Framework for Evaluating National Malaria Programs in Moderate- and Low-Transmission Settings

April 2019
Updated March 2020

Framework for Evaluating National Malaria Programs in Moderate- and Low-Transmission Settings

April 2019
Updated March 2020

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, NC 27516 USA
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This research publication has been supported by the President's Malaria Initiative (PMI) through the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AIDOAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of PMI, USAID, or the United States government. TR-19-334

ISBN: 978-1-64232-129-6

USAID FROM THE AMERICAN PEOPLE
CDC
U.S. President's Malaria Initiative
RBM Partnership To End Malaria
MEASURE Evaluation

ACKNOWLEDGMENTS

This document was initiated by the Evaluation Task Force of the Roll Back Malaria's Monitoring and Evaluation Reference Group. The following task force members contributed as authors of this document:

Yazoumé Yé (MEASURE Evaluation, ICF), Samantha Herrera (MEASURE Evaluation, ICF), Andrew Andrada (MEASURE Evaluation, ICF), Debra Prosnitz (MEASURE Evaluation, ICF), Ruth Ashton (MEASURE Evaluation, Tulane University School of Public Health and Tropical Medicine), Thom Eisele (MEASURE Evaluation, Tulane University School of Public Health and Tropical Medicine), Achuyt Bhattarai (President's Malaria Initiative [PMI]/United States Centers for Disease Control and Prevention [CDC]), Erin Eckert (PMI/United States Agency for International Development [USAID]), Lia Florey (PMI/USAID), Arantxa Roca-Feltrer (Malaria Consortium), Chris Drakeley (London School of Hygiene and Tropical Medicine), Sean Hewitt (VBDC Consulting Ltd), Ryuichi Komatsu (the Global Fund to Fight AIDS, Tuberculosis and Malaria [Global Fund]), Estifanos Shargie (Global Fund), Roger Tine (Université Cheikh Anta Diop de Dakar), Adam Bennett (University of California San Francisco), Cameron Taylor (Demographic and Health Surveys Program/ICF), John Painter (PMI/CDC), Anna Bowen (PMI/CDC), Abdul-Wahid Al-Mafazy (Zanzibar Malaria Elimination Programme), Rebecca Kiptui (National Malaria Control Program, Kenya), Mateusz Plucinski (CDC), Peter McElroy (PMI/CDC), Christelle Gogue (PATH), Misun Choi (PMI/USAID), Frank Chacky (National Malaria Control Program, Tanzania), Manuel Hetzel (Swiss Tropical and Public Health Institute), Sumaiyya Thawer (Swiss Tropical and Public Health Institute/National Malaria Control Program, Tanzania), Tabitha Kibuka (PMI Impact Malaria Project/Population Services International), Bolanle Olapeju (VectorWorks/Johns Hopkins University), Deepa Pindolia (Clinton Health Access Initiative), Inessa Ba (Clinton Health Access Initiative), and Arnaud Le Menach (Clinton Health Access Initiative). The Evaluation Task Force thanks all other individuals who contributed at various stages in the development of this framework.

The authors have tremendous gratitude for those who reviewed this document and provided constructive feedback to ensure that the document is relevant, accurate, and concise: Agbessi Amouzou and Melissa Marx (both of the Johns Hopkins Bloomberg School of Public Health) and Alexander Rowe (CDC).

Special thanks also go to Cindy Young-Turner for editing and proofreading multiple versions of this document.

We thank the knowledge management team of MEASURE Evaluation, University of North Carolina at Chapel Hill, for editorial, design, and production services.

Suggested citation

Evaluation Task Force of Roll Back Malaria's Monitoring and Evaluation Reference Group. (2020). *Framework for Evaluating National Malaria Programs in Moderate and Low Transmission Settings*. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ACCM	all-cause child mortality
ACD	active case detection
ACT	artemisinin-based combination therapy
API	annual parasite incidence
APLMA	Asia-Pacific Leaders Malaria Alliance
CCMm	community case management of malaria
CHW	community health worker
CRVS	civil registration and vital statistics
DHS	Demographic and Health Survey
DiD	difference-in-difference
GDP	gross domestic product
GTS	Global Technical Strategy
HDSS	health and demographic surveillance systems
HIS	health information system
HPM	human population movement
iCCM	integrated community case management
IPTp	intermittent preventive treatment in pregnancy
IRB	institutional review board
IRS	indoor residual spraying
ITN	insecticide-treated net
ITS	interrupted time series
M&E	monitoring and evaluation
MDA	mass drug administration
MERG	Monitoring and Evaluation Reference Group
MICS	Multiple Indicator Cluster Survey
MIS	Malaria Indicator Survey
MOH	Ministry of Health
NMP	national malaria program
NMSP	national malaria strategic plan
OPD	outpatient department

PCD	passive case detection
PCR	polymerase-chain reaction
PfPR	*Plasmodium falciparum* parasite rate
RBM	Roll Back Malaria
RCD	reactive case detection
RDT	rapid diagnostic test
RHIS	routine health information system
SARA	Service Availability and Readiness Assessment
SDG	sustainable development goal
SMC	seasonal malaria chemoprevention
SPA	Service Provision Assessment
VA	verbal autopsy
WHO	World Health Organization

1. INTRODUCTION

1.1. Background and rationale for this framework

The early 2000s brought about a renewed focus on malaria control, after decades of neglect in the 1960s through the late 1990s [1]. Malaria was recognized as a global health priority, and substantial investments in malaria control soon followed. Over the past two decades, these investments have resulted in significant global reductions in malaria cases and deaths, with many malaria endemic countries seeing large decreases in transmission [1, 2]. Although substantial advances were made, by 2015 it was recognized that progress had been uneven and that significant investments were still needed for progress to continue. In a renewed commitment, the Global Technical Strategy (GTS) 2016–2030 was developed and adopted by the World Health Assembly in 2015, setting ambitious goals by 2030 that included reducing malaria case incidence and malaria mortality rates by at least 90 percent (compared to 2015 levels) and eliminating malaria from at least 35 endemic countries [1]. The Sustainable Development Goals (SDGs) were also adopted in 2015 and reflect the same commitment toward ending malaria through Goal 3, which is focused on health and well-being [3].

Given the investments and subsequent scale-up of malaria interventions since the 2000s, the epidemiology of malaria has dramatically changed [2, 4]. Transmission has decreased in many countries, and, as a result, transmission has often become more focalized and heterogeneous. Countries experiencing these epidemiological changes require more granular and finer-scale data on transmission risk and incidence to effectively inform and target their interventions and track their progress. Given this more complex and evolving context, it is widely recognized by the malaria community that our available monitoring and evaluation (M&E) approaches and tools also need to evolve to meet country needs. Refined methods should enable countries to measure the progress and impact of their programs, and, at the global level, to reliably monitor and report on progress toward the goals outlined in the GTS and SDGs.

Existing and relevant guidance is available for carrying out process and outcome evaluations of national malaria programs (NMPs) (through the World Health Organization [WHO] Malaria Programme Reviews) and for conducting impact evaluations in high-transmission settings [5, 6]. In 2014, the Monitoring and Evaluation Reference Group (MERG) of the Roll Back Malaria (RBM) Partnership published a detailed guidance document for conducting impact evaluations in high-burden settings [5, 7], which built upon an initial evaluation framework for high-transmission settings developed in 2007 [8]. The framework and the guidance, however, are specific to high-transmission settings only and do not address the complexity of impact evaluations in countries with more focalized and heterogeneous malaria transmission. Since the evaluation framework for high-burden settings was developed, the landscape has further evolved with new interventions (e.g., seasonal malaria chemoprevention [SMC], mass drug administration [MDA]), more tools to measure prevalence and estimate transmission intensity (with greater sensitivity for low-transmission settings), and a generally greater understanding of and experience in the application of different analytic methods for malaria evaluation that need to be reflected in guidance and tools for countries.

The evaluation framework presented in this document intends to build on existing work by the MERG, expanding it to address settings along the continuum of malaria transmission, with a specific focus on moderate- and low-transmission settings. This framework also emphasizes the importance of process evaluation to impact evaluation, linking implementation processes to implementation strength to then demonstrate program impact on malaria transmission, morbidity, or mortality.

1.2. Scope and objectives of the framework document

This document provides an overarching framework for evaluating NMPs along the continuum of malaria transmission, from high- to low-transmission settings. It provides a theory of change to guide evaluation efforts along the continuum of transmission for NMPs and covers key evaluation objectives, questions, and indicators. The framework is intended to be applicable to malaria endemic countries experiencing transmission along the continuum from high to low and across the different regions of the world. We draw on and present examples from various countries and regions to illustrate the flexibility and adaptability of the framework in different country settings. Given the existing guidance and tools available, particularly for high-burden settings, the focus of this document is to provide detailed guidance on available evaluation designs, indicators, and analytic methods to evaluate the impact of NMPs on malaria morbidity and mortality outcomes in moderate-, low-, and heterogeneous-transmission settings.

In this framework, we use the WHO classifications [9] for the different transmission settings, recognizing that these are fluid and represent a continuum of transmission: high transmission measured as greater than or equal to 35 percent Plasmodium falciparum parasite rate (PfPR) or approximately 450 cases per 1,000 population annual parasite incidence (API); moderate transmission measured as greater than 10 percent but less than 35 percent PfPR or 250–450 cases per 1,000 API; low transmission measured as 1–10 percent PfPR or 100–250 cases per 1,000 API; and very low transmission as more than 0 percent but less than 1 percent PfPR or less than 100 cases per 1,000 API. We also recognize that countries may have differing definitions for transmission stratifications, and this framework is intended to be flexible enough to adapt to country-specific contexts.

We use the term impact evaluation to encompass evaluations that assess the plausible contribution of program interventions to malaria health outcomes and the changes in malaria health outcomes. This broad definition of impact evaluation reflects the different evaluation design options presented in this framework. It also reflects the feasibility and difficulty of conducting strict impact evaluations of NMPs,[1] given the challenges of interventions being implemented at a national scale, multiple concurrent implementation and funding platforms, differences in interventions implemented at smaller scale in countries with heterogeneous transmission, and often the lack of a contemporaneous control group, or counterfactual, to infer causal relationships between the program and health outcomes.

The key objectives of the evaluation framework are to provide the following:

- An overarching framework for evaluating NMPs along the continuum of malaria transmission
- Description of linkages between impact and process evaluation
- Specific recommendations and guidance for conducting impact evaluations in countries with moderate-, low-, and heterogeneous-transmission settings
- Guidance on how to bring together evaluation results at the subnational level to tell a national-level narrative in heterogeneous-transmission settings

This document does not explicitly cover very low and pre-elimination malaria transmission settings[2] because guidance is available for strategic planning and evidence generation to move such areas toward elimination [10, 11]. WHO also has specific guidance on the process for verifying and certifying a country as malaria-free

[1] Here we use the term "strict" to refer to the standard academic definition of an impact evaluation, which is an evaluation that attributes a change in impact measures directly to a program or program interventions (a probabilistic evaluation, or evaluation using a probability design from which causality can be ascertained).

[2] WHO defines very low and pre-elimination malaria transmission settings as greater than 0 but less than 1 percent *P. falciparum* parasite rate or less than 100 API.

[11]. Furthermore, this document is not intended to be an exhaustive resource on process and impact evaluations for all malaria transmission settings or an evaluation methodology. Guidance on process evaluations for malaria programs, impact evaluations in high-transmission settings, and evaluation methodology more broadly is available in other guidance documents and tools (Annex 1), which we highlight and provide relevant references to throughout this document where appropriate.

1.3. Target audience

The audience for this framework is M&E staff of NMPs, Ministries of Health (MOH), donor agencies, and other partners interested in evaluating malaria control and elimination programs in moderate-, low-, and heterogeneous-transmission settings. It is designed specifically for M&E teams within NMPs, to enable them to identify appropriate approaches for evaluating their programs and, as needed, determine whether and what kind of additional technical support may be required for evaluation.

1.4. Process to develop this framework

Under the RBM MERG, a task force was formed to lead the development of this framework. The scope and objectives of this framework were informed through a review and synthesis of existing guidance documents and tools for monitoring and evaluating malaria interventions and programs (Annex 1), as well as through discussions and meetings with the task force and MERG members. The task force comprises a subgroup of MERG members and includes representatives from NMPs, donor agencies, and malaria research and implementing partners. An external review of the final draft of the framework was conducted by malaria experts outside the task force.

1.5. Organization of this document

After this introduction (Section 1), we present an overarching theory of change for evaluating NMPs along the continuum of transmission, discussing key evaluation objectives, questions, and indicators for carrying out process, outcome, and impact evaluations (Section 2). This discussion is followed by a section on evaluation design options for impact evaluations in moderate-, low-, and heterogeneous-transmission settings (Section 3); a section on gathering evidence that discusses key data sources and indicators for measuring intervention coverage, malaria impact measures, and contextual factors (Section 4); a section on data analytic methods, synthesis, and interpretation for impact evaluation (Section 5); and guidance on implementing the evaluation framework (Section 6). The document also contains several annexes, including an annotated bibliography of available guidance documents and tools for malaria evaluation, a detailed indicator reference guide, and case studies that showcase applications of certain aspects of the evaluation framework.

2. INTRODUCTION TO MEASURING NATIONAL MALARIA PROGRAM ACHIEVEMENTS

2.1. Theory of change for national malaria programs across the spectrum of transmission

The theory of change for NMPs outlines the relationships between the program inputs, processes, and outputs of a malaria program, with the expected outcomes and desired impact of the program (Figure 1). In all transmission settings, national programs aim to reduce the number of malaria cases and deaths. The specific interventions and measures of impact depend on the country context. In high-transmission settings, it is not feasible to count malaria deaths nor to, consequently, measure malaria-specific mortality. Instead change in all-cause child mortality (ACCM) is used as a proxy measure of impact on malaria mortality in these settings. The theory of change presented in this framework (Figure 1) is designed to be broad in scope and to apply across different settings, with the intention that it be further adapted, or elaborated with detail, to a country's specific context.

To achieve the desired impact, high coverage of empirically proven malaria control interventions must be achieved among the populations at risk for malaria. The appropriate intervention package for programs to implement should be determined based on evidence and tailored to the country context and its transmission settings.

In high- and moderate-transmission settings, attaining high coverage of key interventions, including vector control (insecticide-treated nets [ITNs], indoor residual spraying [IRS]), intermittent preventive treatment in pregnancy (IPTp), and prompt and effective case management, is critical to achieving impact. These key interventions may be complemented by other new interventions, such as SMC or MDA, to further bolster the impact of the program. In low-transmission settings, the program aim may be to achieve or maintain high coverage of vector control interventions among the populations at risk for malaria, in addition to ensuring high coverage of prompt and effective case management. Community case management of malaria (CCMm) and integrated community case management (iCCM) are strategies that expand access to case management to hard-to-reach areas with poor access to health facility services [12]. It is critical in all settings to use malaria surveillance data of adequate quality to select evidence-based intervention packages. As transmission decreases, and in settings with low transmission, a strong, functioning, and responsive surveillance system will also become increasingly critical to inform evidence-based decision-making, including effective targeting of interventions, to achieve the program's desired impact. This is an iterative process informed by the evaluation of NMPs.

The relationships between the various inputs and processes that need to be put in place and the expected outputs that will enable high coverage of malaria interventions are illustrated in the theory of change. These are relatively similar across transmission settings.

Several contextual factors may affect program implementation (the inputs, processes, and outputs), program outcomes, and program impact throughout the program lifecycle.[3] There are four main categories of contextual factors to consider: (1) health system factors, (2) sociocultural and economic factors at the micro and macro levels, (3) environmental factors, and (4) epidemiological factors. The theory of change does not show an exhaustive list of contextual factors to assess but provides examples of key factors that should be examined to understand how they may have affected program implementation and program outcomes and

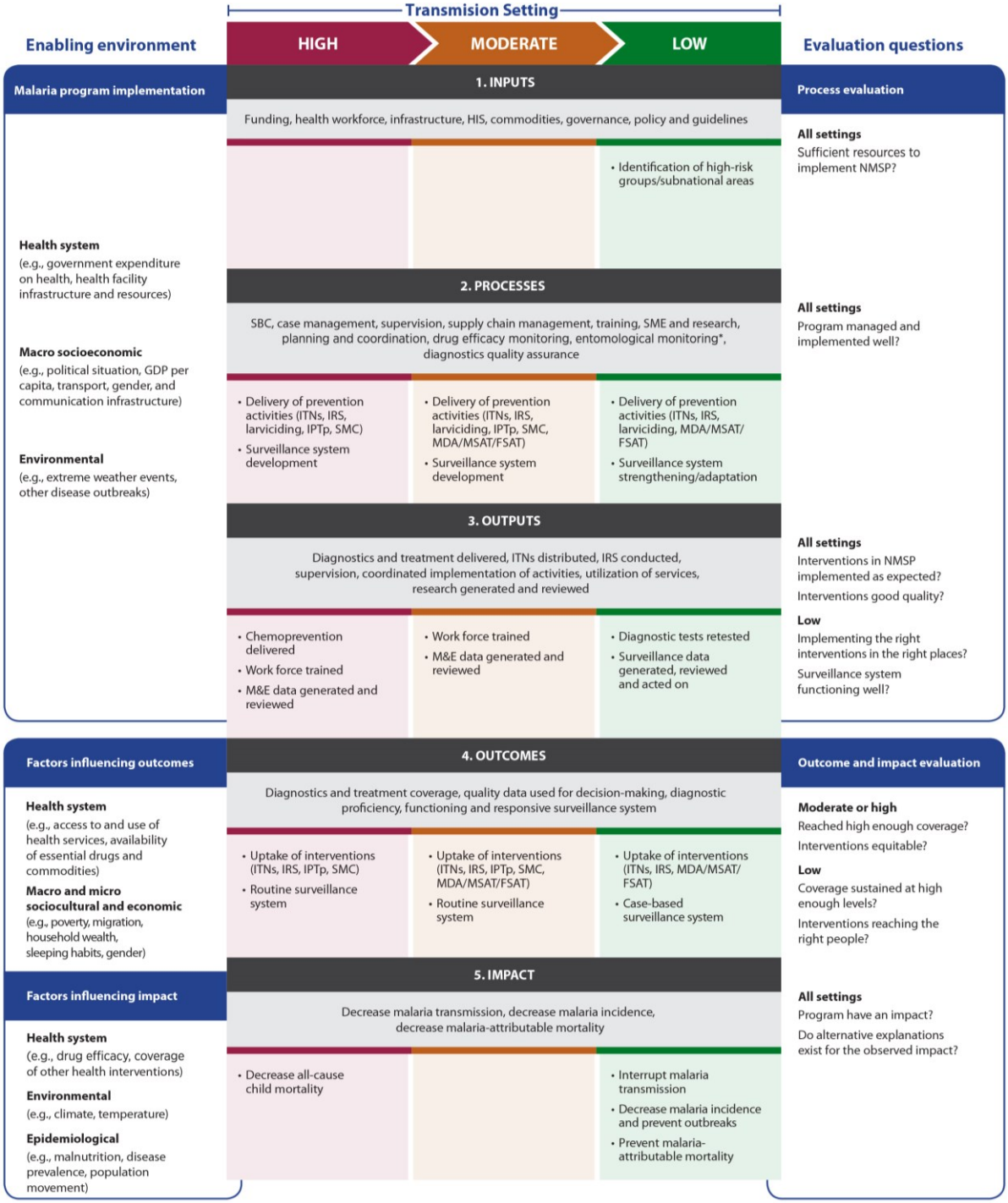
[3] The iterative cycle of designing, implementing, monitoring, and evaluating programs

impact. Therefore, users of this document should articulate the theory of change according to specific country contexts. Health system factors encompass government expenditure on health, health facility infrastructure and resources, access to and use of health services, availability of drugs and commodities, and coverage of other health services or interventions. Key macro-level sociocultural and economic factors include the political situation and stability, gross domestic product (GDP) per capita, poverty, the country's transportation and communication infrastructure, migration and urbanization, and crises and conflicts. At the micro level, the health system includes factors such as household wealth and parental education. Key environmental factors to examine include rainfall, temperature, land cover and vegetation, altitude, and extreme weather events such as floods or droughts. Epidemiological factors refer to the prevalence of other diseases (e.g., HIV), outbreaks of other diseases (e.g., Ebola), or malnutrition in the population at risk for malaria.

How these different contextual factors influence program implementation and program outcomes and impact is important to examine. Different health system, sociocultural and economic, and environmental factors may help enable or negatively affect the implementation of the program and thus affect the coverage of interventions achieved. At the outcome level, key health system and macro-level and micro-level socioeconomic factors, such as gender, access to and use of health services, availability of drugs and commodities, poverty levels, population migration, and household wealth, may directly influence the level of coverage of malaria control interventions achieved. At the impact level, health system, environmental, and epidemiological factors can indirectly or directly influence malaria transmission and malaria-related morbidity and mortality and confound the effect of the program on these impact measures. It is also important to consider that at the outcome and impact levels, the interplay between contextual factors and the transmission setting may be different. In other words, how contextual factors influence outcomes and the impact of a program may vary across transmission settings. Further discussion and detailed examples of the role of contextual factors in evaluation across moderate- and low-transmission settings is provided in Section 4.3.

