **[M]**. This may be influenced by the individual members' specific attributes (such as availability of time and knowledge), previous experience and incentives provided to them **[C]**. This results in a decreased workload for the committee, due to increased collaboration, increased respect by community members and an overall committed committee better able to initiate activities and work towards building community capacity **[O]**."

As with the higher-level program theory, building on existing evidence and theory helps to develop richer insights that are grounded in what is already known, rather than starting from scratch or relying exclusively on how stakeholders think their program works. Building on existing theories can help avoid the tendency for evaluators to preferentially choose certain pathways that the program team wants to see work (or not work) versus others.

<sup>&</sup>lt;sup>4</sup> Gilmore B, McAuliffe E, Larkan F, Conteh M, Dunne N, Gaudrault M, ... and Vallières F. 2016. How do community health committees contribute to capacity building for maternal and child health? A realist evaluation protocol. *BMJ Open.* 6(11): e011885.

### **TESTING THEORY**

STEP 3: Conduct inquiries in relation to C-M-O hypotheses

Testing the initial program theory and associated C-M-O hypotheses can involve collection, analysis, and triangulation of a wide range of data sources to develop a quantitative and qualitative picture of the program in action.

In some cases, this may involve additional analysis and synthesis of data already collected for monitoring program performance, monitoring context, or adaptive learning purposes. In others, it may also require additional targeted primary data collection—designed specifically to explore C-M-O concepts and causal links (see <u>MOMENTUM Knowledge Accelerator Guide to Complexity-Aware Monitoring Approaches for</u> <u>MOMENTUM Projects</u> section on Causal Link Monitoring for additional guidance).

### **REFINING THEORY**

**STEP 4:** Assess how C-M-O hypotheses play out within the program, revisit, and refine initial program theory.

Theory refinement should be an iterative process during which evaluators and stakeholders discuss, review, revise, and reflect on C-M-O configurations to develop higher-level theories. Specifically, evaluation teams should look for different outcome patterns, examining the mechanisms that did or did not influence an outcome, and under which contexts these different pathways occurred.

This process should be repeated among evaluators and stakeholders to best refine and strengthen the configurations and work toward the final goal, a middle-range theory. The middle-range theory is the refined initial program theory that explains the "how" of the outcome results, including the contextual factors and their influence. This can then be used by decision-makers to inform future policy and program efforts in a variety of contexts.

### **STEP 5:** Report findings

Realist evaluation reports should reflect the principles and features of realist evaluation. This means describing theories in terms of C-M-O hypotheses, explicitly discussing disaggregated outcomes for different sub-groups, and presenting a refined program theory along with implications for decision-makers.

### POTENTIAL CHALLENGES AND CONSIDERATIONS

- Realist evaluation can be time and resource intensive. Do not underestimate time and effort required for literature review and analysis.
- Finding the right balance of theory and programmatic emphasis can be challenging, especially with multiple levels of theories (published theories, program theories and C-M-O configurations) to consider.
- Realist evaluations also carry a risk of being too academic to appeal to some stakeholders. This can be mitigated with attention to the balance of theory development and programmatic interpretation, or with a well-developed knowledge translation and uptake strategy for tailoring presentations of findings to different audiences.
- It is easy to be too ambitious in tackling complexity (e.g., planning to test multiple outcomes across multiple cases for multiple respondent groups). Where time and resources are in short supply, it may be more appropriate to use realist evaluation as one component within a larger evaluation investigating a few specific causal links or change processes that are of particular interest.

### **EFFECTIVENESS/IMPACT EVALUATION APPROACHES**

### QUALITATIVE COMPLEXITY-AWARE EVALUATION METHODS

### WHAT IT IS (AND ISN'T)

Causal inferences cannot be drawn from simply observing changes in outcomes before and after program implementation, because there may be factors other than the intervention causing changes over time. When stakeholders are interested in evaluating the effectiveness of a program strategy or intervention, they should choose a quantitative experimental or quasi-experimental method. Sometimes these quantitative methods may not be feasible or appropriate. This is especially true for complex intervention packages and/or interventions or contexts that are changing throughout the project period. In this case, evaluators will want to choose a qualitative impact evaluation method to provide insights on both intended and unintended outcomes of the intervention. Although qualitative impact evaluations are often much less resource intensive than experimental and quasi-experimental studies, and do not require advanced statistical training, the quality of inferences will only be as strong as the effort and skill invested.