The theory of change illustrates key questions for malaria programs to answer along the continuum of their program cycle, to provide the necessary information to guide malaria programmatic decision-making (e.g., course corrections, programmatic or policy changes) and to assess whether the program had the desired impact and met its targets and goals. The key priority evaluation questions for a national program will likely vary, depending on the country context and the transmission settings. For example, in high-transmission settings, the priority for the national program may be in understanding whether the program has achieved high enough coverage of malaria interventions to trigger impact, and if not, why? Conducting an evaluation that assesses the impact of the program may not be a key priority in these settings. For instance, if the program has not achieved sufficiently high coverage of interventions, there may not be a measurable impact. For moderate-transmission settings, if high coverage of interventions has been achieved, a key priority area may be to understand whether at these high levels of coverage, the program is having an impact (and the magnitude of that impact) on malaria transmission and reducing malaria morbidity and mortality. These findings could inform programmatic changes to further bolster impact or trigger investigations to understand better why the program is not having a strong impact. For low-transmission settings, the priority for evaluation may be at the process and output levels and understanding whether program interventions have been implemented as intended and assessing the performance of the surveillance system. Or as in moderate-transmission settings, the priority may be to understand whether the program's interventions are having an impact and, if not, to better understand the reasons behind the lack of impact.

Figure 1. Theory of change for national malaria programs across the transmission spectrum



*May include net durability monitoring and IRS application quality monitoring using cone bioassay. WHO provides specific guidance on entomological surveillance, and intervention monitoring and evaluation available at **http://www.who.int/malaria/areas/vector_control/entomological_surveillance/en/**

Definitions: **NMSP**=national malaria strategic plan; **ITNs**=insecticide-treated nets; **IRS**=indoor residual spraying; **IPTp**=intermittent preventive treatment in pregnancy; **SMC**=seasonal malaria chemoprevention; **SME**=surveillance, monitoring, and evaluation; **SBC**=social behavior change; **HIS**=health information system; **MDA/MSAT/FSAT** =mass drug administration/mass screening and treatment/focal screening and treatment

2.2. Evaluation of national malaria programs

Evaluations of NMPs provide critical information for programmatic and policy decision-making. They help the program and its stakeholders understand the degree to which the national malaria strategic plan (NMSP) is being implemented as intended and whether the program is achieving its desired impact. The theory of change outlines the key evaluation areas for NMPs—process, outcome, and impact—that feed into one another and together generate the necessary information for a comprehensive assessment of the NMP. The theory of change also illustrates the key data elements needed for these evaluations. Quantitative and qualitative data are used to answer many evaluation questions; data sources are discussed in Section 4. This section provides further details on these evaluations and the ideal timing of these evaluations during the NMSP cycle,[4] and discusses the important linkages between these evaluations.

2.2.1. Process evaluation

A process evaluation assesses the degree to which an NMP (and its NMSP) has been implemented as intended—and why or why not. A process evaluation encompasses tracking of program inputs (governance and leadership, policies and guidelines, resources and infrastructure), process indicators (the interventions and activities implemented), and outputs (types and quantities of services delivered, beneficiaries of services, and quality of interventions and services). It focuses on assessing the program's management and operations, implementation, and service delivery. A process evaluation examines questions, such as whether sufficient inputs have been allocated or mobilized for a program, what activities have been undertaken, and who has been reached by the program activities.[5]

Some examples of key process evaluation questions include the following:

- Was the NMSP designed based on evidence?
    - Was it designed based on the malaria epidemiological context?
    - Did the plan call for the appropriate interventions/right things to be done?
- Were the appropriate populations targeted? Was targeting of populations equitable?
    - Did the NMSP target the appropriate populations? (Were marginalized groups targeted?)
- Were there sufficient resources to carry out the NMSP?
- Was there an enabling environment for the program (e.g., supportive governance and leadership, policies and guidelines)?
- Were interventions implemented as expected?
    - Were the interventions delivered with good quality?
    - Were needed equipment and infrastructure available and functional?
    - Were providers available and performing?
- Was the program managed well?
    - Was appropriate training and supervision provided?
    - Was routine monitoring performed and results used for programmatic adjustments or other decision-making?
- Did the program achieve its expected outputs?

[4] A country's NMSP typically covers a 4–6 year period. In this document, we refer to this time period as the NMSP "cycle."

[5] Definition of process evaluation adapted from the Centers for Disease Control and Prevention Evaluation Manual and the United States Agency for International Development MEASURE Evaluation Monitoring and Evaluation of Malaria Programs Course.

Process evaluations are ideally carried out at least mid-way through and toward the end of the NMSP cycle. A mid-term evaluation provides an opportunity to assess whether the program is being implemented well and whether it is on track to meet its expected targets and goals, and to provide sufficient time for any necessary course corrections. A final evaluation toward the end of the NMSP cycle provides an opportunity to assess whether the strategic plan was implemented as intended and whether the program achieved its expected outputs, and to provide information to understand why the program did or did not achieve its targets and goals.

In high- and moderate-transmission settings, process evaluations need to focus on assessing the full package of interventions being implemented to identify bottlenecks for improving program implementation. Process evaluations in historically high-transmission settings that have become low-transmission settings due to malaria control efforts should examine all interventions to ensure that low transmission is maintained. In low-transmission settings with an environment or vector only marginally conducive for malaria transmission, the priority for a process evaluation will likely shift, putting a much stronger focus on assessment of case management implementation and the functioning and performance of the surveillance system.

Linking to outcomes

Process evaluations provide information to understand why a program may or may not have achieved its intended targets for intervention coverage (outcomes). It ties the inputs into a program, the processes or interventions implemented, and the outputs of a program, to the achieved outcomes. As a result, process evaluations can characterize the strength or intensity of program implementation. Poor or inadequate implementation of a program results in a lack of impact. It is important to have a good understanding of why a program did or did not achieve strong implementation and whether and how the level of implementation may have varied across different geographical areas, transmission settings, or different risk groups. This assessment provides for a deeper understanding of why the program did or did not achieve its set outcomes and can elucidate the cause and effect relationships between intervention implementation and achieved outcomes and impact.

2.2.2. Outcome evaluation

At the outcome level, the primary aim of an evaluation is to determine whether the program reached the expected level of coverage of interventions at the population level. Some key outcome evaluation questions to ask are as follows:

- Have malaria knowledge and behaviors changed?
- Has the program achieved population intervention coverage as planned?
- Has the intervention coverage been equitable?
    - Did all populations use services/access interventions?
- Has the surveillance system been properly functioning and responsive?

Key outcomes of interest to assess include the following: malaria knowledge, awareness, and uptake of malaria interventions among populations at risk; coverage of vector control interventions, chemoprevention interventions (as appropriate), and diagnostic testing and treatment; and performance of the surveillance system. The primary data sources for measuring coverage of outcomes are routine health information systems (RHIS) and cross-sectional surveys, such as population-based household surveys, facility surveys, and special surveys of specific populations or geographies. RHIS data, however, are not population-based and therefore do not typically provide accurate denominators for coverage estimates. RHIS data are limited to people accessing services at health facilities and, in some contexts in which community data are included in the

health information system (HIS), those accessing community health worker (CHW) services. Key population-based and facility surveys include the Malaria Indicator Surveys (MIS) [13], Demographic and Health Surveys (DHS) [14], Multiple Cluster Indicator Surveys (MICS) [15], Service Provision Assessments (SPA) [16], and Service Availability and Readiness Assessment (SARA) surveys [17]. If not captured in the HIS or existing population-based surveys, other efforts should be made to collect data on CHW service provision. Detailed indicator guidance is available for measuring several outcomes of interest for malaria programs [9, 18-20].

In high- and moderate-transmission settings, the key focus at the outcome level will be assessing the level of coverage of interventions achieved and differences in coverage levels across different subpopulations (e.g., geographical areas, urban/rural, specific age groups). In these settings, baseline and endline population-based surveys will serve as key data sources, alongside RHIS data. In low-transmission settings, the focus will shift to assessing the coverage of interventions among the populations at risk for malaria, with a much greater emphasis on assessing the coverage of diagnostics and treatment for malaria. Given the more focal nature of malaria in low-burden settings, data from nationally representative population-based surveys will likely be less useful; in these contexts, data from the surveillance system and special surveys, including those targeted among the populations most at risk, will be important data sources.

Linking to impact

An impact evaluation assesses and attributes changes in impact (e.g., malaria morbidity and mortality) to malaria intervention coverage. Before conducting an impact evaluation, it is important to assess whether the NMSP was implemented as intended (process evaluation) and whether implementation was strong enough (e.g., reached high enough coverage) among the population at risk (outcome evaluation) to demonstrate that the program had an impact on malaria transmission, morbidity, or mortality. There are no defined coverage thresholds for programs to use to make this determination [21], but reviewing the duration of implementation and trends in intervention coverage to determine whether coverage has improved or been maintained at high levels is important and informs whether it makes sense to carry out an impact evaluation.

2.2.3. Impact evaluation

The primary aim of an impact evaluation of an NMP is to determine whether the program as a whole had an impact on malaria transmission and malaria-attributable morbidity and mortality, and whether it achieved its goals. Specifically, an impact evaluation aims to assess the changes in impact measures that can be attributed to a particular package of interventions implemented by the NMP. To estimate a program's impact, a counterfactual is required; that is, a control or comparison group (or historical control) to estimate what would have happened to program beneficiaries had they not received the intervention.6 Impact evaluations for interventions with empirical evidence of efficacy and effectiveness need to measure indicators of health impact and often rely on intervention coverage results from an outcome evaluation. It is also essential that contextual factors be accounted for in an impact evaluation, because they can confound the association between the program and its potential impact or modify the effect of the program and affect the conclusion.

[6] Definition of impact evaluation adapted from MEASURE Evaluation's Monitoring and Evaluation of Malaria Programs Course, USAID's Evaluation Policy, and the World Bank's Impact Evaluation in Practice manual.

The main questions examined through impact evaluations of NMPs are as follows:[7]

- Have the goals of the NMP or NMSP been achieved?
- Is it plausible that the program contributed to measured impacted?
    - Was the impact equitable?
    - Do alternative explanations exist for the observed impact?

The indicators selected to measure the impact of a malaria program will be context specific. They will be largely based on the country's malaria epidemiological context, the goals of the NMSP, the availability of data, and the quality of the data. Recommended primary- and secondary-level impact indicators are outlined in Table 1 and detailed further in Section 4 (Table 2). In high-transmission settings, ACCM is the primary recommended indicator; in both moderate- and low-transmission settings, malaria case incidence is the primary recommended indicator. Case incidence is the recommended indicator in these settings for several reasons: it is collected routinely and available at subnational levels, there are no additional costs associated with the collection of the data, and the indicator is more sensitive to short- and long-term changes. This indicator is limited because it captures only those people who are accessing health services; this bias must be adjusted for in the analysis and interpretation. As countries progress toward lower-burden settings, fewer malaria-attributable deaths are expected, and changes in malaria transmission will be more difficult to detect over time. For these reasons, the following key impact indicators are listed as secondary: malaria test positivity rate, proportion of malaria admissions, malaria mortality, number of annual malaria outbreaks, parasite prevalence, and seroprevalence.

Impact evaluations should ideally be prospectively planned at the start of an NMSP cycle and should include evaluation of process and outcomes, the results of which are key inputs to the impact evaluation. Data are needed from the period of NMSP cycle implementation and one transmission after the cycle ends, so the evaluations should be timed accordingly, typically every five years. This timing allows for an assessment of whether the targets and goals of the NMSP were achieved and provides valuable data and learning to inform the next NMSP. Further, given the resources required to undertake an impact evaluation and the length of time for program interventions to have an impact, it is most appropriate to time an impact evaluation for the end of the NMSP cycle. Ideally, at the end of the NMSP cycle, the evaluation will incorporate an assessment of both the processes and the impact of the NMP to be able to fully understand why there was or was not an impact plausibly attributable to the NMP (see Section 3.1).

2.3. Monitoring and evaluation indicators for malaria programs

The theory of change (Figure 1) guides the key data elements and indicators that need to be monitored by a malaria program through the program cycle and that feed into process, outcome, and impact evaluations. Key indicators for malaria programs across high-, moderate-, and low-transmission settings are presented in Table 1. Indicators for high-transmission settings are included for reference;[8] those for very low and pre-elimination transmission settings are not shown, because existing guidance is available for measuring progress toward elimination [10, 11]. A more detailed indicator reference guide is available in Annex 2.

[7] Adapted from Victora, et al., *The Lancet,* 2011.

[8] As noted in Section 1, given the existing guidance available for high-transmission settings, the focus of this document will be on providing guidance on indicators and evaluation methods for moderate- and low-transmission settings only.

Table 1. Monitoring and evaluation indicators for malaria across high-, moderate-, and low-transmission settings

Indicator	High (≥35% PfPR or ≥450 per 1,000 API)	Moderate (>10–<35% PfPR or 250–450 per 1,000 API)	Low (1–10% PfPR or 100–250 per 1,000 API)
I. Input			
Expenditure per capita for malaria control and/or elimination [9]	X	X	X
Human resources: Number of health workers per 10,000 population [22]	X	X	X
Annual number of malaria commodities procured by type	X	X	X
II. Process			
Standards and guidelines development	X	X	X
Delivery of malaria diagnostics and treatment services	X	X	X
Delivery of malaria preventive services to populations at risk (e.g., ITNs, IRS, IPTp, SMC)	X	X	X
Targeted social and behavior change communication	X	X	X
Supervision of health care providers at facilities	X	X	X
Supervision of community health workers	X	X	X
Supply chain management and logistics management	X	X	X
Training of health staff	X	X	X
Surveillance, monitoring, and evaluation	X	X	X
Program planning and coordination	X	X	X
Drug efficacy monitoring	X	X	X
Insecticide efficacy monitoring	X	X	X
III. Output			
Social and behavior change communication			
Number and proportion of population at risk who recall hearing or seeing malaria messages within the past six months	X	X	X
Vector control			
Number of ITNs distributed	X	X	X
Number and proportion of households targeted for IRS that received IRS	X	X	X
Number of areas targeted for larviciding that are covered	X	X	X
Number of entomological monitoring sites	X	X	X
Chemoprevention			
Number of SP doses delivered for IPTp	X	X	
Number of children ages 3–59 months who received the full number of courses of SMC	X	X	
Diagnostic testing			

Indicator	High (≥35% PfPR or ≥450 per 1,000 API)	Moderate (>10–<35% PfPR or 250–450 per 1,000 API)	Low (1–10% PfPR or 100–250 per 1,000 API)
Number and proportion of health facilities with microscopy or RDT capability	X	X	X
Number of blood slides taken and read	X	X	X
Number of RDTs done and read	X	X	X
Number of microscopy slides cross-checked by national reference laboratory			X
Treatment			
Number of first-line antimalarial treatment courses administered	X	X	X
Number of pre-referral treatment courses administered	X	X	X
Number of radical cure treatment courses (primaquine or tafenoquine) administered (*P. vivax* settings)		X	X
Number of single, low-dose primaquine treatment courses administered for *P. falciparum* transmission blocking			X
Number of severe malaria cases referred	X	X	X
Number of antimalarial treatment courses for severe malaria cases administered	X	X	X
Commodities			
Number of health facilities with stockouts of key commodities for diagnostic testing	X	X	X
Number of health facilities with stockouts of key malaria drugs	X	X	X
Surveillance			
Number and proportion of expected health facility reports received on time	X	X	X
Number and proportion of expected health facility reports received that are complete	X	X	X
Training and supervision			
Number and proportion of health facilities with trained clinicians in case management	X	X	X
Number and proportion of health facilities with staff trained in surveillance, monitoring, and evaluation	X	X	X
Number and proportion of health facilities that received supervisory visits in the reporting period	X	X	X
Drug and insecticide efficacy monitoring			
Number of studies of drug efficacy completed	X	X	X
Number of studies of insecticide efficacy completed	X	X	X

Indicator	High (≥35% PfPR or ≥450 per 1,000 API)	Moderate (>10–<35% PfPR or 250–450 per 1,000 API)	Low (1–10% PfPR or 100–250 per 1,000 API)
IV. Outcome			
Malaria knowledge			
Proportion of population at risk who know the main symptom of malaria	X	X	X
Proportion of population at risk who know the treatment for malaria	X	X	X
Proportion of population at risk who know preventive measures for malaria	X	X	X
Vector control			
Proportion of population with access to an ITN in their household	X	X	
Proportion of population at risk that slept under an ITN the previous night	X	X	X
Proportion of population at risk protected by IRS during previous 12 months	X	X	X
Proportion of population at risk with access to an ITN in their household	X	X	X
Proportion of adult female vectors alive after exposure to insecticide (resistance frequency)	X	X	X
Resistance to insecticide status	X	X	X
Chemoprevention			
Proportion of pregnant women who received three or more doses of IPTp	X	X	
Proportion of eligible children ages 3–59 months who received the full number of courses of SMC per transmission season	X	X	
Diagnostic testing			
Proportion of patients tested among all febrile patients	X	X	X
Proportion of cases confirmed by a parasitological test out of all reported cases	X	X	X
Proportion of health facilities without stockouts of key commodities for diagnostic testing	X	X	X
Proportion of microscopy results cross-checked by national reference laboratory			X
Proportion of microscopists achieving both sensitivity and specificity greater than 90 percent during proficiency tests			X
Treatment			
Proportion of children under five with fever in the past two weeks for whom advice or treatment was sought from a health provider	X	X	

Indicator	High (≥35% PfPR or ≥450 per 1,000 API)	Moderate (>10–<35% PfPR or 250–450 per 1,000 API)	Low (1–10% PfPR or 100–250 per 1,000 API)
Proportion of patients with confirmed malaria who received first-line antimalarial treatment according to national policy [9]	X	X	X
Proportion of patients with *P. vivax* or *P. ovale* who received radical cure treatment (primaquine or tafenoquine) [9]		X	X
Proportion of confirmed *P. falciparum* cases who received single, low-dose primaquine			X
Proportion of severe malaria cases that were referred	X	X	X
Proportion of referred patients with severe malaria that received pre-referral treatment	X	X	X
Proportion of health facility months without stockouts of first-line treatments (includes treatment for severe anemia)	X	X	X
Proportion of patients with confirmed malaria with adequate clinical and parasitological response	X	X	X
Surveillance			
Proportion of malaria cases detected by surveillance systems	X	X	X
Annual blood examination rate	X	X	
Proportion of expected health facility reports received	X	X	X
Number and proportion of malaria outbreaks detected within two weeks [23, 24]		X	X
Number and proportion of suspected malaria outbreaks investigated		X	X
Number and proportion of malaria outbreaks responded to in a timely manner		X	X
Proportion of inpatient deaths due to malaria	X	X	X
V. Impact			
Malaria case incidence: number and rate per 1,000 people per year (disaggregate by species)	S	P	P
Malaria test positivity rate	S	S	S
Proportion of admissions due to malaria	S	S	S
Malaria mortality: number and rate per 100,000 people per year	S	S	S
All-cause child mortality (number of deaths among children ages 0–59 months per 1,000 live births)	P	S	
Annual number of malaria outbreaks		S	S

Indicator	High (≥35% PfPR or ≥450 per 1,000 API)	Moderate (>10–<35% PfPR or 250–450 per 1,000 API)	Low (1–10% PƒPR or 100–250 per 1,000 API)
Parasite prevalence: proportion of population with infection with malaria parasites[9]	S	S	S
Seroprevalence			S

P=primary indicator, S=secondary indicator, SP=sulfadoxine-pyrimethamine, RDT=rapid diagnostic test

NOTES: Moderate- and low-transmission criteria for PfPR and API defined by WHO.

Outcome and impact indicators should be disaggregated by age, sex, wealth, and geography as data availability allows.

[9] In areas with low prevalence (<1–2 percent), this is best done with qPCR.

3. EVALUATION DESIGN

Designing an impact evaluation requires taking a range of factors into consideration. The timing of an impact evaluation is discussed in Sections 2.2.3 and 6.2, and ideally it would coordinate with the Malaria Program Review cycle. The evaluation design will be influenced by the priority impact questions of the malaria program and other stakeholders, transmission settings of the country, data sources available (and quality of the data), and interventions applied and strategies used to introduce or scale up these interventions. The theory of change for the evaluation setting (Section 2.1) is also an important component in the early stages of evaluation design, and its development will assist in defining indicators and data sources for the evaluation.

3.1. Experimental, quasi-experimental, and non-experimental methodologies

In an impact evaluation, we strive to understand the causal relationship between the program activities applied and the change in malaria burden. Experimental methods, such as randomized controlled trials, have generally been considered the gold standard in demonstrating causal attribution and are generally reserved for evaluation of new tools or interventions with unknown efficacy. However, a study in which randomization is used to assign those who receive the program activities and "controls" who do not receive the program activities is not practical in situations in which malaria programs are scaled up to cover the entire population at risk, and these approaches may not be required or ethical in settings in which efficacy of tools or interventions have been previously demonstrated. Randomization can be applied in stepped-wedge study designs, in which programs are gradually scaled up (Section 5.4).