When to use qualitative complexity-aware methods	When NOT to use qualitative complexity-aware methods
• When there is an interest in assessing both intended and unanticipated outcomes.	<ul> <li>When there is interest in quantifying an outcome measurement.</li> </ul>
• When prospective data collection is not an option, and baseline data are not available.	• Where there is only interest in whether a change occurred, but not about the how or why.
• When more traditional effectiveness/impact evolution methods are not practical with the resources and time available.	<ul> <li>When there is not sufficient time and resources to engage with a variety of stakeholders.</li> <li>When the evaluation team lacks familiarity or comfort with qualitative and complexity-aware evaluation methods.</li> </ul>

### WHEN IT SHOULD (AND SHOULDN'T) BE USED

### HOW IT WORKS

Examples of complexity-aware evaluation techniques that are well suited for qualitative impact evaluations include:

- **OUTCOME HARVESTING** Outcome Harvesting is a participatory impact evaluation method that is most useful when decision-makers are interested in learning what was achieved but anticipated outcomes are not clearly specified at intervention start-up, or intervention approaches evolve over the course of implementation in complex settings with high levels of uncertainty. Outcome Harvesting collects evidence from program stakeholders of what has changed. Then working backwards, evaluators determine whether and how an intervention has contributed to these changes.
- QUALITATIVE IMPACT PROTOCOL The QUIP is a form of impact evaluation wherein independent evaluators collect narrative accounts from intended beneficiaries about what caused changes in specified areas of their life over a defined period. QUIP evaluations tend to be guided by a desire to understand if a change happened and why, rather than to quantify an effect. Data collectors are not provided information

about the project, and are trained to use probing questions to collect as much information as possible about changes observed and reasons why change has happened. Interview transcripts generate "change stories" that are reviewed by analysts who are familiar with the project's theory of change and can identify and code evidence of change directly linked to the program or intervention, as well as to other causes.

### POTENTIAL CHALLENGES AND CONSIDERATIONS

- Non-experimental, qualitative, and complexity-aware evaluation methods require flexibility, adaptability, and responsive thinking at all stages of design and implementation, which is not possible without good team communication and management.
- Despite being less resource intensive than experimental designs, qualitative outcome/impact evaluations still require dedicated time, resources, and technical expertise. If well-planned, however, optimizing data collected in an ongoing manner from performance monitoring and pause-and-reflect activities over the course of implementation may minimize the time and resources needed for additional summative data collection.
- These methods may be unfamiliar to many implementing partners and stakeholders; therefore, there may be resistance to their use initially. Involving key stakeholders in evaluation design and planning processes can help mitigate resistance and facilitate interest and uptake of findings.

### EXPERIMENTAL AND QUASI-EXPERIMENTAL EVALUATION METHODS

### WHAT IT IS (AND ISN'T)

There are numerous resources available on experimental and quasi-experimental designs in health services, and health systems and policy research, as well as international development more broadly. If stakeholders identify quantitative evaluation of the effectiveness of a program strategy or intervention as a priority, we recommend consulting with experienced evaluation experts as early as possible in the program design process to determine whether such an evaluation is appropriate and feasible with time and resources available.

#### WHEN IT SHOULD (AND SHOULDN'T) BE USED

When to use experimental/quasi-	When NOT to use experimental/quasi-
experimental methods	experimental methods
<ul> <li>When there is an explicit demand to quantify intervention effectiveness is controlled (experimental) or real-world (quasi- experimental) settings.</li> <li>When evaluators are involved in design and implementation from the beginning of the project.</li> <li>When there are resources to engage individuals or organizations with experience conducting experimental and quasi-experimental studies.</li> </ul>	<ul> <li>When the program or intervention will not have been implemented for long enough to observe anticipated outcomes.</li> <li>Few USAID-funded global health programs will have the time and resources to conduct an experimental or quasi-experimental evaluation. Select projects may expect to conduct this type of evaluation, but these "special studies" are typically highlighted in the initial project description and resourced accordingly.</li> <li>When it is unethical to assign participants to an intervention or control group.</li> </ul>

### HOW IT WORKS

We present a few selected experimental and quasi-experimental approaches for evaluation of program strategy and intervention outcomes. We believe that these are particularly well suited to being incorporated into the fabric of the program design:

• Experimental stepped-wedge trials are a form of cRCT with a crossover design whereby all clusters start in the control arm and the intervention is introduced by random allocation and at regular intervals either to one cluster at a time or in small groups of clusters, until all clusters eventually receive the intervention. Researchers often think of conducting a stepped-wedge cluster randomized controlled trial (SW-cRCT), rather than a traditional cRCT with parallel treatment arms, for one of three reasons: First, the appeal of phased implementation to avoid the challenges of introducing an intervention to a large number of clusters simultaneously. Second, ethical concerns about withholding an intervention with clear benefits from the control arm for the full duration of the trial. Third, where a policy decision has already been made to implement the intervention and simultaneous rollout is impractical. All three of these conditions often are fulfilled in "real-world" programming, where an intervention will roll out in waves. This programming design can be taken advantage of to use a stepped-wedge evaluation design.



### FIGURE 1. STEPPED-WEDGE CLUSTER RANDOMIZED CONTROLLED TRIAL DESIGN

The design has numerous methodological complexities, including the possibility of time-varying treatment effects, changes in correlation structures over time, and the possibility of within-cluster contamination that require detailed documentation and advanced statistical expertise to address. Data collection and quality assurance costs can also be a barrier to using the SW-cRCT design for evaluation of many USAID global health programs, given that they require consistent data from all intervention and control clusters in all time periods.

• Quasi-experimental DiD estimation is an analytical technique that estimates the effect of an intervention by comparing changes in the outcome over time between a population that is enrolled in a program (the intervention group) and a population that is not (comparison group). The DiD technique is often used when randomization at the individual level isn't an option, which makes it a useful technique for large-

scale health and development programs. It requires longitudinal cohort data, or repeated cross-sectional data, from pre- and post-intervention and that certain assumptions about composition of intervention/comparison groups and trends in outcomes are met. The parallel trend assumption, one of the core assumptions for this technique, states that in absence of the intervention or program, the difference between the intervention group and the comparison group is constant over time, as shown in the figure below (labeled "constant difference over time"). This assumption can be difficult to validate, though having multiple time points pre- and post- intervention is recommended in order to visually inspect whether the assumptions holds true. It is also critical that the group assignment was not influenced by the outcome measurement at baseline and that the group composition is stable over time.



#### FIGURE 2. DID ESTIMATION APPROACH

• Quasi-experimental ITS analyses. The ITS method is used to estimate the effect of interventions by examining the change in the level and trend of an outcome after an intervention is introduced. It can be used in a situation when comparison sites are not available as all eligible places receive the intervention. This method requires data from a number of time points to be collected before and after the intervention is introduced (at least eight points before and eight points after), to allow modeling of what the trend in the outcome would have been if the intervention was not introduced. The model is compared to what actually occurs. Any change in the level of the outcome or in the rate of change over time, compared to the model, can be interpreted as the effect of the intervention. For video tutorials on conducting this type of analysis, please see the link in <u>Additional Resources</u> at the end of the document.

### **FIGURE 3. ITS DESIGN**



### POTENTIAL CHALLENGES AND CONSIDERATIONS

- It can be challenging to establish a counterfactual in many settings and adequately account for potential sources of bias. In these cases, an experimental or quasi-experimental evaluation may not be an appropriate use of resources.
- Experimental and quasi-experimental evaluations often require advanced statistical analysis (oftentimes requiring academic partnerships or consultants). This has contractual and resource implications and, if necessary, should be included in program work plans and budgets.
- Experimental and quasi-experimental evaluations of complex programs can be incredibly time and resource intensive, often well beyond the scope of many global health program evaluations.