In practice, evaluation approaches most suited to moderate- and low-transmission settings are quasi-experimental methods, which use non-randomized exogenous variation in the exposure of interest to estimate effect sizes, for example, by taking advantage of natural variation in exposure to or uptake of the program or interventions, perhaps as a result of policy changes or intervention timing [25]. These quasi-experimental methods are generally perceived as the mid-point between classical randomized experiments and observational studies, because the non-randomization (at the group level) of interventions still opens the possibility for confounded effect estimates, but through rigorous design and analytical methods, quasi-experimental methods can better account for these threats to interval validity than observational studies [26]. Recent studies have indicated that interrupted time series (ITS) analyses, a type of quasi-experimental design, are particularly strong and may be included in some Cochrane reviews [27, 28]. Most designs discussed in this document are of the quasi-experimental type [26].

The use of counterfactuals, which describe the outcome in the absence of the program being evaluated, are important for establishing the impact of the program. Experimental methods, such as randomized controlled trials, estimate the counterfactual from the control group. Quasi-experimental and observational studies use various methods and assumptions to estimate the counterfactual (Section 5).

3.2. Stratification

The WHO Surveillance Manual describes the process of stratification as characterizing receptivity to transmission of malaria and the population at risk to target appropriate malaria interventions [9]. Malaria risk components that may be relevant to consider in stratification include not only API, but also metrics such as vectorial capacity, case treatment rates, importation rates, and severe malaria incidence [29].

In countries that experience heterogeneous transmission and have already defined malaria transmission strata (e.g., Senegal), these strata should be considered in the impact evaluation design. This may involve conducting the analysis within each stratum to understand the stratum-specific program impact. Alternatively, it may be appropriate to use different impact evaluation analyses within each stratum, because the interventions applied and data available may differ between strata. Where local-level stratification has been used to target specific intervention packages (e.g., to certain villages or health facility catchment areas), but interventions are likely to have community effects beyond the specific target areas, these "strata" should not be considered in impact evaluation design.

Furthermore, strata should not be considered as simply geographical units. In settings where malaria risk is linked with demographic and behavioral factors, a stratified analysis among higher-risk and lower-risk populations may be appropriate.

When deciding how to conduct stratum-specific analysis, the statistical power within each stratum should be considered, as well as the scale at which decisions regarding control program activities and interventions are made. As an illustration, individual but under-powered impact estimates for 10 strata may be less informative for decision-making than an analysis that combines these strata into two or three groups.

If a country does not have previously defined strata but is known to be a heterogeneous setting, an interim approach is to use baseline impact indicator values (e.g., confirmed malaria incidence) to define levels for a sub-group or stratified analysis. The number of groups or strata will vary according to the country context; two strata may be appropriate, or dividing into quartiles of incidence may be preferred. As stated previously, other characteristics may also need to be considered in this stratification approach, such as malaria importation rate or presence of special populations with different behaviors or risks.

It is beyond the scope of this document to provide detailed information about the process to develop a national malaria stratification, but further documentation regarding stratification procedures can be found in the WHO Surveillance, Monitoring and Evaluation manual [9].

3.3. What data are available for the evaluation and what are their quality?

The types of available data will influence which impact evaluation designs are feasible. In settings where an evaluation is being prospectively planned, the preferred evaluation design can inform the types of data collected. Potential impact indicators are introduced in the theory of change and further discussed in detail in Section 4, with a full listing and definitions in Annex 2. Some analytical approaches (e.g., segmented regression of ITS) require data collected over the evaluation period; therefore, either routine surveillance data or case management data from sentinel sites will be needed for these longitudinal analyses. Other analytical approaches (e.g., difference-in-differences [DiD]) can use cross-sectional survey data; these may be MIS, DHS, or other population-based surveys or surveys targeting specific populations or geographies.

The specific impact measures that are appropriate to use differ according to the transmission setting. For example, ACCM is not recommended as an impact indicator in low-transmission settings because only a small proportion of ACCM will be due to malaria; however, ACCM can be informative in moderate-transmission settings, particularly if RHIS data of adequate quality are not available. Confirmed malaria incidence is the primary recommended indicator in moderate- and low-transmission settings, and seroprevalence is particularly useful in in low-transmission settings due to its greater sensitivity to detect changes in transmission over time. Selection of impact indicators is discussed further in Section 4.2.

When identifying and assessing existing survey data for inclusion in the impact evaluation, it is essential to understand the sampling frame used, any sampling weights that will be required for secondary analysis, and

the analysis unit to which the survey is powered (e.g., national-level, regional-level, epidemiological stratum). Geostatistical models have been developed by the Malaria Atlas Project from these national survey data that estimate age-standardized P. falciparum prevalence, among other indicators, at resolutions of 5x5km or lower [30-33]. These model estimates are periodically updated with new survey data and are available for download by analysts and NMPs.

Cross-sectional surveys are often a valuable source of data that may be included as contextual factors, describing variables such as access to health services (see Section 4.3 for more detail on contextual factors and confounding). Several analytical approaches require baseline data or measuring a long period of pre-intervention outcome data. Data describing potential confounding factors are often also necessary if baseline data are being retrospectively compiled. If evaluations are planned and completed in sync with five-year NSP cycles, it may be possible for one evaluation's end line to serve as the following evaluation's baseline.

Data quality is an important component to consider, because use of poor-quality data can result in misleading or incorrect evaluation findings. It is, however, important to understand that data, particularly HIS data, do not need to be perfect to be used, just of "adequate quality." Although defining a cut-off for what is considered "adequate quality" is impractical, problems with missing or incomplete data, creation of new health facilities, and roll-out of confirmatory malaria diagnosis can be accounted for in data analysis. Concerns regarding falsified data or incorrect interpretation of indicators by the staff collating and reporting data are harder to adjust for and may require affected facilities or districts to be excluded from analysis. Data Quality Assessments can provide further information about data quality [34-37], and some suggestions on minimizing bias when using health management information system (HMIS) data in impact evaluation have been proposed by Ashton, et al. [38]. Addressing bias in HMIS data used for impact evaluation is also discussed in Section 4.1.1.

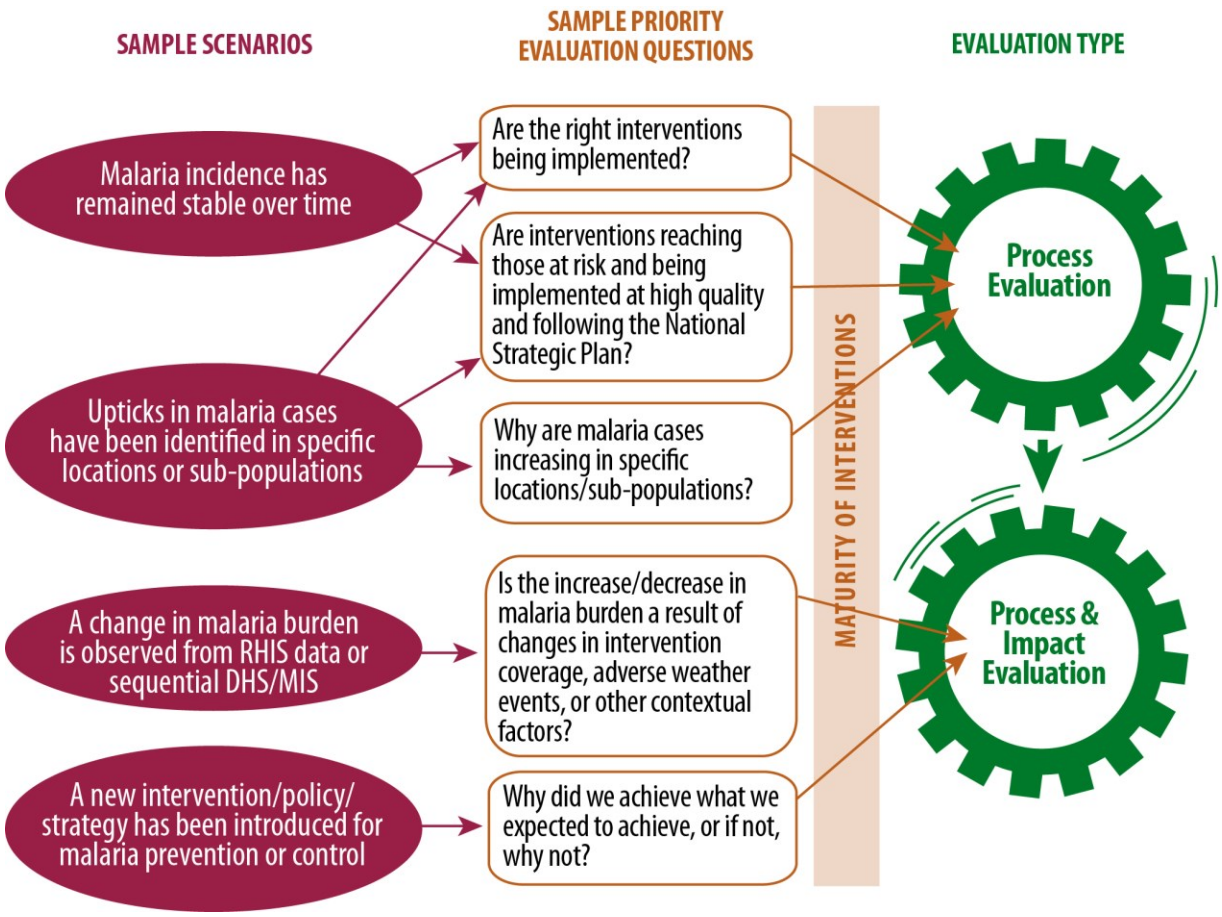
In designing the impact evaluation analysis approach and identifying data sources, it should be noted that including multiple data sources and running multiple analyses, as well as a range of techniques to address gaps and biases in the data, can improve the plausibility of findings using quasi-experimental evaluation approaches [39]. This approach is often described as triangulation. For example, if both cross-sectional survey data and longitudinal surveillance data from sentinel sites are available, it would be possible to run one analysis using impact indicators from the cross-sectional data and another using the surveillance data. If both analyses (each with relevant confounding factors) indicate that the program had an impact, this strengthens the plausibility that the program did have an impact.

3.4. What strategies have been used to introduce and scale up activities and interventions?

The approach used to introduce or scale up specific interventions can also influence the evaluation design. If an intervention was introduced using a phased approach, then it may be possible to consider this approach as a natural experiment and use a stepped-wedge approach [40]. If a policy change occurred, such as a new first-line drug was implemented over a very short period of time, then this rapid change could be presented using a classical ITS approach [41], which does not require a separate control group or area. It is also possible to use ITS studies for interventions that were not introduced over a very short time period, but the scale-up timing is known and can therefore be included in the analysis. For situations in which there is no clearly defined control area where the intervention or program was not implemented, analysis approaches such as dose-response or propensity score matching can be used to take advantage of varying levels of intervention coverage at a district or individual level in the population being evaluated [39, 42].

The evaluation scenario diagram (Figure 2) presents illustrative examples of evaluation questions and shows that the combination of transmission setting and intervention mix drive decisions on which type of evaluation to implement. The scenario diagram specifies that in low-, moderate-, and heterogeneous-transmission settings, there is a need to ensure that intervention coverage and maturity have reached a level sufficient to trigger a decline in malaria incidence (impact indicator); only then may an impact evaluation be implemented. If intervention coverage and maturity are below optimum, the focus should be on process evaluation.

Figure 2. Evaluation Scenario Diagram



Impact evaluations investigating the impact of a program or intervention on malaria burden require different approaches if the allocation of program activities has been decided according to existing data on malaria burden. An example of this "endogeneity" is a situation in which IRS is targeted to N districts in the country with the highest recorded malaria incidence. If this targeting approach is not adequately considered in the analysis, then the evaluation findings may falsely indicate that IRS is associated with an increased risk of malaria. Analysis approaches such as regression discontinuity methods and instrumental variables potentially address endogeneity issues (see Section 5.5). The MERG suggests that countries facing this challenge seek assistance from evaluation specialists to ensure that endogeneity is adequately addressed.

3.5. Addressing changes in reporting methods and denominators

An additional challenge in many settings is a situation in which indicator definitions or reporting methods have changed over time. This is particularly relevant if routine surveillance data are used. This section describes several common scenarios and potential solutions.

- **Scenario 1:** Introduction of rapid diagnostic tests (RDTs) resulted in increased access to confirmatory diagnosis. How should an impact indicator of confirmed malaria incidence over this period of diagnostic scale-up be used in analysis?

  In this kind of time-series analysis, access to malaria diagnostics must be incorporated into the analysis. Variables describing access to malaria diagnostics include the total number of individuals tested by RDT or microscopy, or the proportion of all-cause outpatient department (OPD) visits that were at facilities with malaria confirmatory testing. Including variables that capture changes in access to malaria diagnostics directly in analysis models as potential confounders is the optimal approach to minimize potential bias. Failing to account for increasing access to confirmatory diagnosis could bias impact estimates downward, because increases in confirmed malaria cases resulting from increased access to confirmatory testing are interpreted as true increases in malaria incidence.

- **Scenario 2:** The HMIS previously reported a combined indicator of presumptive and confirmed malaria incidence, but since the introduction of RDTs, it now includes only confirmed malaria.

  A first step in exploring this change in indicator definitions is to prepare plots of malaria incidence over time and include a reference line indicating when incidence was restricted to confirmed cases only. Presenting trends in secondary indicators (e.g., number of patients tested by RDT or microscopy) over this period, or complementing the analysis of HMIS data with cross-sectional survey data describing changes in population prevalence of malaria, can assist in understanding the trend in malaria incidence following this change in definition. A further alternative is to complete analyses separately for the period in which incidence included presumptive and confirmed malaria, and for the period in which incidence represented confirmed malaria only. In settings with a stable test positivity rate, an estimate of confirmed cases for the period before the introduction of RDTs could be generated by multiplying the number of presumed cases (assumed to be the number of fever cases) by test positivity. Failure to account for this change in malaria case definition in the HMIS would likely bias impact estimates upward, because an apparent decrease in malaria case count would be attributed to the program being evaluated, rather than to the change in case definition from presumptive to confirmed malaria.

- **Scenario 3:** Survey sampling frames were changed from the pre-intervention survey to the most recent survey, so that indicators reported by region do not reflect the same geographical area.

  If you aim to compare indicators for the same locations between the two surveys, it is important to use the raw data to generate weighted estimates for the same area. You should also note that the survey may not be powered to compare indicators for small areas and that changes in strata or region may have also resulted in changes to the control program activities or interventions implemented. Depending on the changes made to the sampling frame, failure to account for this change could bias impact estimates either up or down.

- **Scenario 4:** Shifting denominators (e.g., re-stratification resulting in a change in the population at risk of malaria)

Similar to the recommendations for Scenario 3, it is important to use raw data to be able to compare API with the same denominator definition over time. For example, if one region is no longer considered at risk of malaria, generating estimates of API both including and excluding this region over the time periods being evaluated can assist in explaining changes over time and where they occurred.

4. GATHERING EVIDENCE

Gathering evidence for evaluation entails defining evaluation indicators, identifying and gathering relevant data sources and datasets, assessing the quality of the data identified, and analyzing, triangulating, and interpreting the data. Results of process evaluations are integral to impact evaluations, providing evidence that ties program inputs to outputs and outcomes. In some cases, and particularly if there has not been a recent process evaluation, impact evaluation will entail the collection of primary data, depending on the proposed objectives of the evaluation and the resources available for data collection. In this section, we review the key data sources for intervention coverage, impact measures, and contextual factors, and discuss the key strengths and limitations of the sources.

4.1. Overview of key data sources

4.1.1. Routine health information systems

RHIS are an important source for malaria intervention data, providing information on malaria case management and malaria in pregnancy care, facility performance, and, in some contexts, CHW performance. Some countries have multiple sources of RHIS data, including the main RHIS that captures most routine health data across the public health system, and disease-specific surveillance and vital report systems that may be limited to certain geographic areas. Many countries are using DHIS 2 as a platform for the main RHIS.[10] RHIS data are generated at the health facility level, and in contexts in which CCMm or iCCM strategies are in place, from the community level; they are limited in that they capture data on only people who sought and accessed care. This includes information on malaria diagnostic testing, malaria test positivity, malaria treatment provision, IPTp, facility-based distribution of ITNs, commodity stockouts, facility and CHW reporting rates, and detection and timely response to epidemics. RHIS may also collect several key impact indicators on malaria morbidity and mortality relevant for moderate- and low-transmission settings; these include the number of malaria cases, the malaria test positivity rate, the proportion of inpatient admissions due to malaria, the number of malaria deaths, and the proportion of inpatient deaths due to malaria. Health facility catchment area population data drawn from census or vital registry data can be used to estimate the malaria case incidence rate and the malaria mortality rate. If the program includes active case detection (ACD) or reactive case detection (RCD), passively detected cases (passive case detection [PCD]) and cases detected actively or reactively should be reported separately. If both cases are combined to calculate malaria case incidence, it is important to report the indicator disaggregated by PCD and ACD/RCD because ACD/RCD rates are likely to be higher than those from PCD.

Several potential biases can be introduced when using RHIS data to measure coverage of interventions and impact. Foremost among these is reporting bias, because RHIS data capture only individuals who have sought or accessed care from a CHW or health facility. These biases result from various factors, including the following: challenges of defining health facility catchment population size; fluctuations in care-seeking and health facility use; changes in definitions of indicators over time; changes in health facility reporting rates; incomplete recording; fluctuations in the availability of diagnostic tests; and the availability, use, and reporting of cases from the private sector [5, 38]. In countries with low or more focalized malaria transmission, data on malaria testing may not be reported consistently at the facility level [5]. For malaria mortality indicators specifically, underreporting of malaria deaths and misclassification of cause of death can result in biased estimates of malaria mortality [5]. Further, deaths occurring outside public health facilities, including those in

[10] Owing to the nascent implementation of DHIS 2, data history might be limited. Gaining access to DHIS 2 data is also a potential limitation to their use for evaluations.

the community and in the private sector, are likely not captured in the RHIS. Malaria deaths that occur in public facilities may also not be representative of all malaria deaths [5]; thus changes in trends in malaria deaths observed at health facilities may not represent changes in malaria deaths occurring outside public health facilities.

To account for biases introduced through changes in health facility use and recording completeness, it is important to examine trends in all-cause outpatient and inpatient numbers (or attendance for non-malaria cases), and the proportion of malaria cases out of all-cause outpatient visits, inpatient admissions, and hospital deaths. Health facility reporting rates, if available, should also be examined and used to interpret the changes in coverage and impact indicators over time. Other data sources, notably population-based household surveys, such as DHS, MIS, or MICS, can be used to estimate health facility use by looking at the percentage of children under five for whom care or treatment was sought from public health facilities. These data can also provide information on health-seeking behavior in the private sector.

To help account for bias introduced by changes in diagnostic testing over time, indicators on malaria cases, inpatient cases, and malaria deaths should be stratified, as feasible and where available, by whether they were parasitologically confirmed. It is important to examine these indicators together with the malaria test positivity rate and the proportion of suspected cases that were tested to help with the interpretation of the data and trends over time. Some methodological considerations to account for biases in RHIS data during evaluations of NMPs are described in greater detail elsewhere [38, 43].

Although RHIS data may often be imperfect due to missing data and the potential biases mentioned above, an impact evaluation does not require perfect quality RHIS data. Statistical approaches (imputation methods) are available that can simulate missing data [43], and by accounting for various assumptions of the type of missingness and misclassification of outcomes, impact evaluation analyses can still produce valid estimates of causal effects. Quasi-experimental evaluation methods support the use of multiple sources of data and analyses, as well as the use of multiple techniques to address potential bias, to maximize the plausibility of the findings. As such, analyses using imperfect RHIS data can be complemented and triangulated by other analyses using different data sources, such as cross-sectional survey data. Facility-based case-control data may also be used to look at risk factors and associations with intervention coverage, and this could include the use of community controls.

Community health information systems

In countries with CCMm or iCCM, data from these programs can be an additional data source for malaria case incidence. In some countries, data from CCMm and iCCM programs are integrated in the RHIS and aggregated with health facility data in the system. In these contexts, it is important to account for the timing of when CCMm and iCCM programs began and their scale, because these factors can have a large impact on the interpretation of trends in malaria cases over time. For example, the roll-out of a CCMm or iCCM program will likely result in an increased number of reported malaria cases diagnosed and treated. This does not necessarily mean that malaria incidence or treatment rates have changed, but that they are more likely a reflection of wider access to services. Data on community health programs can also be used to help contextualize differences in access to health services in different settings, and thus may serve as an important contextual factor data source in an impact evaluation.

4.1.2. Surveys

Household surveys

Population-based household surveys, such as DHS, MIS, and MICS, are important sources for obtaining nationally representative information on malaria knowledge levels, coverage of malaria interventions, and parasite prevalence, in addition to other important contextual factors, such as access to health facilities, health-seeking behavior, and other demographic and health-related indicators of interest. These surveys are usually conducted every three to five years in a country. Notably, DHS and MICS surveys are typically conducted during the non-peak malaria transmission season, whereas MIS surveys are typically conducted during the peak malaria transmission season.

The RBM MERG has provided guidance on key malaria intervention coverage indicators that can be measured from household survey data, which encompasses coverage of malaria prevention and case management interventions [18]. Prevention indicators include those measuring ITN ownership and use for vector control and IPTp uptake among pregnant women. The case management indicators encompass prevalence of fever, treatment-seeking behavior, diagnostic testing, and appropriate first-line treatment among children under five years of age with fever in the last two weeks. IRS coverage indicators are no longer recommended for inclusion in household surveys by the RBM MERG due to the focal deployment of IRS, making it difficult to generate representative estimates of IRS coverage from surveys [18].

Several considerations should be made when interpreting coverage indicators. First, these surveys are generally powered to provide reasonably precise estimates at the national or subnational level (typically at the first administrative subnational level, such as region or province). In settings with more heterogeneous transmission and more focalized prevention efforts, these surveys will likely underestimate the level of coverage of interventions among high-risk populations surveyed in a nationally representative sample [5] and therefore may be less useful. Bayesian geostatistical modeling methods may help produce finer estimates [33]. The timing of the survey is also important to consider, because surveys are not always conducted during the same seasons. Using time series HIS data (e.g., from sentinel sites or reference hospitals), if available, can help contextualize survey data that were in-season or out-of-season for a particular year. Interventions may be used more or less depending on the season and the perceived risk of malaria, and therefore this could affect the interpretation of trends across surveys. Other important considerations include changes in the case management indicators over time and issues with recall bias [44, 45]. Previous case management indicators used fever as a proxy for malaria, but due to the scale-up of parasitological confirmation, these have since been updated. These indicators should be evaluated based on the country's case management intervention scale-up [5]. Indicators retrospectively assessing the type of treatment and receipt of IPTp during the woman's last pregnancy are subject to recall bias and may result in an underestimation or overestimation of coverage [46].

These surveys also provide national and subnational estimates of parasite prevalence among children ages 6–59 months, typically using RDTs and microscopy. The estimates of parasitemia prevalence from these surveys are a relevant source for measuring impact in high- to moderate-transmission settings, although these indicators need to be interpreted by taking into consideration important factors such as the timing and season of the survey. In low-transmission settings, particularly in settings with prevalence of less than 3 percent, these estimates become less useful for measuring impact for a number of reasons. The diagnostic tools used in these settings—RDTs and microscopy—have their limitations. They are unable to detect submicroscopic infections, which are proportionally more common in low-transmission settings [47], and some RDTs are not able to detect all species of Plasmodium infections or mixed infections [47]. This can result in an underestimation of the true prevalence in the population. In addition, in these settings, the burden typically

shifts away from young children and into older populations, which these surveys tend not to capture as effectively [18]. Further, as malaria transmission decreases, it becomes more focalized in nature, and population-based surveys are usually not sufficiently powered to provide robust estimates at lower subnational levels.

Seroprevalence is also collected in some household surveys. Although they are not primary impact indicators, serological indicators of recent and historical exposure to Plasmodium offer another method to assess malaria transmission and can be valuable in impact evaluation [48, 49]. Multiplex bead assays allow blood samples (e.g., from a dried blood spot on filter paper) to be tested against a large number of malaria antigens [50] and can provide information about population-level exposure as well as estimate seroconversion rates from age-seroprevalence curves [51, 52].

Special surveys

Special surveys are a potential data source for measuring malaria parasite prevalence and intervention coverage at lower subnational levels or among specific high-risk populations for malaria (e.g., settings with large mobile or migrant populations). Special surveys can make use of different designs or sampling techniques to provide estimates at a finer geographic scale and to target specific groups of people who are at higher risk of malaria. This includes surveys carried out among easy access groups, such as school children or health facility attendees, or among subgroups attending public health activities, such as vaccination campaigns [53]. These types of surveys are often logistically easier and more affordable to implement and can be suitable for reaching at-risk populations; however, careful assessment of the potential selection bias introduced using this type of design is required [53]. Surveys can also make use of different sampling techniques for specific subgroups, such as respondent-driven sampling or venue-based and time-location sampling. These methods allow for more targeted sampling of subgroups, mainly mobile or migrant populations that are more difficult to reach given their high mobility, their limited access to or use of public health facilities, and, in some settings, a greater likelihood of engagement in illegal behaviors [54-56]. These sampling techniques can also introduce selection bias and may not be representative of the at-risk group as a whole [55].

Health facility surveys

Health facility surveys typically do not provide data for measuring impact, but they can provide an important source of information for contextualizing impact indicators. In other words, these surveys provide information about the health facility environment (e.g., availability of commodities) that may modify or explain intervention coverage trends or results of other impact indicators. Common health facility surveys include the SPA and SARA, both of which gather information to assess the availability of different health services and the extent to which facilities are ready to provide those services [16, 17]. For malaria specifically, these surveys examine the availability of malaria diagnostic and treatment services, antimalarials, and laboratory diagnostic capacity [16]. The SPA often includes measures of quality of care through observations of client consultations [16]. It is important to note that service readiness is not equivalent to quality of care [57]. Health facility surveys that include data collation from facility registers and records could use the collated data to measure impact indicators or as a source of information to assess the quality of RHIS data. Separate quality of care assessments may be useful to include in process evaluations, given that quality of care is correlated with positive health outcomes [58]. These assessments can provide critical context to measured impact.

4.1.3. Health and demographic surveillance sites or sentinel sites

Many countries have set up health and demographic surveillance systems (HDSSs) or sentinel sites to record and monitor longitudinal health and demographic data in geographically defined areas. Data on impact

indicators related to malaria morbidity, malaria mortality, and, in some cases, all-cause mortality are often collected in these sites and can be used to show impact within the geographically defined sites. Because the data in these sites are collected for relatively small, defined populations, the data are not typically generalizable to other populations and therefore cannot be used to demonstrate impact on a broader scale. In addition, given the lower number of malaria cases and malaria-attributable deaths in moderate- and low-transmission settings, these sites may be less useful in collecting these data than national program evaluations. Despite these limitations, data from these sites can be used to contextualize broader findings on impact in a country, either at a national or subnational level.

4.1.4. Verbal autopsy

Verbal autopsy (VA) involves the collection of data on probable causes of death through interviews with primary caregivers or family members of recently deceased individuals on the circumstances and symptoms experienced prior to the death [59]. The information gathered during the interview is interpreted by a trained physician or automated methods (such as data-derived algorithms, expert algorithms, or computer-based modeling) to ascertain the cause of death using the International Classification of Diseases [60]. VA data collection is typically nested with HDSSs, population-based surveys (e.g., DHS), civil registration and vital statistics (CRVS), or a census survey. In countries without a functioning CRVS system and lacking quality mortality data, VA is thought to provide the best alternative method to estimate cause-of-death patterns. VA may be subject to bias, however, and may result in either an underestimation or an overestimation of the malaria burden [61]. Despite these limitations, VA has an acceptable level of diagnostic accuracy in high- and moderate-transmission settings. As countries continue to develop their CRVS systems, VA provides the best intermediary to fill the cause-of-death patterns gap in countries with poor CRVS systems [62-64]. VA, coupled with data from other sources, can provide better malaria mortality estimates to measure program impact.

4.1.5. Civil registration and vital statistics systems

In countries with national-scale functioning CRVS systems, these systems may serve as a good source for malaria mortality and all-cause mortality data. This requires high coverage of malaria diagnostic testing, of performance of medical autopsies, and of recording of cause of death. In these settings, CRVS systems may be a valuable data source for tracking trends in all-cause and malaria-specific mortality by different age groups. Many CRVS systems in low- and middle-income countries, however, are not adequately functioning at the national level and do not have high coverage of autopsies and cause-of-death reporting, and thus are unable to provide reliable mortality estimates. In some settings, the CRVS system may provide useful population estimates that can be used for calculating malaria case incidence and mortality using RHIS data.

4.1.6. Entomological surveillance

Entomological surveillance conducted at sentinel sites or transmission foci can provide valuable information for impact evaluations at the subnational level. The most common entomological impact measures include the sporozoite rate and the entomological inoculation rate, which provide proxy measurements for malaria transmission intensity [9]. The main limitation of entomological surveillance data is that they are relevant for assessing trends over time and impact only within specific sites; they are not generalizable at higher levels. Further, entomological surveillance may be of limited use in settings with low vector density or low incidence rates because of the lack of vectors and parasites, which results in less precise and unreliable estimates [9, 49]. Designing and implementing an entomological survey with probability sampling is difficult in areas with low vector density. Other entomological data on vector species composition and abundance, vector biting behavior, and vector insecticide resistance can provide useful contextual information for impact evaluation,

including information on whether effective vector control interventions were deployed in the country [9]. This information is valuable in providing context to understanding the causal pathway for impact and specifically for interpreting impact evaluation findings.

4.2. Summary of recommended impact indicators

A summary of the recommended impact indicators is presented in Table 2, which includes information on the appropriate transmission settings, potential data sources, relevant stratifications, strengths, and limitations and potential biases for each indicator. It is important to note that although microscopy and RDTs continue to be the recommended diagnostic method for identification and management of clinical malaria, alternative tools have been developed that detect parasite DNA, antigen, or antibodies produced in response to Plasmodium exposure. A range of polymerase-chain reaction (PCR) assays are known to be able to identify low-density infections that may be missed by RDT or microscopy, and they have increasingly been used in research in low-transmission settings [65]. Loop-mediated isothermal amplification is a method that has comparable sensitivity to standard nested PCR assays and is feasible in low-resource settings [66]. Serological indicators can provide information about population-level exposure as well as estimate seroconversion rates from age-seroprevalence curves [51] and can be valuable in impact evaluation [48, 49].

These new diagnostic tools and indicators continue to be refined and evaluated in different settings and are expected to be particularly valuable for impact evaluation in moderate- and low-transmission settings. We encourage readers to examine the recent literature for current best practice in detecting parasite DNA or anti-parasite antibodies and to consider including these additional indicators in impact evaluations.

Table 2. Summary of impact indicators for malaria programs

	Indicator	Transmission setting	Data sources	Stratification	Strengths	Limitations/potential biases
1	Malaria case incidence: number and rate per 1,000 people per year*	Moderate, low	RHIS, CHIS, HDSS/ sentinel sites	Subnational area (e.g., district or health facility catchment area), age group	• Collected routinely • Available at subnational level	• Challenge of defining health facility catchment area population size • Potential fluctuations in health facility use over time • Changes in health facility reporting rates and incomplete recording over time • Changes in availability of diagnosis tests over time • Private sector and community data may not be captured in the RHIS • Quality (accuracy and completeness) of RHIS data • Incorporation of malaria cases detected through active surveillance
2	Malaria test positivity rate*	Moderate, low	RHIS, CHIS, HDSS/ sentinel sites	Subnational area, age group	• Collected routinely • Available at subnational level	• Potential fluctuations in health facility use over time • Seasonal fluctuations in prevalence of other febrile illnesses (influencing frequency of malaria testing) • Changes in health facility reporting rates and incomplete recording over time • Changes in availability of diagnostic tests over time
3	Proportion of admissions for malaria	Moderate, low	RHIS, HDSS/ sentinel sites	Subnational area, age group	• Collected routinely • Available at subnational level • Provides information on the level of severe malaria	• Potential fluctuations in health facility use over time • Changes in health facility reporting rates and incomplete recording over time

	Indicator	Transmission setting	Data sources	Stratification	Strengths	Limitations/potential biases
4	Malaria mortality: number and rate per 100,000 people per year	Moderate, low	RHIS, HDSS/ sentinel sites, CRVS	Subnational area, age group	• Collected routinely • Available at subnational level	• Potential fluctuations in health facility use over time • Changes in health facility reporting rates and incomplete recording over time • Underreporting of malaria deaths or misclassification of malaria deaths • Deaths in the community are not captured; deaths occurring in private sector facilities may not be captured in RHIS • Deaths occurring in public facilities may not be representative of all malaria deaths
5	Proportion of inpatient deaths due to malaria	Moderate, low	RHIS, HDSS/ sentinel sites	Subnational area, age group	• Collected routinely • Available at subnational level	• Potential fluctuations in health facility use over time • Changes in health facility reporting rates and incomplete recording over time • Underreporting of malaria deaths or misclassification of malaria deaths • Does not capture deaths occurring in the private sector or in the community • Does not capture fluctuations in deaths for other reasons (outbreak, displacement, conflict)
6	All-cause child mortality (Number of child deaths per 1,000 live births)	Moderate	Population-based household survey (DHS, MICS), census data, CRVS, HDSS/sentinel sites	Age group	• Robust estimates for national level	• Not collected routinely • Many factors may influence child mortality • Retrospective nature of data (estimates more robust for more recent deaths) • Precise estimates not available for subnational level • Does not capture fluctuations in deaths for other reasons (outbreak, displacement, conflict) • In settings approaching lower transmission, all-cause child mortality will no longer be an

	Indicator	Transmission setting	Data sources	Stratification	Strengths	Limitations/potential biases
						appropriate measure of impact due to the low proportion of deaths due to malaria
7	Annual number of malaria epidemics	Moderate, low	RHIS, program data	Not applicable	• Collected routinely • Available at subnational level	• Challenge of defining an epidemic (threshold of malaria cases beyond what is considered normal) • Contextual (external) factors may influence epidemics
8	Parasite prevalence*	Moderate, low	Population-based household survey (DHS, MIS, MICS)	Subnational area, age group	• Provides information at the population level • Estimates typically available at the first subnational level (e.g., region or province)	• Can fluctuate a lot within a year, particularly in areas with seasonal transmission • Typically available for children under five and not entire population • Not collected routinely; typically few data points to assess trends over time • Submicroscopic infections, which are more common in low-transmission settings, will not be detected using microscopy and RDTs
9	Seroprevalence	Low	Population-based household survey	Subnational area, age group	• Greater sensitivity to detect changes and variations in malaria transmission over time [67] • Measurement reflects exposure of an extended period; not prone to seasonal fluctuations [67] • Estimates available at subnational level	• Not collected routinely; typically few data points to assess trends over time • Methodology not entirely worked out

CHIS=community health information system

*Disaggregated by vector species, if possible

4.3. Contextual factors

4.3.1. Types of contextual factors

To evaluate the relationship between malaria interventions and impact indicators, it is essential to determine the degree to which contextual factors could explain the observed changes in the impact indicator. Contextual factors may modify the effects of malaria interventions and can confound the association between interventions and their health impact. Considering the contextual factors is therefore likely to increase the validity of evaluations [68].

Types of contextual factors that countries may need to consider during an evaluation include the following (Figure 1):

- **Health system factors.** Per capita expenditure on health, government expenditure on health as a percentage of total government expenditure, availability of drugs and commodities, quality of commodities, intervention delivery quality, and population access to and use of health interventions and services
- **Sociocultural and socioeconomic factors at the micro and macro levels.** Household assets and income, housing construction, gender, parental education, political situation and stability, GDP per capita, country's transport and communication infrastructure, proportion of the population living below the poverty line, migration, and crises and conflicts
- **Climate factors.** Attributes such as total precipitation, intermittency of rainfall, air temperature, and extreme weather events such as floods
- **Environmental factors.** Characteristics such as land cover and vegetation, and altitude
- **Epidemiological factors**. Prevalence of other diseases (e.g., HIV) and particularly other febrile illnesses, outbreaks of malaria in neighboring countries, outbreaks of other non-malaria diseases, immunity, HRP2 gene deletion, and malnutrition (stunting and growth retardation) in the population at risk for malaria

Climatic and environmental factors can be sourced from national meteorological agencies and from satellite-derived products produced by agencies such as the National Oceanographic and Atmospheric Administration. Satellite-derived products are often available at high resolution (often less than 5km^2); this is likely to be more appropriate for an evaluation in moderate- and low-transmission settings, in which subnational analysis will be performed, than data from national meteorological agencies, which may provide data for specific weather stations or as national-level summaries only. A full description of the impact of variations of climatic and environmental variations on malaria transmission is presented in the evaluation guidance for high-transmission countries [5].

Commonly used satellite-derived climate and environmental data sources are as follows:

- The U.S. Geological Society Famine Early Warning Systems Network has easy-to-navigate rainfall and vegetation indicators for Africa, Central America and the Caribbean, Central Asia, the Middle East, and South Asia: https://earlywarning.usgs.gov/fews.
- The National Aeronautics and Space Administration Moderate Resolution Imaging Spectroradiometer includes temperature, vegetation indices, and land cover types: http://modis.gsfc.nasa.gov.

- The National Oceanographic and Atmospheric Administration Advanced Very High Resolution Radiometer has vegetation and land cover characterization data at https://www.usgs.gov/centers/eros/science/usgs-eros-archive-avhrr-normalized-difference-vegetation-index-ndvi-composites?qt-science_center_objects=0#qt-science_center_objects and temperature data at http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html.

Key migration-related factors of interest include within-country human population movement (HPM), cross-border HPM, the seasonality of HPM, and sociodemographic breakdown of HPM. HPM data can be derived from a combination of data sources, including census data, household surveys (e.g., migration and remittances surveys, labor force surveys, World Bank Living Standard Measurement Surveys, DHS, MIS), small-scale or sub-population surveys, phone records, hospital records (e.g., travel histories taken from hospitalized patient records), and surveillance systems [69, 70].

The DHS, MICS, and MIS include questions on non-malaria programs and factors of interest, including socioeconomic status, health and nutritional indicators, coverage of health care services and care-seeking behavior, and immunization coverage. Specific guidance for analysis of indicators from DHS, MIS, and MICS datasets are presented in the relevant survey documentation on the institution websites. Other sources of contextual factor data include WHO and World Bank reports, the United Nations Children's Fund, the Joint United Nations Programme on HIV/AIDS, and country-specific reports. Additional data sources may include other country-specific surveys and datasets, such as HDSS data, and data identified through discussion with the MOH and other programs in the MOH, as well as other partners with knowledge of the country's health system.

During the design of an evaluation, it is important to list potentially relevant contextual factors and to include as part of the analytic plan an assessment as to whether and to what degree changes in these factors could have affected intervention coverage and impact indicators. Section 5 provides further discussion and examples of how to account for contextual factors in the analysis.

4.3.2. Organizing contextual factors for analysis and interpretation

In preparation for analysis and interpretation, contextual factors should be used to generate a causal diagram to describe the hypothesized relationship between the contextual or confounding factors and program coverage and impact indicators. For example, data on total precipitation are in the category of climate factors, and inclusion of these trend data in the evaluation is justified because they affect mosquito breeding and malaria transmission and thus may cause some of the observed changes in outcomes over time or geography [5]. For example, countries in the Greater Mekong Subregion experiencing an increase in malaria incidence along their borders should examine HPM, climate factors, epidemiological factors, and health-seeking behaviors, as well as seek to collaborate with neighboring NMPs to review data and trends. It is also important to consider anomalies (e.g., extreme weather events and political instability) and qualitative contextual data for a comprehensive evaluation.

Quantitative contextual data can be analyzed using standard statistical methods. Basic univariate analyses can be used to examine trends in contextual data (e.g., climate data) against trends in key malaria indicators to determine whether trends are similar or inverse. Statistical methods may be used in multivariate analyses to assess confounding and effect modification [5, 71]. Qualitative contextual data, including data that are more anecdotal (e.g., security event, or incidence of conflict, that disrupted delivery of interventions), can be used to inform a plausibility argument.

5. DATA ANALYSIS, SYNTHESIS, AND INTERPRETATION

5.1. Summary of study designs and methods

The choice of analytic method should be determined by a combination of the evaluation questions, the implementation approach used in the area under evaluation (e.g., phased introduction), data availability (temporal and spatial resolution, quality), and the indicators available for inclusion. The following sections briefly describe the main analytic approaches that may be relevant for impact evaluation in moderate- and low-transmission settings, with illustrative examples from the literature or country evaluation experiences. These analytic approaches are not necessarily specific to low- and moderate-transmission settings only, but they are presented to demonstrate the breadth of analyses that are possible with the types of data likely to be available in these settings. Readers are encouraged to seek further information about the analytic approaches from the referenced sources or by contacting evaluation specialists for additional support. In addition to these analytic approaches using quantitative data, impact evaluation also requires qualitative data, such as a narrative description of the program over the evaluation period, timeline of key activities, policy changes, or other contextual factors. This is described further in Section 6.1.2.

In selecting an analytic approach, it is important to consider the internal and external validity of different study designs and analytic approaches. Internal validity refers to the ability to attribute differences in the observed outcome between intervention and control groups to the program or intervention being investigated. External validity refers to the extent to which the findings from the study can be generalized to other populations or locations. Most designs discussed in this document are of the quasi-experimental type [26].