### **MODIFYING AN EVALUATION APPROACH**

Most programmatic evaluations will employ modifications of the approaches described here in order to balance rigor with feasibility and programmatic utility. In many cases, a "textbook" approach will not be possible due to constraints, such as limited time or funding, or due to contextual factors, such as a challenging geographic or political environment. Combining approaches is also possible. This can yield richer findings and more robust recommendations to guide programmers. We outline several key points to consider if adapting or combining evaluation approaches to fit program-specific needs and resource constraints.

#### Effectiveness/impact evaluations and theory-based evaluations are not mutually exclusive approaches.

Increasingly, outcome/impact evaluations are paired with qualitative or mixed methods implementation research techniques to provide insights into how and why interventions do or do not work in a given setting. Just to pick one of many possible examples, a recently published paper used a mixed methods evaluation in

Malawi to examine whether and how performance-based financing (PBF) affected intrinsic motivation, or the health worker's innate willingness to perform well.<sup>5</sup> The authors used a quasi-experimental DiD analysis technique to examine changes in intrinsic motivation at baseline and two years after the PBF intervention was implemented. The analysis was grounded in self-determination theory (SDT), informed by in-depth interviews with health workers in both intervention and control facilities. The authors sought to examine motivation-related elements and their impact on health workers' psychological needs for competence, autonomy, and relatedness, which SDT suggests are central to intrinsic motivation. The theory-based qualitative component was critical for elucidating reasons that the PBF intervention did or did not have an effect on intrinsic motivation.

# Most approaches described here can be adapted to suit program needs, particularly non-experimental and theory-based approaches.

This guidance is intended to serve as a resource for teams considering or planning to conduct a summative program evaluation. In many cases, approaches may be simplified or adapted to meet program needs. In such instances, it is important to accurately describe the method used. If the evaluation team has significantly deviated from the defining characteristics of a method, it may be more accurate to describe an evaluation approach as "realist inspired" or as a "simplified contribution analysis," for instance.

As an example of such a modified approach, if a program in rural Bangladesh wanted to evaluate a multifaceted intervention on infant and young child feeding involving community health worker training and community outreach to understand why it was more effective in improving exclusive breastfeeding practices in some areas compared to others, the team might choose a theory-based evaluation approach. Suppose the team identified realist evaluation as most appropriate for the evaluation question, but only followed steps 1 through 3 of a standard realist evaluation approach. That is, identifying a program theory, developing C-M-O hypotheses, and conducting inquiries to test these hypotheses. The team then reports selected findings along with recommendations for program modifications.

The team might have decided that this was the most feasible and programmatically useful way to carry out the evaluation, while still preserving enough rigor to yield results of enough validity on which to base programming decisions. In this case, since the team did not use findings to revisit and further refine the initial program theory, it may be more accurate to describe such an evaluation approach as "realist inspired" rather than a full realist evaluation.

Determining when to modify evaluation approaches and if methods used have deviated enough from established techniques to merit a modified description is a judgment call. Such decisions should be made in a group that includes both those with programming responsibilities and those with evaluation expertise.

# The validity of theory-based evaluation findings is greatly improved if the program was well-designed and implementation well-documented.

A good design includes a well-thought-out theory of change and documentation of the key strategies contained within it. Well-documented implementation includes documenting adaptations made to those same key strategies; ongoing collection of key indicators (performance, context, sentinel); complementary

<sup>&</sup>lt;sup>5</sup> Lohmann J, Muula AS, Houlfort N, De Allegri M. 2018. How does performance-based financing affect health workers' intrinsic motivation? A Self-Determination Theory-based mixed-methods study in Malawi. *Soc Sci Med*. 208: 1-8.

qualitative information that may explain variations in performance; and documentation of any pause-and-reflect meetings. A rich information base will help stakeholders to consider and test alternative hypotheses.