Table 3. Summary of study designs and methodologies

Methodology/study design	When is it useful?	What types of data can be used?	How robust is the design?
Interrupted time series	Policy change or other intervention introduced on a known date. Useful when no underlying contemporaneous control group, but can be adapted to include a control group.	Time-series data (retrospective or prospective), ideally RHIS	Good. Considers trend and confounding factors, counterfactual can be estimated.
Dose-response	When no clear intervention and comparison areas, but intervention at varying levels of intensity by district	Subnational data (e.g., district-level) describing intervention, impact indicator, and potential confounders. Ideally RHIS. Requires data on process and activities to define "intensity."	Moderate, if high spatial and temporal resolution and confounders included. Can estimate counterfactuals for alternative program coverage levels. Prone to confounding because intensity of intervention or program applied may be related to impact outcome.
Constructed controls (matching or discontinuity designs,	When no clear intervention and comparison areas, but differences in individual use and access to	Individual-level data from cross-sectional survey data with large sample size, and all possible	Moderate. Limited by availability of data from which to estimate controls. Often uses data from a single cross-sectional

Methodology/study design	When is it useful?	What types of data can be used?	How robust is the design?
instrumental variables)	interventions, or eligibility criteria determine whether an individual or area received interventions. Useful for inference at the individual level.	confounders measured	survey, and evaluation may have low power to identify changes where cross-sectional RDT positivity is the primary impact indicator.
Stepped-wedge	Phased introduction of program with or without randomization	RHIS or repeat cross-sectional surveys	Moderate. Important to account for other programs or contextual changes occurring during the phased roll-out of program being evaluated.

5.2. Interrupted time series

The ITS approach requires longitudinal datasets, such as surveillance data, which are often aggregated to specified units of time and space (e.g., surveillance data summarized by health district and month). The ITS approach involves comparison of the level and mean trend in outcome indicators before and after a breakpoint [38], which could be the introduction of a new intervention or a change in policy. The ITS method is generally used to evaluate changes in outcome due to interventions introduced over a short time period with consistent intensity [72, 73]. It can also be adapted to incorporate lags between intervention and effect on outcome or to include roll-out periods by increasing the number of breakpoints in the time series. Evaluators should be aware that increasing the number of breakpoints in ITS analyses does increase the probability of observed changes in outcome being attributed to confounding factors [74]. ITS analysis must account for secular trends in the statistical analysis (in fact, the ability to account for secular trends is a major strength of ITS) and can also directly incorporate data on potential confounders, such as climate and environmental data.

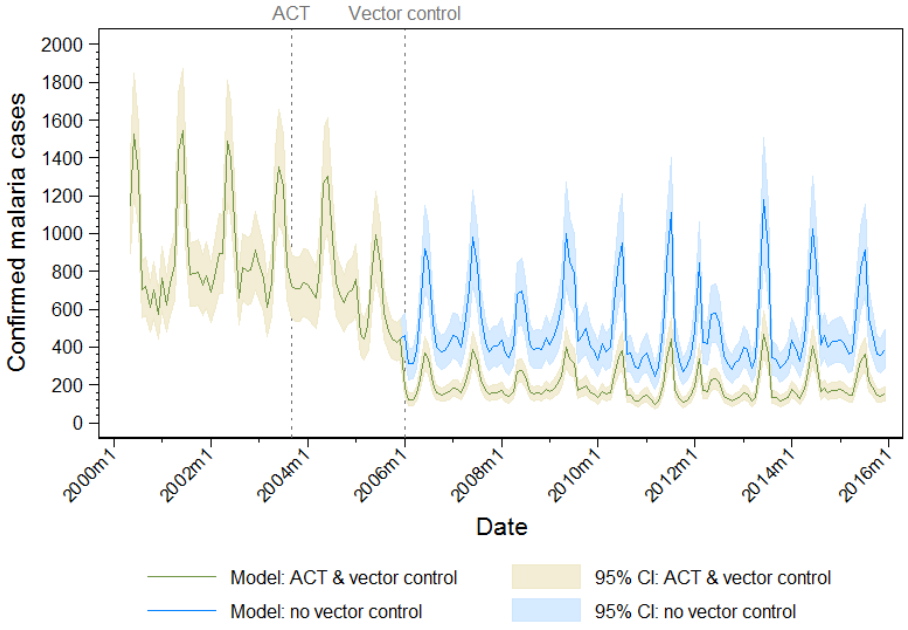
ITS analyses can be performed on areas that received the program activities, and in equivalent comparison areas, to describe the changes in level and mean trend in outcome between the program and control areas. Without a contemporaneous control group, such designs usually require investigators to have measured a long period of pre-intervention outcomes to model underlying temporal trends and seasonal variables. When no equivalent control area exists, a counterfactual can be estimated by fitting the ITS model and then predicting the outcome with just a continuation of the baseline level and trend (and other covariate data). The modeled counterfactual must be interpreted with contextual information to check the plausibility that changes could be explained by factors other than the program being evaluated. A number of publications discuss the design and implementation of ITS designs using segmented regression models in more detail [41, 75-77].

5.2.1. Example: Evaluating changes in malaria incidence in Zanzibar over 16 years

In Zanzibar, routinely reported surveillance data from public health facilities were used to estimate the confirmed malaria incidence each month from 2000 to the end of 2015 [78]. Over this period, substantial decreases in malaria incidence were observed. Artemisinin-based combination therapies (ACTs) were introduced in late 2003, and a combination of IRS and mass long-lasting ITN distribution began in 2006. The introduction of these interventions at the same time across Zanzibar, as well as the availability of outcome data (confirmed malaria incidence) for several years prior to the introduction of ACT, made using an ITS approach to fit a segmented regression model particularly appropriate for this impact evaluation. The study found evidence for a decrease in malaria incidence during the period of ACT roll-out compared to pre-

intervention, and that the decrease continued, but at a slower rate, following introduction of vector control. Figure 3 uses vertical dashed lines to indicate the three periods of the ITS: (1) prior to introduction of ACT as first-line treatment; (2) during the period of ACT availability but prior to vector control scale-up; and (3) during the period of ACT availability, mass distribution of ITNs, and implementation of IRS.

Figure 3. Graph describing the modeled number of confirmed malaria cases occurring across Zanzibar (green line) with 95 percent confidence interval (green shading), and a modeled counterfactual predicting the number of cases that would have occurred if vector control had not been scaled up in 2006 (blue line) with 95 percent confidence interval (blue shading)



CI=confidence interval

5.2.2. Example: Evaluating changes in health facility use during introduction of iCCM in Uganda

An ITS design was used in a district of Uganda to evaluate the impact of CHW scale-up on OPD attendance by children under five [79]. Prior to the introduction of CHWs in this district in May 2010, there were no other community case management services operating or providing malaria diagnosis and treatment services. The authors describe the division of the time series into pre-intervention and intervention, and the specification of their segmented regression model used to estimate whether there was a change in slope and level of the outcome (OPD visits) following introduction of CHWs, after accounting for the overall secular trend. The study found a 64 percent decrease in malaria OPD visits following CHW introduction but found no evidence for any change in numbers of non-malaria OPD visits.

5.3. Dose-response

Dose-response studies, also termed a "national-evaluation platform approach," make use of impact indicators available at a subnational level (e.g., district) and a varying intensity of the intervention or program at a subnational level to examine a dose-response relationship between the intervention and impact indicator, ideally over a period of time [39, 80]. As a result, this approach is suitable in settings in which high spatial and

temporal resolution data are available, and where similarly detailed information can be found at a subnational level describing the intervention intensity. Examples of "dose" indicators could be coverage of ITNs, or proportion of health facilities providing effective malaria case management, and the "response" is generally the impact indicator (e.g., confirmed malaria incidence). By specifically investigating associations between varying intervention intensity and impact indicator, there is no need for an observed control area that did not receive the intervention to serve as a counterfactual. Counterfactuals can be generated to describe alternative scenarios (e.g., of lower intervention coverage, if a net distribution had not been performed) by simply fitting the model and predicting the outcome using the estimated intervention or program coverage for the counterfactual scenario.

5.3.1. Example: Impact of malaria control activities in Zambia

This study used district-level surveillance data from a period of rapid diagnostic and reporting scale-up to evaluate the association between ITN program intensity and monthly confirmed malaria incidence [43]. The analysis was conducted at the district level by month and used a geostatistical model to estimate longitudinal ITN coverage from a combination of survey and programmatic data. The model exploring the association between ITN coverage and monthly malaria incidence included confounders such as climate variables, reporting, testing, and treatment-seeking, and it accounted for spatial and temporal heterogeneity. The model estimated that an increase in district ITN coverage of one net per household was associated with a 27 percent reduction in overall confirmed malaria case incidence; it was associated with a 41 percent reduction in case incidence in areas of lower malaria burden.

5.3.2. Example: A district-level ecological analysis between household ITN coverage and ACCM in Malawi

This analysis took advantage of two household surveys in Malawi (MICS 2006 and DHS 2010) that had very large sample sizes and were consequently weighted to produce valid district-level estimates of household ITN coverage. Most standard national surveys such as the DHS and MIS are only powered to generate regional estimates or estimates by ecological or risk zone strata. The analysis used two different methods to generate district-level household ITN ownership estimates (dose) for each year from 2006 to 2010, based on estimates from each cross-sectional survey. The impact indicator (response) of ACCM was generated from birth history data in each cross-sectional survey (note that ACCM is not recommended as an impact indicator in low-transmission settings). The study found that higher levels of ITN ownership were significantly associated with lower ACCM [81].

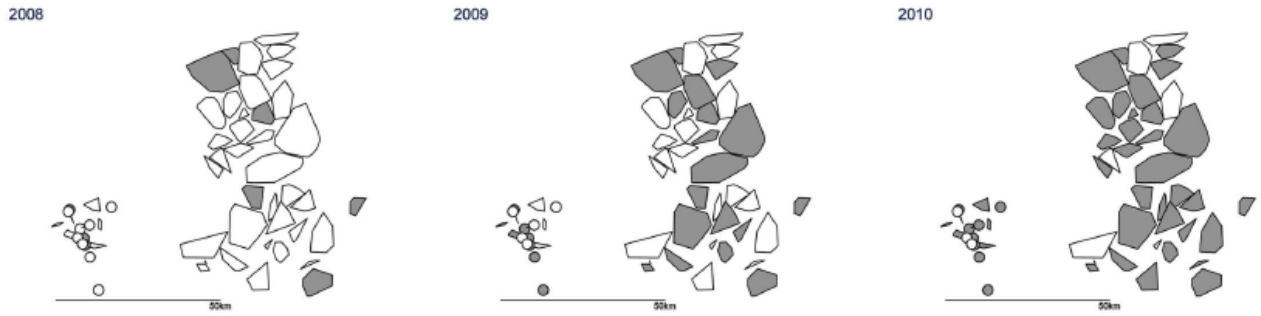
5.4. Stepped-wedge

In settings in which an intervention has had a phased introduction, whereby different locations receive the intervention at different times, it is possible to take advantage of this staggered deployment by comparing intervention and non-intervention areas using a stepped-wedge design (also sometimes called a multiple baseline design). If a country or program wants to generate evidence of impact of an intervention or program being introduced, implementation units can be randomized to each phase of roll-out to maximize the internal validity of this design [40, 82]. This approach takes advantage of the logistical justification to conduct a phased introduction of new interventions or programs and addresses ethical concerns by ensuring that all populations receive the intervention; however, data are required at each "step" of the roll-out. A DiD analysis is often used for this type of design (Section 5.5.1). Analysis of stepped-wedge designs can become complex, and it is recommended that specialized statistical support be sought to ensure that analysis is appropriate.

5.4.1. Example: Introduction of SMC in Senegal

A study in Senegal aimed to evaluate the cost, acceptability, and impact on mortality of SMC [83]. As a new intervention without an established delivery channel, a stepped-wedge evaluation design enabled a gradual introduction of the intervention. The evaluation team considered that an uncontrolled pre-post design would be inappropriate due to variability in malaria transmission intensity over time. The SMC intervention was introduced over three years, with the phased introduction schedule determined by randomization of 54 health post areas; 9 areas completed SMC in 2008, an additional 18 began in 2009, and another 18 began in 2010. The remaining nine health post areas did not begin SMC during the study period and served as controls. The primary trial indicators were all-cause mortality among children 3–59 months of age and RDT-confirmed malaria among outpatients. Parasite prevalence among children 3–59 months of age measured by surveys at the end of transmission season was a secondary indicator. The study found that introduction of SMC was associated with a 60 percent decrease in malaria incidence among children under 10, but no differences in ACCM were observed (note that ACCM is not recommended as an impact indicator in low-transmission settings) (Figure 4).

Figure 4.Overview of the stepped-wedge design used by Cisse, et al. in Senegal, whereby SMC was introduced in phases to 45 health posts over 3 years



Year	Zone# 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
2008	SMC*					
2009	SMC	SMC	SMC			
2010	SMC	SMC	SMC	SMC	SMC	

Fig 2. Stepped-wedge introduction of SMC in 45 health post areas over three years. The upper maps show the health district boundaries (the district of Fatick was subdivided in 2010 to create a new district of Niakhar); land cover and water bodies; and the location of villages and health facilities. In the lower maps, the polygons show the catchment areas of the 54 health posts (drawn as the convex hull of the village coordinates). # Each zone comprised nine health post areas. The nine health post areas in Zone 6 remained untreated. * In 2008, SMC was delivered to children aged 3–59 months; in other years, SMC was provided for children aged 3–119 months.

doi:10.1371/journal.pmed.1002175.g002

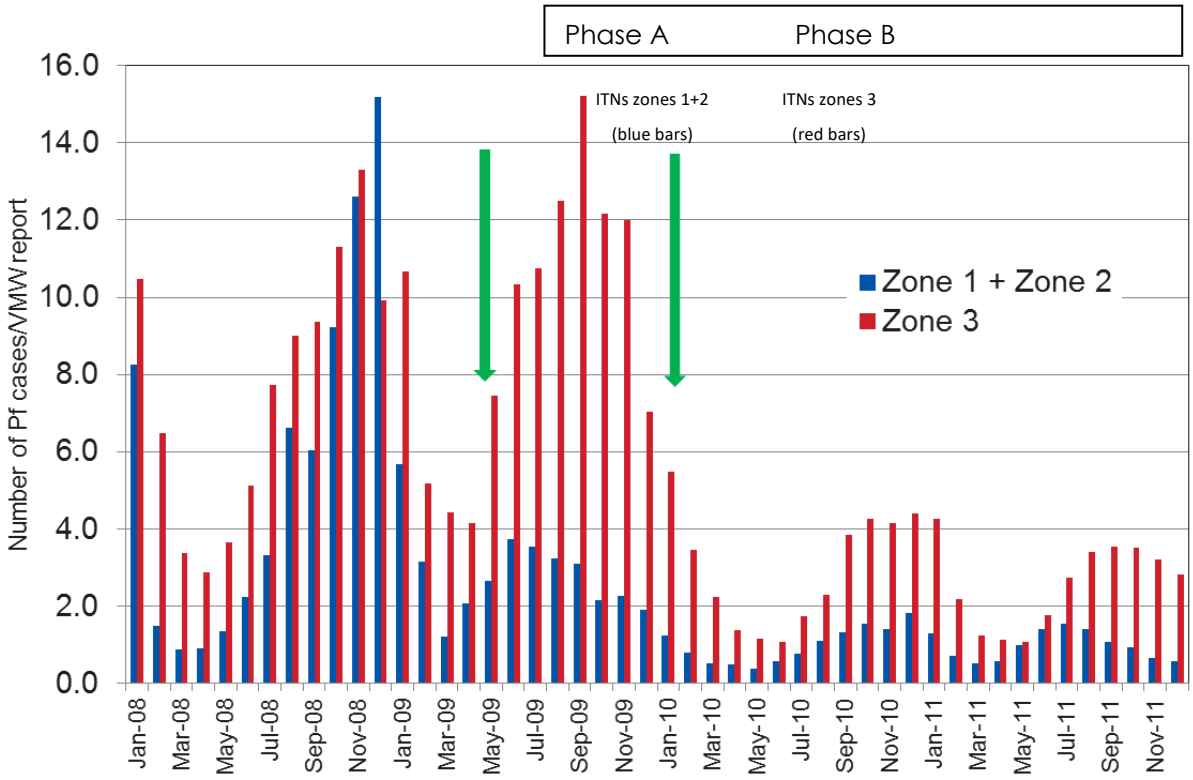
Note: The map indicates the phased introduction to health zones from 2008 to 2010.

Source: Cisse B, Ba EH, Sokhna C, JL ND, Gomis JF, Dial Y, Pitt C, M ND, Cairns M, Faye E, et al: **Effectiveness of Seasonal Malaria Chemoprevention in Children under Ten Years of Age in Senegal: A Stepped-Wedge Cluster-Randomised Trial**. *PLoS Med* 2016, **13:**e1002175.

5.4.2. Example: Impact of phased ITN distribution and village malaria worker introduction in Cambodia

As part of the Cambodia Malaria Program Review 2012, an impact assessment was introduced to measure coverage of key interventions over time at national and subnational levels and to quantify the relationship between resources and changes in intervention coverage using statistical models. Initial descriptive analysis in Cambodia provides a compelling illustration of an association between phased mass ITN distribution and estimates of malaria burden from the passive surveillance system, adjusted for the changes in diagnosis coverage and reporting practices due to the roll-out of village malaria workers (trained community members who provide malaria diagnosis and treatment) over time. A large reduction in the number of cases reported in zones 1 and 2 occurred after these areas received ITNs in May 2009 (Phase A), but zone 3 case reports during the 2009 transmission season were at a level similar to the previous year (Figure 5). Following the ITN distribution in zone 3 in early 2010 (Phase B), the number of cases reported in zone 3 decreased significantly compared to the previous two years.

Figure 5. Bar chart displaying average number of P. falciparum cases reported by each village malaria worker, each month, in areas receiving the first phase of interventions (zones 1 and 2, blue) and in zones receiving the second phase of interventions (red)



5.5. Analytic techniques relevant to impact evaluation in low-transmission settings

Table 4. Summary of major analytic techniques

Technique	How does it work?	What type of question can it answer?	What type of settings is it appropriate?
Difference-in-differences	Compares before-after changes in intervention group with before-after changes in a comparison group	Whether a pilot program in district A had an impact on malaria incidence, compared to an equivalent district B, which did not receive the pilot program	Only where the "parallel trends" assumption holds, meaning that in the absence of the intervention, both groups would have equal trends
Instrumental variables	Uses another variable (the instrument), which is associated with the intervention or program, but not with the outcome, to account for unobserved correlation between intervention and impact	Estimating the impact of an intervention in the context of endogeneity (e.g., when allocation of the intervention is associated with the impact indicator)	When there is time-varying selection bias (differences between recipients and non-recipients that changes over time). In practice, it is often challenging to find an appropriate variable to include as the instrument.
Matching to construct controls (exact, group, and propensity score matching)	Uses data on observed individuals who did not receive the intervention to estimate the impact of the intervention at the population level	Estimating the impact of an intervention (with imperfect adherence or uptake) under routine conditions, using survey data	Where individual-level data are available on the intervention, the impact indicator, and a sufficient number of other indicators to define matching

5.5.1. Difference-in-differences

The DiD estimator compares changes in the impact indicator over time between a population that received the intervention and a population that did not receive the intervention, in cases in which interventions are not randomly assigned or based on cut-off values of another indicator [84]. DiD is commonly used in conjunction with stepped-wedge designs or other pre-post comparisons using cross-sectional survey or time-series data, and it allows the evaluation to take into account underlying trends in the outcome level over time [38]. The main limitation of DiD is the "parallel trends" assumption; that is, the trend in the impact indicator over time would be the same in both the intervention and comparison areas if there was no intervention. The DiD method attributes any differences in trend between intervention and comparison areas to the effect of the intervention, and, as a result, this method will generate biased estimates of intervention impact if any other factors are present that influence the impact indicator and differ between intervention and comparison areas. DiD analysis has been used to investigate the impact of IRS in Mali [85] and to complete a multi-country analysis of the impact of President's Malaria Initiative activities on ACCM [86].

An alternative to the parallel trends assumption is that potential outcomes are independent of treatment status, conditional on past outcomes, using the synthetic control method. The synthetic control method allows for effects of unobserved variables to change over time by constructing a comparator as a weighted average of the available control units [87, 88].

5.5.2. Use of instrumental variables to address endogeneity

Instrumental variable methods involve finding a variable (the instrument) that is highly correlated with uptake or allocation to the intervention or program of interest and is not correlated with unobserved characteristics affecting the impact indicator [89]. Instrumental variable methods allow for endogeneity in allocation to or participation in the intervention or program (e.g., evaluating the impact of malaria programs including IRS if IRS was targeted to districts according to malaria incidence). When using panel data (e.g., observing at the district level over time), instrumental variable methods can allow for time-varying selection bias (e.g., where there are systematic differences between intervention recipients and non-recipients, and these differences change over time), which is not possible with methods such as DiD [89].

The main limitation is in identifying a good instrument: the instrument must be strongly correlated with allocation to or uptake of the intervention or program, but it cannot be correlated with the outcome through an effect on other variables. Use of instruments that are correlated with unobserved characteristics affecting the outcome will lead to biased effect estimates, and the use of instruments that only weakly correlate with the intervention variable will result in reduced precision of the impact estimate.

Analysis using instrumental variables is particularly relevant in dealing with endogeneity and widely used in the econometrics and broader public health literature. Few examples of analysis using instrumental variables exist in the malaria literature, however, due to the difficulty in identifying appropriate instruments. One example of the use of instrumental variables was in an evaluation of vector control in the Solomon Islands, in which calendar month was the instrumental variable [90]. The researchers justified the use of month (which is often strongly associated with malaria incidence) as an instrumental variable by also including rainfall in regression models and explaining that after accounting for rainfall, there was no reason to expect month to be associated with malaria incidence.

5.5.3. Matching methods to construct controls

In settings in which an observed comparison area that did not receive interventions is unavailable, but individual-level data are available from both individuals who did and did not receive the intervention, matching methods can be used to simulate a statistical comparison group [89]. The aim of matching methods is to generate a comparison group that is as similar as possible to the intervention group in terms of the observed variables available; the impact of the intervention is defined as the average difference in outcomes across the two groups. Exact matching, group matching, and propensity score matching are all potential methods to construct a control from available data.

Exact matching has been used in a multi-country analysis of the impact of net use using cross-sectional survey data, whereby children using nets were matched to children not using nets using data on age, mother's education, urban or rural residence, and malaria transmission intensity [91]. Logistic regression analysis was then performed on the matched dataset for each survey, including the matching factors and other variables as potential confounders.

Propensity score matching uses a two-stage regression approach. First, a regression model is used to predict the probability of receiving the intervention for each analysis unit (e.g., person, household, health facility catchment area) using a set of observed variables (e.g., confounders of the intervention and outcome). This propensity score is then used to match those who received and did not receive the intervention, or the inverse of these propensity scores can be used to create weights for each analysis unit [89]. A second regression is used to model the outcome based solely on the intervention, using only the matched subjects (for propensity score matching) or weighting each analysis unit by its inverse probability weights. Propensity score methods work best with large sample sizes, when the intervention is common but the outcome is rare, and

investigators can assume that no further unmeasured confounding variables exist that predict the propensity of receiving the intervention or are strongly correlated with the outcome of interest [74]. Propensity score matching has been used in evaluations of behavior change messaging for net use [92] and impact of SMC [93].

5.5.4. Advanced techniques to estimate impact and causal effects

Regression discontinuity is a method that until now has not been widely used in public health evaluation, and it is relevant in situations in which receipt of the intervention program is determined by placement above or below an arbitrary cut-off value for an assignment value [84]. The cut-off value should not be decided by biological plausibility or known interaction. Regression discontinuity analysis assumes that individuals immediately above and below the cut-off score are similar in measured and unmeasured confounders [94]. The impact of the intervention is estimated by generating regression lines for each group (intervention and non-intervention) and then comparing the level and slope of the regression lines between the two groups at the cut-off value of the assignment variable [74].

Although traditional regression analyses seek to describe how an outcome is associated with an exposure and mediating factor, statistical mediation analysis aims to understand and quantify the relative magnitude of different possible mechanisms by which an exposure effects an outcome [95, 96]. Recent developments in mediation analysis within the causal inference literature have allowed for the estimation of causal effects through the use of causal path analysis and structural equation modeling [97]. Mediation analysis is particularly appropriate where there is an interest in understanding how much of the exposure-outcome relationship is explained by the mediating pathway, which may be particularly relevant if the investigator has the opportunity to intervene at the level of the mediator. Mediation analysis and other new causal inference methods remain a subject of debate in the literature, with disagreement about the extent to which these methods permit researchers to infer causality in the absence of classical experiments involving randomization [98, 99].

Oaxaca-Blinder decomposition is a method that has been frequently used in economics literature to describe gender gaps in wages, health and racial disparities in health outcomes, and environmental epidemiology [100]. The method combines cross-sectional conditional correlations with information on changes in the underlying variables, to present a statistical decomposition of changes in the mean of an outcome variable. The purpose of such analyses is mainly descriptive and is used to explore potential mechanisms, because strong assumptions are required to estimate causal relationships. If causal relationships can be adequately identified, results from the Oaxaca-Blinder may aid in guiding policies that can reduce health disparities. An example from Rwanda uses decomposition analysis to describe interventions and contextual factors contributing to decreases in ACCM [101].

5.6. Linking process and impact evaluation findings

As discussed in the theory of change (Section 2.1), linking process evaluation findings with impact evaluations can provide for more meaningful and actionable impact evaluations. Impact evaluations by their nature focus on examining the relationship between the coverage of interventions (outcomes) and the desired impact of the program. Without process evaluation findings, it can be unclear as to why the program has or has not achieved its impact. Process evaluations provide this critical information on the *why* and *how* a program worked, and therefore are valuable in providing the necessary context to elucidate the relationships between intervention implementation and achieved outcomes and impact. This linkage of qualitative information from

process evaluations with impact findings is crucial to allow NMPs to generate evidence for decision-making and to identify actions required for continued program improvement and impact.

In an impact evaluation of a national program, the linkage of process evaluation findings becomes even more important, because the evaluation examines the impact of a package of interventions, rather than assessing the impact of a specific intervention. Methods such as decomposition analysis can assist in exploring which interventions and contextual factors contribute to observed changes in an impact indicator [101]. Many impact evaluations also encounter data availability constraints, particularly if they rely on survey data. In these situations, there may only be a few data points to examine changes in impact over the evaluation period. It is also important to consider that many of the impact indicators can fluctuate considerably over time. These constraints make it even more challenging to interpret why the program did or did not achieve its desired impact. For these reasons, understanding what the program actually entailed and the quality and intensity of intervention implementation over the evaluation period can help better contextualize the impact evaluation results and understand which may have contributed more (or less) to any observed impact.

In situations in which a program is found to have no impact, or a negative impact, findings from a process evaluation can help determine whether this was due to an issue in the program's theory of change or issues with program implementation [102]. For example, in the situation of negative program impact on malaria incidence, reviewing complementary data on therapeutic efficacy and insecticide resistance monitoring may also be key in determining the plausibility of impact evaluation findings. Lastly, process evaluations can provide valuable information on factors that positively affected or hindered implementation, which can be useful for translating the implications of the findings for other settings [102], in addition to improvement of the program.

5.7. Building a national-level impact narrative

Where impact evaluation analysis has been stratified according to differential risk areas, risk populations, or different intervention packages, it is often valuable to compile these individual findings into a descriptive overall narrative at national level. This national-level narrative may be particularly relevant to advocate continued funding of the malaria program, both from national and international sources.

A common tool used to compile multiple findings and indicators is the use of a scorecard. Scorecards are already used by initiatives such as the "Elimination 8" countries aiming to eliminate malaria in southern Africa and by the Asia Pacific Leaders Malaria Alliance (APLMA) [103, 104]. These scorecards provide a qualitative interpretation (red, yellow, green) of key indicators. Classifications of impact should be decided according to local targets and the type of impact evaluation performed. Classifications could take the following forms:

- Red: evidence for negative impact (malaria increase)
- Grey: no evidence for impact
- Yellow: weak evidence for positive impact
- Green: strong evidence for positive impact

Where possible, scorecards should be harmonized with existing tools. This can improve the interpretation of impact evaluation findings, along with indicators relating to malaria program performance and processes that are being tracked over time.

In addition to a scorecard, a short descriptive summary should be prepared that is easily understood by non-specialist audiences. This can describe the main findings of the impact evaluation in each relevant

stratum or risk population, with the inclusion of any potential unmeasured confounders or limitations in the analysis that may alter interpretation of the findings.

Example scenario: IRS was completed in selected districts, but impact evaluation found no change in malaria incidence compared to unsprayed areas.

In this context, programmatic information about the IRS campaign, as well as supporting entomological information, is crucial to understanding why no impact was observed. Entomological monitoring site data can be used to determine whether the insecticide used is effective against the locally relevant *Anopheles* species. Coverage data, as well as training and supervision data from the IRS campaign, will indicate whether there were operational problems with reaching all sprayable structures and whether spray teams operated according to standard operating procedures (e.g., sufficient insecticide was sprayed on walls). As is recommended for all impact evaluations, additional contextual information should be collected from both intervention and comparison areas to understand other programs that may influence the impact indicator, as well as changes in reporting, access, climate, etc.

5.7.1. Elimination 8 scorecard example

The scorecard shown in Figure 6 has been developed by the alliance of eight countries aiming to eliminate malaria in Southern Africa: Angola, Botswana, Mozambique, Namibia, South Africa, Swaziland, Zambia, and Zimbabwe. The scorecard is completed annually and uses a key set of indicators covering epidemiology, vector control, financing, policy, and program management.

Each indicator is classified as one of the following:

- Red: Not on track
- Yellow: Some progress
- Green: On track
- Grey: No data available or not applicable

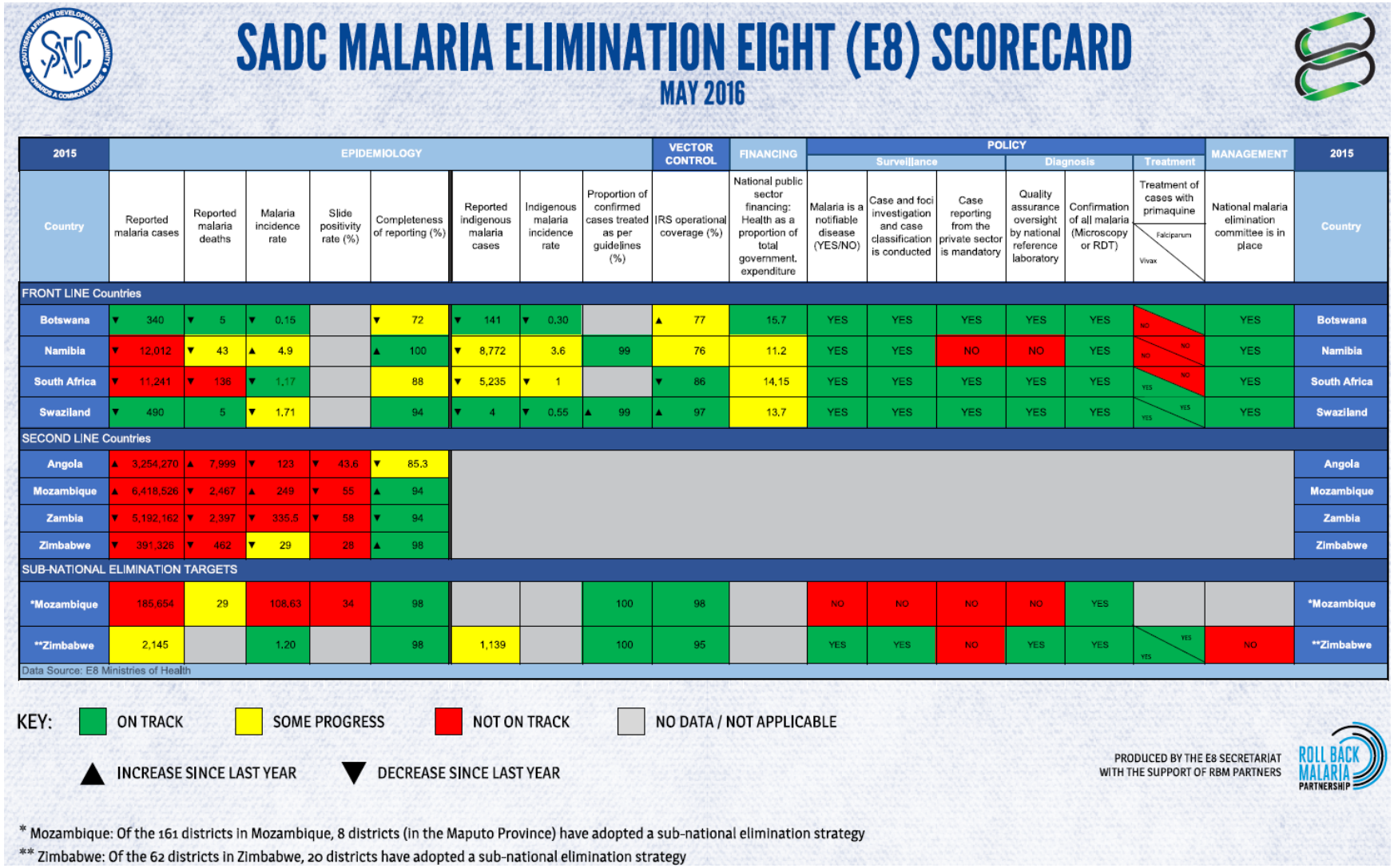
5.7.2. APLMA scorecard example

APLMA is an affiliation of Asian and Pacific heads of government, formed to accelerate progress against malaria and eliminate malaria in the region by 2030. Two scorecards are produced annually to describe progress of APLMA countries (Figure 7). The first has a summary of progress against six priority actions and gives a policy-level view of progress. A second technical scorecard has more detail on specific targets under epidemiology, surveillance, vector control, resistance, financing, and policy themes.

Each indicator is classified as one of the following:

- Red: Not on track
- Yellow: Progress but more effort needed
- Green: On track
- Grey: No data available or not applicable

Figure 6. Malaria Elimination Eight Scorecard



SADC MALARIA ELIMINATION EIGHT (E8) SCORECARD

MAY 2016

2015	EPIDEMIOLOGY								VECTOR CONTROL	FINANCING	POLICY						MANAGEMENT	2015
											Surveillance			Diagnosis		Treatment		
Country	Reported malaria cases	Reported malaria deaths	Malaria incidence rate	Slide positivity rate (%)	Completeness of reporting (%)	Reported indigenous malaria cases	Indigenous malaria incidence rate	Proportion of confirmed cases treated as per guidelines (%)	IRS operational coverage (%)	National public sector financing: Health as a proportion of total government expenditure	Malaria is a notifiable disease (YES/NO)	Case and foci investigation and case classification is conducted	Case reporting from the private sector is mandatory	Quality assurance oversight by national reference laboratory	Confirmation of all malaria (Microscopy or RDT)	Treatment of cases with primaquine (Falciparum / Vivax)	National malaria elimination committee is in place	Country
FRONT LINE Countries																		
Botswana	▼ 340	▼ 5	▼ 0.15		▼ 72	▼ 141	▼ 0.30		▲ 77	15.7	YES	YES	YES	YES	YES	NO	YES	Botswana
Namibia	▼ 12,012	▼ 43	▲ 4.9		▲ 100	▼ 8,772	3.6	99	76	11.2	YES	YES	NO	NO	YES	NO / NO	YES	Namibia
South Africa	▼ 11,241	▼ 136	▼ 1.17		88	▼ 5,235	▼ 1		▼ 86	14.15	YES	YES	YES	YES	YES	NO / YES	YES	South Africa
Swaziland	▼ 490	5	▼ 1.71		94	▼ 4	▼ 0.55	▲ 99	▲ 97	13.7	YES	YES	YES	YES	YES	YES / YES	YES	Swaziland
SECOND LINE Countries																		
Angola	▲ 3,254,270	▲ 7,999	▼ 123	▼ 43.6	▼ 85.3													Angola
Mozambique	▲ 6,418,526	▼ 2,467	▲ 249	▼ 55	▲ 94													Mozambique
Zambia	▼ 5,192,162	▼ 2,397	▼ 335.5	▼ 58	▼ 94													Zambia
Zimbabwe	▼ 391,326	▼ 462	▼ 29	28	▲ 98													Zimbabwe
SUB-NATIONAL ELIMINATION TARGETS																		
*Mozambique	185,654	29	108.63	34	98			100	98		NO	NO	NO	NO	YES			*Mozambique
**Zimbabwe	2,145		1.20		98	1,139		100	95		YES	YES	NO	YES	YES	YES / YES	NO	**Zimbabwe

Data Source: E8 Ministries of Health

KEY: ON TRACK | SOME PROGRESS | NOT ON TRACK | NO DATA / NOT APPLICABLE

▲ INCREASE SINCE LAST YEAR ▼ DECREASE SINCE LAST YEAR

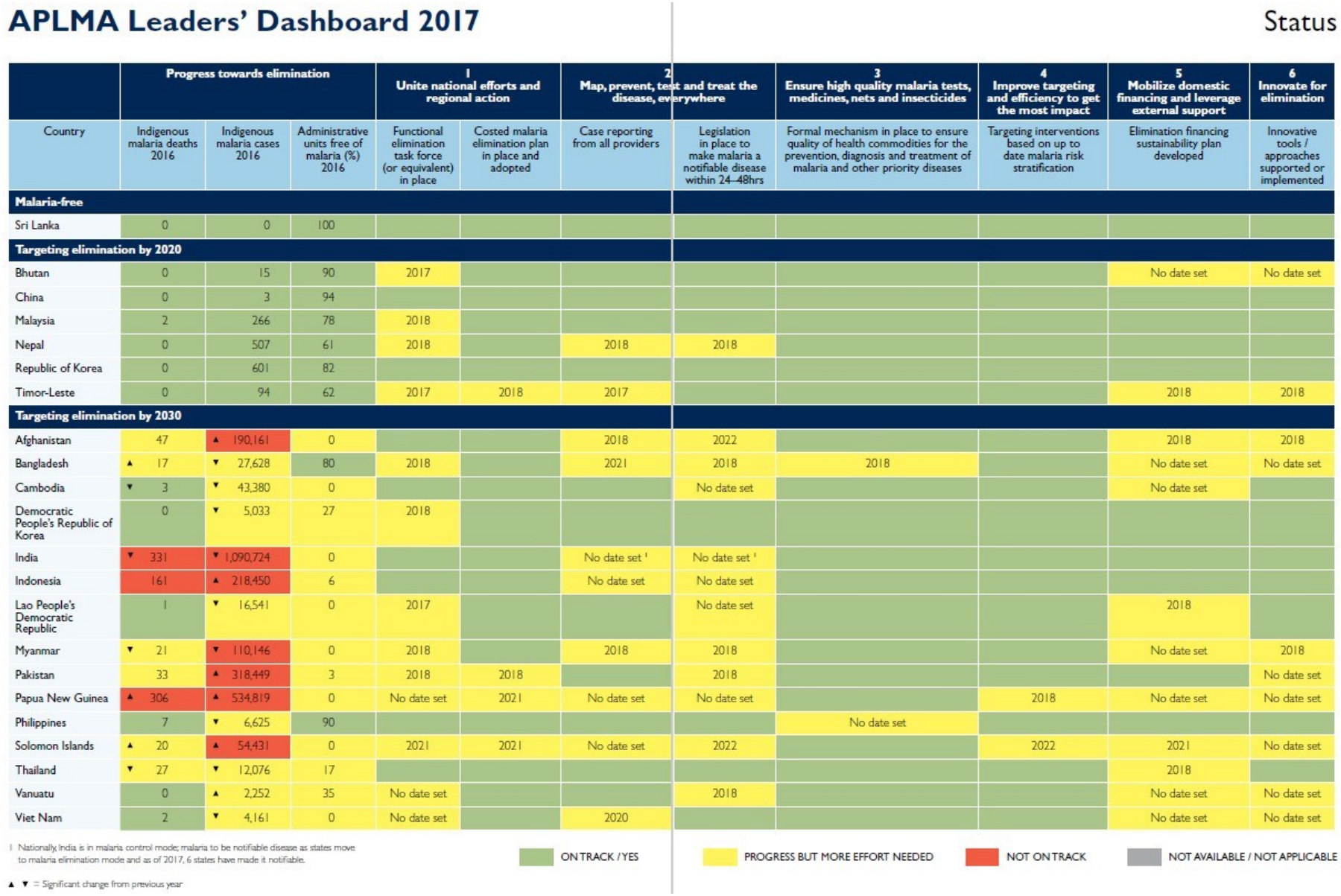
PRODUCED BY THE E8 SECRETARIAT WITH THE SUPPORT OF RBM PARTNERS

ROLL BACK MALARIA PARTNERSHIP

* Mozambique: Of the 161 districts in Mozambique, 8 districts (in the Maputo Province) have adopted a sub-national elimination strategy

** Zimbabwe: Of the 62 districts in Zimbabwe, 20 districts have adopted a sub-national elimination strategy

Figure 7. APLMA Leaders' Dashboard



APLMA Leaders' Dashboard 2017

Status

	Progress towards elimination			1 Unite national efforts and regional action		2 Map, prevent, test and treat the disease, everywhere	3 Ensure high quality malaria tests, medicines, nets and insecticides		4 Improve targeting and efficiency to get the most impact	5 Mobilize domestic financing and leverage external support	6 Innovate for elimination
Country	Indigenous malaria deaths 2016	Indigenous malaria cases 2016	Administrative units free of malaria (%) 2016	Functional elimination task force (or equivalent) in place	Costed malaria elimination plan in place and adopted	Case reporting from all providers	Legislation in place to make malaria a notifiable disease within 24–48hrs	Formal mechanism in place to ensure quality of health commodities for the prevention, diagnosis and treatment of malaria and other priority diseases	Targeting interventions based on up to date malaria risk stratification	Elimination financing sustainability plan developed	Innovative tools / approaches supported or implemented
Malaria-free											
Sri Lanka	0	0	100								
Targeting elimination by 2020											
Bhutan	0	15	90	2017						No date set	No date set
China	0	3	94								
Malaysia	2	266	78	2018							
Nepal	0	507	61	2018		2018	2018				
Republic of Korea	0	601	82								
Timor-Leste	0	94	62	2017	2018	2017				2018	2018
Targeting elimination by 2030											
Afghanistan	47	▲ 190,161	0			2018	2022			2018	2018
Bangladesh	▲ 17	▼ 27,628	80	2018		2021	2018	2018		No date set	No date set
Cambodia	▼ 3	▼ 43,380	0				No date set			No date set	
Democratic People's Republic of Korea	0	▼ 5,033	27	2018							
India	▼ 331	▼ 1,090,724	0			No date set [1]	No date set [1]				
Indonesia	161	▲ 218,450	6			No date set	No date set				
Lao People's Democratic Republic	1	▼ 16,541	0	2017			No date set			2018	
Myanmar	▼ 21	▼ 110,146	0	2018		2018	2018			No date set	2018
Pakistan	33	▲ 318,449	3	2018	2018		2018				No date set
Papua New Guinea	▲ 306	▲ 534,819	0	No date set	2021	No date set	No date set		2018	No date set	No date set
Philippines	7	▼ 6,625	90					No date set			
Solomon Islands	▲ 20	▲ 54,431	0	2021	2021	No date set	2022		2022	2021	No date set
Thailand	▼ 27	▼ 12,076	17							2018	
Vanuatu	0	▲ 2,252	35	No date set			2018			No date set	No date set
Viet Nam	2	▼ 4,161	0	No date set		2020				No date set	No date set

1 Nationally, India is in malaria control mode; malaria to be notifiable disease as states move to malaria elimination mode and as of 2017, 6 states have made it notifiable.