# If there are not sufficient time and resources to do an evaluation well, it may not be appropriate to conduct a formal summative program evaluation.

In this case, the donor may commission an external evaluation, or may rely on performance monitoring and program learning to assess project success or inform decisions about future investments.

# **ADDITIONAL RESOURCES**

Each of these is a link to the named resource:

How-To Note: Activity Monitoring, Evaluation and Learning (MEL) Plan

**USAID Evaluation Policy** 

USAID Learning Lab Evaluation Toolkit

**OECD Evaluation Criteria** 

USAID Guidance Note: Participatory Evaluation – Locally Led Approaches to Evaluation

Basic Toolkit for Adaptive Learning, MOMENTUM Country and Global Leadership

A Guide to Complexity-Aware Monitoring Approaches for MOMENTUM Projects

Video Tutorials: Statistical Methods When Using Routine Data in Evaluation (Interrupted Time Series)

# **APPENDIX A. EVALUATION CONCEPT NOTE TEMPLATE**

### **PART 1: CONCEPT NOTE**

Study Title	
Evaluation/Learning questions to be addressed	
Country(ies) where the activity will be implemented	
Principal investigator name(s), title/organization	
Co-investigator names, title/organization	
Date completed	

### I. LEARNING OBJECTIVES

*If relevant, include subsidiary learning objectives as bullets that give more detail than the main learning question you listed in the table above. Otherwise, just restate the learning question.* 

### II. BACKGROUND AND RATIONALE

In two to three paragraphs describe:

- <u>Rationale</u>: Why are you engaging in this evaluation? What is already known about the issues you will address?
- <u>Intervention/program description</u>: What is the intervention/program? What problems or issues are being addressed by the intervention or program?
- <u>Timing and scope of the evaluation</u>: When do you plan to conduct this evaluation? What stages of the intervention/program will it evaluate?
- Expected outcomes: What is the implication for practice of the findings/results of the study?

### **III. PROPOSED METHODS AND DATA SOURCES**

**Methods:** Describe the methods you will use to gather data. Will you use quantitative secondary data from health information systems or project monitoring and evaluation? Primary data collection from facility assessment or household survey? Qualitative (e.g., most significant change, pulse polls, after action review, learning meeting minutes, focus group discussions, key informant interviews)?

**Study population and setting:** Include information on where the study will be conducted, who the study population is, how many sites the data will be collected from. Please note if this involves any vulnerable populations (e.g., younger age, disease status, marginalized social group).

**Recruitment:** Include information on how study participants will be recruited/invited to participate. Please describe as a step-by-step process for each participant group. Will there be informed consent? Please describe the process and attach the consent form with this application. *Sampling:* Include the sample size and sampling frame where applicable for each method.

Data collection: Describe how data will be collected and reported.

### IV. PROPOSED TIMEFRAME FOR THE EVALUATION ACTIVITY

Please describe when you will do the four main phases of the evaluation activity. That is, planning (including obtaining any required reviews and approvals), data collection, analysis/writing, and dissemination. Use this Gantt chart and add more detail if necessary. Replace the Quarters (Q1, Q2, etc.) with actual dates.

Activity	Q1	Q2	Q3	Q4	Q5	<b>Q</b> 6	Q7	Q8
Planning / preparation								
Data collection								
Analysis / writing								
Dissemination								

#### V. EXPECTED PRODUCT(S) AND DISSEMINATION PLAN

### 1. To whom will you disseminate the findings of the evaluation activity?

Please state your primary and secondary audiences and briefly state how the results will be disseminated to them.

Primary audience	
Secondary audience(s), if applicable	

### 2. Expected product(s) to disseminate the evaluation activity

PRODUCT	CHECK ALL APPLICABLE
Peer-reviewed journal article	
National/international conference presentation	
Dissemination workshop (with MOH and/or other partners)	
Study brief	
Other (please specify):	

### PART 2: ADDITIONAL INFORMATION NEEDED FOR NON-HUMAN SUBJECTS RESEARCH/NON-RESEARCH/HUMAN SUBJECTS RESEARCH DETERMINATION

Note: This section of the template should be reviewed and adapted to align with organization-specific procedures for ethical review of primary data collection activities.