▲ ▼ = Significant change from previous year

ON TRACK / YES

PROGRESS BUT MORE EFFORT NEEDED

NOT ON TRACK

NOT AVAILABLE / NOT APPLICABLE

6. IMPLEMENTING THE EVALUATION FRAMEWORK

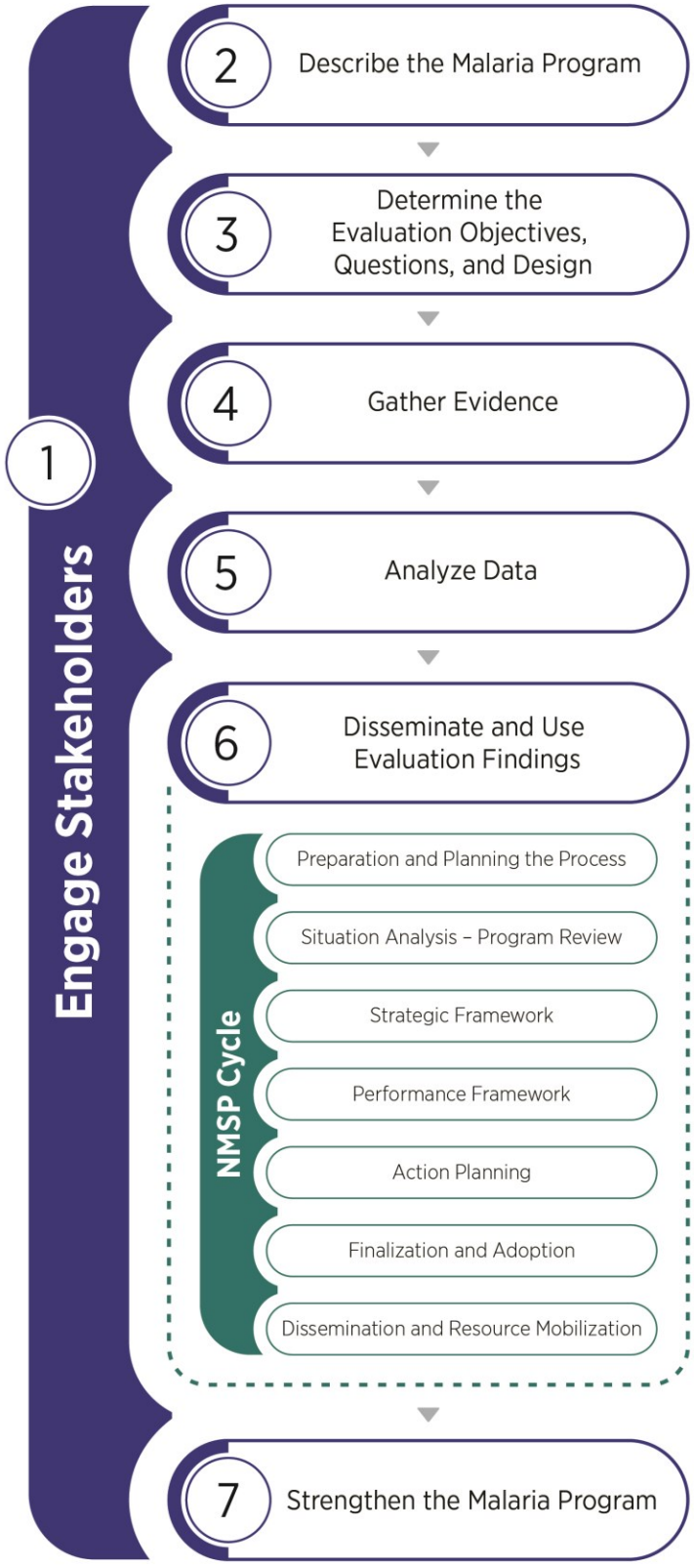
6.1. Steps for implementing the evaluation framework

The implementation of the evaluation framework involves seven interrelated steps, which are as follows: (1) engaging stakeholders; (2) describing the malaria program; (3) determining the evaluation objectives, questions, and design; (4) gathering evidence; (5) analyzing the data; (6) disseminating and using the evaluation findings; and (7) improving and strengthening the national malaria program (Figure 8).[11] These steps promote country ownership, partner coordination, and the dissemination of evidence for use in a timely manner [21]. As shown, the framework also encourages the prospective planning of evaluations at the start of the NMSP lifecycle in tandem with the development of the national malaria M&E plan. This timing allows for greater stakeholder involvement and buy-in for evaluation from the beginning of the program, promotes timely implementation of evaluations, and helps ensure that data needed for evaluation will be collected.

This framework is applicable for all organizations that may lead a process or impact evaluation, whether it be the NMP, the MOH, donors, implementing partners, or academic institutions. The NMP and the MOH should lead the coordination of the evaluation process in country to ensure country ownership and use of results, even if an external organization or independent evaluator is brought in to conduct the evaluation. Engaging an independent evaluator can be valuable to help maintain impartiality and maximize accountability.

[11] Framework adapted from: **Mortality Task Force of Roll Back Malaria's Monitoring and Evaluation Reference Group. 2014. Guidance for Evaluating the Impact of National Malaria Control Programs in Highly Endemic Countries**. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina.

Figure 8. Implementation framework for evaluating national malaria control programs



6.1.1. Engage stakeholders

Stakeholder engagement is a crucial first step in the implementation of an evaluation and should be as inclusive as possible. Key stakeholders in the NMP likely include national and subnational program staff from government institutions, policymakers, implementing partners, advocacy groups, program beneficiaries, academic and research institutions, and funding partners. Engaging stakeholders is critical for promoting local ownership of the evaluation process and results, and thus ensuring that evaluation results will be used. Engaging stakeholders throughout the evaluation process helps ensure that the right questions are asked, needed data can be accessed, and findings are validated and consequently considered credible. It is important that through this engagement process, stakeholders come away with a common understanding of the evaluation purpose, objectives, design and methods (and their limitations), roles and responsibilities of the various stakeholders, and how the findings will be disseminated and used for strengthening the NMP.

A stakeholder analysis should be conducted to inform which partners to engage in the evaluation. A good starting point for this analysis is to review the partners that are involved in the national malaria strategic planning process. Then consider whether any key partners were missing from that process, and whether their role in malaria prevention and control efforts in country could help inform the evaluation process or could be benefitted by the evaluation findings. If that is the case, those partners should be included among the evaluation stakeholders. If the NMP and the MOH are not leading the implementation of the evaluation, it is recommended that a specific point of contact be appointed from each institution for the evaluation.

Different stakeholders will have varying levels of expertise to offer, availability, and interest in an evaluation of the NMP. It may be beneficial to designate different groups or stakeholders for an evaluation. These groups could include, for example, the core evaluation team tasked with implementing the evaluation, a steering committee that oversees and provides guidance for the evaluation in country, and a broader stakeholder group that provides input into the evaluation design, interpretation of the findings, and development of recommendations based on the findings [21].

Existing mechanisms, such as M&E steering committee meetings and NMP and MOH technical working groups, are good mechanisms to use for stakeholder engagement and avoid parallel or duplicative mechanisms.

6.1.2. Describe the malaria program

The second step in the implementation of the evaluation is to describe the NMP. A recent NMSP may have much of the information needed for this description. This description includes the following: the overall goal of the NMP; target populations or stratification areas; strategies and their rationale; implementation plans, including resources and inputs; and activities and outputs, their expected outcomes, and their interrelationships. This description can be mapped out visually through a logical model, theory of change, or an impact model (see example in Annex 4).

The description should also include a historical timeline of key programmatic milestones. For example, these may include large-scale ITN distributions or changes in diagnostic or treatment policies, such as the introduction of RDTs or pre-referral rectal artesunate. The NMP description provides a clear picture of how program activities were expected to lead to outcomes and impact and what happened during the period of evaluation. This description highlights linkages between the program design and the evaluation design. See the case studies in Annex 3 for examples.

6.1.3. Design the evaluation

The design of the evaluation specifically involves the development of the evaluation objectives and questions, the methods, and the timeframe for the evaluation. As such, the first step in designing the evaluation is to determine the key objectives of the evaluation, the specific evaluation questions, and what type of evaluation will be conducted (process, outcome, or impact). It is critical to engage stakeholders to get their input and buy-in to the evaluation objectives and questions, and to clearly state how the information generated by the evaluation will be used by the NMP and stakeholders. Organizing a stakeholder meeting at the start of an evaluation can be a great opportunity to get input on the evaluation objectives and questions, identify potential data sources, and review stakeholder roles and responsibilities in the implementation of the evaluation (e.g., accessing data, analyzing data, and writing or reviewing the evaluation report).

Considerations for the data needed to answer the evaluation questions, the resource requirements for implementing the evaluation, and other feasibility issues (e.g., the timeframe needed or analytical requirements) all need to feed into the development and refinement of the evaluation questions and methods. The evaluation team must clearly describe the methods selected to address the evaluation questions based on the available data sources and must document the ethical considerations for using each of these data sources. The timeframe for the evaluation should have a defined baseline and end line, which should take into consideration the NMSP lifecycle, the donor funding cycles, when malaria interventions were implemented, the temporal dynamics of malaria transmission, and what data sources are available and what periods of time those data sources represent (e.g., when surveys were conducted).

After the evaluation objectives, questions, and overall design have been agreed upon, a detailed evaluation protocol outlining the evaluation objectives and questions, methods, data sources, and analysis plan should be developed to guide the evaluation team and ensure transparency to stakeholders.

6.1.4. Gather evidence and conduct the analysis

This step consists of defining the evaluation indicators, compiling secondary data and collecting primary data (as necessary), assessing the quality of the data, conducting data analysis, and writing the evaluation report. The core evaluation team should identify the key indicators that answer each of the evaluation questions (recommended indicators for process and impact evaluation are listed in Section 4, Table 2, with a detailed indicator reference guide in Annex 2). Stakeholders should be engaged to identify appropriate and available data sources to measure the indicators and to assess and verify the quality of those data sources. Creating stakeholder trust in the data is critical for evaluation credibility.

To the extent possible, existing data should be used for process and impact evaluations. The evaluators should make sure that the overall evaluation timeline includes time for gaining access to datasets. Obtaining existing datasets for secondary analysis may require a data use agreement or memorandum of understanding. If the evaluation team is unable to access certain datasets, values for indicators of interest may be extracted from published or unpublished survey reports or other studies. The process of identifying and obtaining existing data should include identification of key data gaps and quality issues. For process evaluations, some primary data collection may be necessary, such as key informant interviews, to more fully understand what happened during program implementation. If primary data collection—qualitative or quantitative—is done for the evaluation, all data collection tools should be field tested before administration.

Primary data collection may require review and approval by an ethical review committee or institutional review board (IRB). In many cases, evaluation data may be determined exempt by the ethical review board. It is critical that a review and approval or determination of exemption is obtained before proceeding with any

primary data collection. In some cases, an exemption for secondary data analysis may also be required. Adequate time for any ethical review must be built into the evaluation timeline.

The analysis plan developed during the evaluation design phase should specify the data sources and analyses, and it should identify who will conduct the analyses. As new evidence emerges, the analysis plan may require adjustment. An analysis workshop with a small group of key stakeholders will add to the credibility of the analysis. A dissemination workshop with stakeholders to discuss the findings and interpretation of results may further bolster credibility of the results and consequent action planning.

In initial stakeholder discussions and engagement meetings, a process for reviewing and finalizing the evaluation report should be discussed. This includes the different stages of review and a proposed timeline for providing input. All stakeholders should have an opportunity to review the evaluation report. It is important to factor in the time for the different stages of review and consequent revisions.

6.1.5. Use and disseminate the evaluation findings

Engaging stakeholders from the start of the evaluation in the design and gathering of evidence will help set the stage for dissemination of the evaluation and promote the use of the findings. Early in the evaluation process, the evaluation team should establish with stakeholders that the evaluation findings will be released, regardless of whether the findings are positive (e.g., whether findings demonstrate that the NMSP was implemented as intended or that there was demonstrated impact by the program). The evaluators should provide ample opportunity for review and discussion of findings among stakeholders, as well as for development of stakeholder-generated recommendations. This should be part of the process of writing and finalizing the report and part of the dissemination of the results with all stakeholders. In addition to the full evaluation report, a summary of key findings (e.g., an evaluation brief) should be prepared because this will be more easily digestible to a broader audience.

Results can be disseminated through different avenues to different audiences. Dissemination may include targeted dissemination events such as meeting presentations and action-planning workshops, development of policy briefs and factsheets, and publication in peer-reviewed journals. For process evaluations in particular, an action plan to address identified gaps should be developed as part of dissemination activities.

Evaluation results are intended for use by multiple stakeholders. The findings should inform the NMP's development of the next national NMSP. Specifically, the results should be used to update, adapt, and strengthen the strategies; increase the effectiveness of activities and interventions; and either maintain or accelerate progress toward goals. In addition, gaps in data identified during the evaluation can be used to inform data collection efforts and updates to the NMP M&E strategy. Donors and other financial partners can also use the findings to demonstrate their contributions and guide funding decisions.

6.2. Evaluation timeline

A realistic timeline for carrying out the evaluation should be developed at the onset of the evaluation, to set expectations (14 months, with range of 12 to 18 months) for when the different stages of the evaluation will be completed and when the results will be available. Several considerations should be taken into account when developing the timeline. It is critical that the evaluation is carried out in a timely manner to ensure that the results are relevant and useful for informing adjustments to the NMP. For example, process evaluations should be performed to feed into a mid-term review and allow adequate time for course corrections during the last years of the NMSP. A process or impact evaluation will also ideally be conducted toward the end of the NMSP cycle, to provide findings that can feed into the development of the next NMSP. The timing of

these evaluations, as feasible, should take into consideration important funding cycles to allow the results to inform the funding application development process.

It is important to plan ample time for stakeholder engagement in the evaluation design, discussion of preliminary findings, and reviews of the evaluation report. The timeline will also need to plan for time needed to access datasets and to get the necessary approvals from the NMP, the MOH, and the IRB.

Timelines should be developed for each phase of the evaluation, including the design, analysis, report writing, and dissemination phases. Deadlines should be set and adhered to for writing and reviewing each draft of the evaluation report. Table 5 provides a sample timeline for conducting an evaluation.

Table 5. Illustrative timeline for conducting a process or impact evaluation

Activity	Estimated timeframe	Months 1–4	Months 5-8	Months 9–12	Months 13–16	Months 17–18	Stakeholders involved
Conduct initial stakeholder engagement	2–3 weeks	Phase 1: Plan and design the evaluation					NMP, MOH, funding partners
Describe malaria program	2 weeks						Evaluation team
Identify and contract evaluation team	6 weeks						Evaluation team
Design evaluation: develop evaluation protocol, inclusive of the evaluation methodology, data collection tools (if needed), analysis plan, and work plan and task matrix	2–3 weeks						Evaluation team
Ethical review of evaluation protocol	4–12 weeks						Evaluation team
Kick off the evaluation with a stakeholder meeting	2 weeks						Evaluation team, funding partners, government and key stakeholders
Identify existing data and solicit access to datasets	2–4 weeks						Evaluation team
Data collection	4-8 weeks		Phase 2: Gather evidence and conduct analysis				Evaluation team
Conduct preliminary analysis	3–4 weeks						Evaluation team
Develop report outline	1 week						Evaluation team
Complete analysis	4 weeks						Evaluation team

Activity	Estimated timeframe	Months 1–4	Months 5-8	Months 9–12	Months 13–16	Months 17–18	Stakeholders involved
Develop draft report and share with core stakeholders	4 weeks				Phase 3: Use and disseminate findings		Evaluation team
Review draft report	4 weeks						Key stakeholders
Convene consultative meeting to present preliminary results	1 week						Key stakeholders, evaluation team
Revise evaluation report and incorporate feedback	4 weeks						Evaluation team
Allow external reviewers to comment on report	4 weeks						Selected stakeholders
Finalize evaluation report, including editing and formatting	4 weeks						Evaluation team, editor, proofreader, graphic designer
Print report	1 week						Printer
Hold dissemination event to share findings and perform preliminary action planning	2 weeks						Evaluation team, key stakeholders

Adapted from MEASURE Evaluation, United States Agency for International Development, and Roll Back Malaria. 2014. **Guidance for Evaluating the Impact of National Malaria Control Programs in Highly Endemic Countries**.

6.3. Resource requirements for evaluation

An important part in planning for an evaluation is determining the human resources and skills and financial costs required for conducting the evaluation. The evaluation team should comprise individuals with a solid understanding of malaria epidemiology and of the malaria program in the country, strong quantitative and qualitative research and analytic skills, knowledge of data quality dimensions and how to assess data quality, and skills in data visualization and writing evaluation findings. Depending on the evaluation questions and methods used, it may be necessary to bring in individuals with expertise in geographic information systems, modeling or advanced analytic methods, and analysis of specific contextual factors (e.g., climate and environmental factors).

A detailed evaluation budget should be prepared before the evaluation is implemented. Ideally, if evaluations are prospectively planned at the start of the NMSP cycle, they can be incorporated into the NMP's costed work plan or costed M&E plan to ensure that funds are set aside for evaluation. The evaluation budget should consider the costs incurred during each phase of the evaluation, including the cost of staff time to

coordinate the evaluation, analyze the data, write the report, and liaise with stakeholders. If any primary data collection will be done, that will require a fieldwork budget inclusive of IRB review fees, transport, and possibly small stipends for key informant interview respondents. If datasets are not publicly available, it may be necessary to pay for access. Seemingly minor budget additions such as licenses for updated data analysis software can add up, so this should be considered carefully when developing the budget at the outset of the evaluation. The budget should include costs for human resources (evaluation staff and subcontracts), stakeholder meetings, payment for datasets if needed, translation, and dissemination needs, including printing and workshop or meeting costs. A sample budget template with key items is provided in Table 6.

Table 6. Sample malaria evaluation budget template

Activity	Cost	Details/notes
Technical partners (international or local)		
Management of the evaluation process		
Compilation of data in country		
Data analysis		
Report writing		
Stakeholder meetings		
Access and analysis of meteorological data		
Translation services		
Printing (full report and key findings report)		
Dissemination meeting		
TOTAL		

Adapted from: [21]

7. CONCLUSIONS

The complex and ongoing changes in malaria epidemiology require different and more detailed data on transmission and risk so countries can effectively strengthen their NMPs and track and report on progress. Evaluations of NMPs provide critical information for programmatic and policy decision-making. Undertaking process, outcome, and impact evaluations in a step-wise and cyclical manner is important for generating useful information for decision-making. Results of process evaluations provide essential evidence for impact evaluations.

Quasi-experimental study designs are well-suited to evaluate malaria programs and interventions and can provide adequate evidence. Confirmed malaria incidence is considered the most appropriate impact indicator for evaluations in moderate- and low-transmission settings, and this framework document presents several evaluation designs that use routine surveillance data and include approaches to reduce bias in these data. Prospective planning of evaluations is suggested to ensure that all relevant data are likely to available for the evaluation.

Inclusion of contextual factors, either explicitly in impact analysis, or qualitatively in interpretation of impact estimates, is crucial to understanding whether contextual factors may have confounded the association between the program and impact indicator. Contextual factors may include health system factors, sociocultural and socioeconomic factors, climate factors, environmental factors, and epidemiological factors.

Data sources for impact evaluations include household surveys, facility surveys, routine health information including from community health information systems, and special studies targeting particular risk groups or geographical areas.

Interrupted time series, dose-response, use of constructed controls, and stepped-wedge designs are proposed as the major evaluation designs and analytic techniques relevant for low- and moderate-transmission settings. This framework presents descriptions and examples of these methods, along with information about supplementary analytic approaches to address issues such as endogeneity and methods for secondary analysis of cross-sectional data (e.g., DiD) in impact evaluation.

Implementing the evaluation framework requires stakeholder engagement and well-planned timelines for each activity in each phase, beginning with initial stakeholder engagement prior to evaluation design through dissemination of results and action planning to address the findings.