### VI. NON-RESEACH / NON-HUMAN SUBJECTS RESEARCH JUSTIFICATION

# 1. Please provide a brief statement describing your rationale that the outlined activities are not or may not be human subject research.

Describe the intent of the outlined activity, who it is intended to inform, and how the findings will be used. Describe any MOH involvement in the activity including if MOH requested or initiated the activity.

Examples of activities that are "not research" are data collection for routine reports to donors, assessments conducted in the context of health service quality improvement initiatives and to answer questions specific to program sites (not intended to be generalizable).

2. Will the findings from this assessment be relevant only to the local authorities or beyond to other researchers and to other settings (outside of this country), and perhaps be generalizable?

If you believe the findings will not be generalizable (e.g. replicable or transferable) to other settings, please explain why not.

### 3. Are you collecting data from human subjects for the purpose of research?

Is the information collected from humans but not about the individuals from whom the data are being collected? Interviews collecting information on personal opinions (as opposed to professional opinions) are not key informant interviews.

# 4. If you are analyzing *secondary* data, will the data come from information that is publicly available? YES / NO

### If YES, please give website address:

If NO, please fill in this table.

Α.	Does the secondary dataset you will use contain identifiers or links to identifiers?	
В.	Did you or any team member originally collect the data AND are able to make linkages to identifiers?	
с.	Do you have plans to have access to the identifiers?	

### VIII. PARTICIPANT DATA SECURITY AND CONFIDENTIALITY

# 1. Will you interact with people while implementing the proposed study activities? YES / NO

Will you interview people, give them a survey, or collect biological specimens? If yes, state the expected interaction(s) and type(s) of data to be collected by filling in this table.

Α.	Will any of the data you plan to collect include personally or politically sensitive topics? If yes, provide details.	
	Consider factors such as communicable disease status, alcohol/drug use or sexual behavior. Could the data you collect, if disclosed, put the respondent at risk of civil or criminal liability? Would it damage their financial standing, reputation, or employment?	
В.	Will you or others implementing the learning activities gather, access, or use participant identifying information *(PII)?	<b>YES / NO</b> If YES, specify the type of PII and how collected.

\*Information considered to be PII includes any of the following:

- Name
- Geographic information smaller than state
- Elements of dates (birth date, admission date, date of death)
- Telephone numbers
- Electronic mail (e-mail) addresses
- Government identification numbers
- Medical record numbers

- Certificate or license numbers
- Device identifiers or serial numbers
- URLs
- Biometric identifiers
- Full face photographic images or comparable images
- Any other unique identifying number, characteristic, or code

# 2. If you will gather, access, or use PII, please describe how you will ensure confidentiality of the data and how you will protect PII.

Consider whether it will be possible to reconnect the study data to the individual who provided the information and how to minimize the risks of identification.

#### IX. FUNDING, ROLES, AND RESPONSIBILTIES

#### 1. Funding for the proposed activities.

Will the program give a sub-award to an implementing partner(s) for research, including implementing partners based in country? Do all implementing partners engaged in research have an active FWA (US Federalwide Assurance for the protection of human subjects)?

**2.** Role of project, implementing partners paid through the project, and MOH involved in this learning activity. *Please complete the chart even if the implementing partner has not yet been identified.* 

Role during study/project implementation	Program staff (direct hires) Please specify if it is a consultant.	Partner/ sub-awardee A (enter name)	Partner / sub-awardee B (enter name)	мон
Design / planning				
Data collection from human beings				
Take informed consent				
Use data with personal identifiers				
<b>Other</b> (i.e., technical expertise, developing protocol and tools, etc.)				





### www.usaidmomentum.org

- @USAID\_MOMENTUM
- f @USAIDMOMENTUM
- in @usaid momentum