8. GLOSSARY

Term	Definition
All-cause child mortality rate	Probability of dying from any cause between the first and fifth birthday per 1,000 children who survived to age 12 months [5]
Autocorrelation	Correlation of consecutive observations over time
Bias	Systematic error in an estimate or inference
Civil registration and vital statistics system	Continuous, permanent, compulsory, and universal recording of the occurrence and characteristics of vital events (live births, deaths, fetal deaths, marriages, and divorces) and other civil status events pertaining to the population as provided by decree, law or regulation, in accordance with the legal requirements in each country [105]
Contextual factors	Non-malaria programs and other factors, such as rainfall, socioeconomic status, urbanization, and policy changes, that could confound the association between an intervention and its potential health impact or modify the effect of the intervention and affect the conclusion [5]
Counterfactual	The state of affairs that would have happened in the absence of the cause
Cure, radical	Elimination of both blood-stage and latent liver infection in cases of *P. vivax* and *P. ovale* infection, thereby preventing relapses [9]
Dose-response	A study that investigates the relationship between the observed outcomes and different levels of the presumed cause
Drug efficacy	Capacity of an antimalarial medicine to achieve the therapeutic objective when administered at a recommended dose, which is well tolerated and has minimal toxicity [106]
Endogenous variable	A variable that is caused by other variables in the model
Epidemic	Occurrence of a number of malaria cases highly in excess of that expected in a given place and time [106]
Evaluation	A comprehensive assessment of a program, normally undertaken at discrete points in time and focused on the longer-term outcomes and impacts of programs
Evaluation, impact	Method of assessing the changes in an outcome that can be attributed to a particular intervention or package of interventions, such as a project, program, or policy; seeks to answer cause-and-effect questions [107]
Evaluation, process	Method of assessing how a program is being implemented; focuses on the program's operations, implementation, and service delivery [107]
Exogeneous variable	A variable that is not caused by other variables in the model
Experimental methods	Methods that involve random assignment of the program or intervention under investigation, so that outcomes can be

Term	Definition
	compared between individuals or groups that did and did not receive the program or intervention. By randomizing assignment, the distribution of observed and unobserved confounders is assumed to be similar across each group. The most common experimental methodology is the randomized controlled trial.
Focal screening and treatment	Screening of a population in a defined geographical area, testing individuals at risk and treating those with a positive malaria test result [106, 108] (*adapted from WHO*)
Focus (Foci)	A defined circumscribed area situated in a currently or formerly malarious area that contains the epidemiological and ecological factors necessary for malaria transmission [9]
Health and demographic surveillance site	A set of field and computing operations to handle the longitudinal follow-up of well-defined entities or primary subjects (e.g., individuals, households) and all related demographic and health outcomes in a clearly circumscribed geographic area [107]
Impact indicator	Indicator that describes health effects
Input indicator	Indicator that describes the basic needs for a program: policy, financing/money, infrastructure
Insecticide resistance	Property of mosquitoes to survive exposure to a standard dose of insecticide; may be the result of physiological or behavioral adaption [106]
Instrumental variable	A variable or set of variables that is correlated with outcome only through an effect on other variables
Integrated community case management	An equity-focused strategy that complements and extends the reach of public health services by providing timely and effective treatment of malaria, pneumonia, and diarrhea to populations (especially children under five) with limited access to facility-based health care providers
Intermittent preventive treatment in infants	A full therapeutic course of sulfadoxine-pyrimethamine delivered to infants in co-administration with DTP2/Penta2, DTP3/Penta3 and measles immunization, regardless of whether the infant is infected with malaria [106]
Intermittent preventive treatment in pregnancy	A full therapeutic course of antimalarial medicine given to pregnant women at routine prenatal visits, regardless of whether the woman is infected with malaria [106]
Interrupted time series	A design in which a string of consecutive observations is interrupted by the imposition of a treatment to see whether the slope or intercept of the series changes as a result of the intervention
Intervention maturity	Duration of intervention implementation vis-à-vis previously demonstrated ability to have the intended effect; for example, whether sufficient time has elapsed since ITN

Term	Definition
	distribution for nets to have been hung in households and a consequent reduction in number of infectious mosquito bites to have occurred
Loop-mediated isothermal amplification	A method that amplifies DNA under constant temperature
Malaria annual parasite incidence	Number of confirmed malaria cases (via rapid diagnostic test or microscopy) during one year, measured per 1,000 population
Malaria confirmed case	Occurrence of malaria infection in a person in whom the presence of malaria parasites in the blood has been confirmed by a diagnostic test [9]
Malaria mortality, direct	Deaths in which malaria was the underlying cause [5]
Malaria mortality, indirect	Deaths in which malaria was a contributing cause and the death was categorized as a non-malaria death [5]
Malaria parasite prevalence	Proportion of a specified population with malaria infection confirmed by a diagnostic test at one point in time [106] (*adapted from the World Health Organization*)
Malaria-related mortality	Deaths in which malaria was the underlying cause or a contributing cause; sum of direct and indirect malaria mortality [8]
Malaria suspected (or presumed) case	Case suspected of being malaria that is not confirmed by a diagnostic test [106]
Mass drug administration	Administration of antimalarial treatment to all age groups of a defined population or every person living in a defined geographical area (except those for whom the medicine is contraindicated) at approximately the same time and often at repeated intervals [106]
Mass screening and treatment	Screening of an entire population for risk factors, testing individuals at risk and treating those with a positive malaria test result [106]
Mediator	A third variable that comes between a cause and effect and that transmits the causal influence from the cause to the effect
Monitoring	A continuous process of gathering and using data on program implementation, with the aim of ensuring that programs are proceeding satisfactorily, and making adjustments if necessary. The monitoring process often uses administrative data to track inputs, processes, and outputs, although it can also consider program outcomes and impacts [9].
Non-experimental methods	Observational studies (individual-level and group-level) where the presumed cause and effect are measured, but the investigator has not assigned the intervention. Non-experimental methods include multivariable regression, decomposition analysis, and mediation analysis.

Term	Definition
Null hypothesis	The hypothesis being tested; that there is no relationship between the variables
Outcome indicator	Indicator that describes coverage, or exposure to activities or interventions
Output indicator	Indicator that describes what the activity did, or the result of the activity
Population at risk	Population living in a geographical area where locally acquired malaria cases have occurred in the past three years [9]
Power	The probability of correctly rejecting a false null hypothesis
Process indicator	Indicator that describes an activity or program action
Quasi-experimental methods	Studies whereby interventions are assigned, but not in a randomized way. Assignment may be made by administrator selection (e.g., the national malaria program selects certain districts to receive indoor residual spraying) or as a national policy (e.g., national insecticide-treated net distribution). Under a various set of assumptions, quasi-experimental methods can provide valid estimates of causal effects. Quasi-experimental methods include interrupted time series, regression discontinuity, dose-response, stepped-wedge, propensity score matching, difference-in-differences, and instrumental variables.
Regression discontinuity design	An experiment in which units are assigned to conditions based on exceeding a cut-off on an assignment variable
Seasonal malaria chemoprevention	Intermittent administration of full treatment courses of an antimalarial medicine during the malaria season to prevent malarial illness. The objective is to maintain therapeutic concentrations of an antimalarial drug in the blood throughout the period of greatest risk for malaria. Note: This intervention is recommended only for areas with highly seasonal malaria, where transmission occurs during a few months of the year [106].
Selection bias	When selection results in differences in unit characteristics between conditions that may be related to outcome differences
Sentinel surveillance	The ongoing, systematic collection, analysis, and interpretation of health data undertaken in a limited number of health facilities [107]
Serological assay	Procedure used to measure antimalarial antibodies in serum [106]
Slide positivity rate	Proportion of blood smears found to be positive for *Plasmodium* among all blood smears examined [106]

Term	Definition
Sporozoite	Motile stage of the malaria parasite that is inoculated by a feeding female anopheline mosquito and may cause infection
Stepped-wedge design	An experiment in which some or all clusters begin by experiencing the control condition, then cross over to receive the intervention condition. The schedule for the intervention condition may be randomized or decided according to logistical factors.
Surveillance	Continuous, systematic collection, analysis, and interpretation of disease-specific data and use in planning, implementing, and evaluating public health practices [9]
Transmission intensity	Frequency with which people living in an area are bitten by anopheline mosquitos carrying human malaria sporozoites Classifications of high-, moderate-, low-, and very low-transmission estimated through two proxy measures, including annual parasite incidence and *P. falciparum* parasite rate [106]
Treatment, first-line	Treatment recommended in national treatment guidelines as the medicine of choice for treating malaria [106]
Treatment, radical	Treatment to achieve complete cure; applies only to *vivax* and *ovale* infections and consists of the use of medicines that destroy both blood and liver stages of the parasite [106]
Vectorial capacity	Number of new infections that the population of a given vector would induce per case per day at a given place and time, assuming that the human population is and remains fully susceptible to malaria

9. REFERENCES

1. World Health Organization (WHO): **Global technical strategy for malaria 2016-2030.** Geneva, Switzerland: WHO; 2016.

2. World Health Organization: **World Malaria Report 2016.** Geneva, Switzerland: World Health Organization; 2016.

3. **United Nations Sustainable Development Goals** [https://sustainabledevelopment.un.org/sdgs]

4. World Health Organization: **World Malaria Report 2017.** Geneva, Switerzland: World Health Organization (WHO); 2017.

5. Mortality Task Force of the Roll Back Malaria's Monitoring and Evaluation Reference Group: **Guidance for evaluating the impact of national malaria control programs in highly endemic countries.** Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina; 2014.

6. World Health Organization (WHO): **Malaria programme reviews: A manual for reviewing the performance of malaria control and elimination programs.** Geneva, Switzerland: WHO; 2010.

7. Yé Y, Eisele TP, Eckert E, Korenromp E, Shah JA, Hershey CL, Ivanovich E, Newby H, Carvajal-Velez L, Lynch M, et al: **Framework for evaluating the health impact of the scale-up of malaria control interventions on all-cause child mortality in sub-Saharan Africa.** *Am J Trop Med Hyg* 2017, **97:**9-19.

8. Rowe AK, Steketee RW, Arnold F, Wardlaw T, Basu S, Bakyaita N, Lama M, Winston CA, Lynch M, Cibulskis RE, et al: **Viewpoint: Evaluating the impact of malaria control efforts on mortality in sub-Saharan Africa.** *Tropical Medicine & International Health* 2007, **12:**1524-1539.

9. World Health Organization (WHO): **Malaria surveillance, monitoring & evaluation: A reference manual.** pp. 105. Geneva: WHO; 2018:105.

10. World Health Organization (WHO): **From malaria control to malaria elimination: A manual for elimination scenario planning.** Geneva, Switzerland: WHO; 2014.

11. World Health Organization (WHO): **A framework for malaria elimination.** Geneva, Switzerland: WHO; 2017.

12. World Health Organization (WHO), UNICEF: **Integrated Community Case Management (iCCM): An equity-focused strategy to improve access to essential treatment services for children.** WHO/UNICEF Joint Statement. 2012.

13. **Roll Back Malaria Partnership (RBM) Malaria Indicator Surveys** [http://www.malariasurveys.org/]

14. **The DHS Program: The Demographic and Health Surveys Program** [https://dhsprogram.com/]

15. **UNICEF Multiple indicator cluster surveys** [http://mics.unicef.org/]

16. **The DHS Program Service Provision Assessment (SPA) Overview** [https://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm]

17. World Health Organization (WHO): **Service availability and readiness assessment (SARA): An annual monitoring system for service delivery: Reference manual.** Geneva, Switzerland: WHO; 2013.

18. MEASURE Evaluation, MEASURE DHS, President's Malaria Initiative, Roll Back Malaria Partnership, UNICEF, World Health Organization: **Household survey indicators for malaria control.** Chapel Hill, NC, USA: MEASURE Evaluation, ICF; 2018.

19. Maternal and Child Health Integrated Program (MCHIP): **Indicator guide: Monitoring and evaluating integrated community case management.** Washington, DC, USA: MCHIP; 2013.

20. Roll Back Malaria Partnership (RBM): **Malaria behavior change communication (BCC) indicator reference guide.** Geneva, Switzerland: RBM; 2014.

21. Hershey CL, Bhattarai A, Florey LS, McElroy PD, Nielsen CF, Yé Y, Eckert E, Franca-Koh AC, Shargie E, Komatsu R, et al: **Implementing impact evaluations of malaria control interventions: Process, lessons learned, and recommendations.** *The American Journal of Tropical Medicine and Hygiene* 2017, **97:**20-31.

22. World Health Organization (WHO): **Monitoring the building blocks of health systems: A handbook of indicators and their measurement strategies.** Geneva, Switzerland: WHO; 2010.

23. DaSilva J, Garanganga B, Teveredzi V, Marx SM, Mason SJ, Connor SJ: **Improving epidemic malaria planning, preparedness and response in Southern Africa.** *Malaria Journal* 2004, **3:**37.

24. World Health Organization (WHO): **Malaria epidemics: Forecasting, prevention, early detection and control. From policy to practice.** Leysin, Switzerland: WHO; 2003.

25. Barnighausen T, Rottingen JA, Rockers P, Shemilt I, Tugwell P: **Quasi-experimental study designs series-paper 1: Introduction: Two historical lineages.** *J Clin Epidemiol* 2017, **89:**4-11.

26. Barnighausen T, Tugwell P, Rottingen JA, Shemilt I, Rockers P, Geldsetzer P, Lavis J, Grimshaw J, Daniels K, Brown A, et al: **Quasi-experimental study designs series-paper 4: Uses and value.** *J Clin Epidemiol* 2017, **89:**21-29.

27. Cochrane Effective Practice and Organisation of Care (EPOC): **What study designs can be considered for inclusion in an EPOC review and what should they be called? EPOC resources for review authors.** Oslo: Norwegian Knowledge Centre for the Health Services; 2017.

28. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, Goodacre S, Herrin J, Kerse N, McKinlay RJ, et al: **A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation.** *J Clin Epidemiol* 2015, **68:**324-333.

29. Cohen JM, Le Menach A, Pothin E, Eisele TP, Gething PW, Eckhoff PA, Moonen B, Schapira A, Smith DL: **Mapping multiple components of malaria risk for improved targeting of elimination interventions.** *Malar J* 2017, **16:**459.

30. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, Zimmerman PA, Weiss DJ, Cameron E, Gething PW, Howes RE: **Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess *Plasmodium falciparum* endemicity trends between 2011 and 2016.** *BMC Med* 2018, **16:**71.

31. **Malaria Atlas Project** [www.map.ox.ac.uk]

32. Pfeffer DA, Lucas TCD, May D, Harris J, Rozier J, Twohig KA, Dalrymple U, Guerra CA, Moyes CL, Thorn M, et al: **malariaAtlas: An R interface to global malariometric data hosted by the Malaria Atlas Project.** *Malar J* 2018, **17:**352.

33. Giardina F, Gosoniu L, Konate L, Diouf MB, Perry R, Gaye O, Faye O, Vounatsou P: **Estimating the burden of malaria in Senegal: Bayesian zero-inflated binomial geostatistical modeling of the MIS 2008 data.** *PLoS One* 2012, **7:**e32625.

34. **MEASURE Evaluation Data Quality Review Toolkit** [https://www.measureevaluation.org/our-work/data-quality/data-quality-review]

35. World Health Organization (WHO): **Data Quality Review (DQR) Toolkit Module 1: Framework and metrics.** Geneva, Switzerland: WHO; 2017.

36. World Health Organization (WHO): **Data Quality Review (DQR) Toolkit Module 2: Desk review of data quality.** Geneva, Switzerland: WHO; 2017.

37. World Health Organization (WHO): **Data Quality Review (DQR) Toolkit Module 3: Data verification and system assessment.** Geneva, Switzerland: WHO; 2017.

38. Ashton RA, Bennett A, Yukich J, Bhattarai A, Keating J, Eisele TP: **Methodological considerations for use of routine health information system data to evaluate malaria program impact in an era of declining malaria transmission.** *Am J Trop Med Hyg* 2017, **97:**46-57.

39. Victora CG, Black RE, Boerma JT, Bryce J: **Measuring impact in the Millennium Development Goal era and beyond: A new approach to large-scale effectiveness evaluations.** *The Lancet* 2011, **377:**85-95.

40. Spiegelman D: **Evaluating public health interventions: 2. Stepping up to routine public health evaluation with the stepped wedge design.** *Am J Public Health* 2016, **106:**453-457.

41. Bernal JL, Cummins S, Gasparrini A: **Interrupted time series regression for the evaluation of public health interventions: A tutorial.** *Int J Epidemiol* 2017, **46:**348-355.

42. Cousens S, Hargreaves J, Bonell C, Armstrong B, Thomas J, Kirkwood BR, Hayes R: **Alternatives to randomisation in the evaluation of public-health interventions: Statistical analysis and causal inference.** *J Epidemiol Community Health* 2011, **65:**576-581.

43. Bennett A, Yukich J, Miller JM, Vounatsou P, Hamainza B, Ingwe MM, Moonga HB, Kamuliwo M, Keating J, Smith TA, et al: **A methodological framework for the improved use of routine health system data to evaluate national malaria control programs: Evidence from Zambia.** *Population Health Metrics* 2014, **12:**30.

44. Eisele TP, Macintyre K, Yukich J, Ghebremeskel T: **Interpreting household survey data intended to measure insecticide-treated bednet coverage: Results from two surveys in Eritrea.** *Malaria Journal* 2006, **5:**36-36.

45. Ashton RA, Doumbia B, Diallo D, Druetz T, Florey L, Taylor C, Arnold F, Mihigo J, Koné D, Fomba S: **Measuring malaria diagnosis and treatment coverage in population-based surveys:**

A recall validation study in Mali among caregivers of febrile children under five years. *Malar J.* 2019 Jan 3;**18**(1):3. doi: 10.1186/s12936-018-2636-3.

46. Allcock SH, Young EH, Sandhu MS: **A cross-sectional analysis of ITN and IRS coverage in Namibia in 2013.** *Malaria Journal* 2018, **17:**264.

47. Murray CK, Gasser RA, Magill AJ, Miller RS: **Update on rapid diagnostic testing for malaria.** *Clinical Microbiology Reviews* 2008, **21:**97-110.

48. Helb DA, Tetteh KK, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H, Ssewanyana I, Kamya MR, Beeson JG, et al: **Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum* exposure for individuals and communities.** *Proc Natl Acad Sci U S A* 2015, **112:**E4438-4447.

49. Tusting LS, Bousema T, Smith DL, Drakeley C: **Measuring changes in *Plasmodium falciparum* transmission: Precision, accuracy and costs of metrics.** *Adv Parasitol* 2014, **84:**151-208.

50. Rogier E, Moss DM, Chard AN, Trinies V, Doumbia S, Freeman MC, Lammie PJ: **Evaluation of immunoglobulin G responses to *Plasmodium falciparum* and *Plasmodium vivax* in Malian school children using multiplex bead assay.** *Am J Trop Med Hyg* 2017, **96:**312-318.

51. Ondigo BN, Hodges JS, Ireland KF, Magak NG, Lanar DE, Dutta S, Narum DL, Park GS, Ofulla AV, John CC: **Estimation of recent and long-term malaria transmission in a population by antibody testing to multiple *Plasmodium falciparum* antigens.** *J Infect Dis* 2014, **210:**1123-1132.

52. Kerkhof K, Sluydts V, Willen L, Kim S, Canier L, Heng S, Tsuboi T, Sochantha T, Sovannaroth S, Menard D, et al: **Serological markers to measure recent changes in malaria at population level in Cambodia.** *Malar J* 2016, **15:**529.

53. Sesay SSS, Giorgi E, Diggle PJ, Schellenberg D, Lalloo DG, Terlouw DJ: **Surveillance in easy to access population subgroups as a tool for evaluating malaria control progress: A systematic review.** *PLoS One* 2017, **12:**e0183330.

54. Jacobson JO, Cueto C, Smith JL, Hwang J, Gosling R, Bennett A: **Surveillance and response for high-risk populations: What can malaria elimination programmes learn from the experience of HIV?** *Malaria Journal* 2017, **16:**33.

55. Ly P, Thwing J, McGinn C, Quintero CE, Top-Samphor N, Habib N, Richards JS, Canavati SE, Vinjamuri SB, Nguon C: **The use of respondent-driven sampling to assess malaria knowledge, treatment-seeking behaviours and preventive practices among mobile and migrant populations in a setting of artemisinin resistance in Western Cambodia.** *Malaria Journal* 2017, **16:**378.

56. Khamsiriwatchara A, Wangroongsarb P, Thwing J, Eliades J, Satimai W, Delacollette C, Kaewkungwal J: **Respondent-driven sampling on the Thailand-Cambodia border. I. Can malaria cases be contained in mobile migrant workers?** *Malaria Journal* 2011, **10:**120.

57. Leslie HH, Sun Z, Kruk ME: **Association between infrastructure and observed quality of care in 4 healthcare services: A cross-sectional study of 4,300 facilities in 8 countries.** *PLoS Medicine* 2017, **14:**e1002464-e1002464.

58. Rowe AK, Labadie G, Jackson D, Vivas-Torrealba C, Simon J: **Improving health worker performance: An ongoing challenge for meeting the sustainable development goals.** *BMJ (Clinical research ed)* 2018, **362:**k2813-k2813.

59. World Health Organization (WHO): **Verbal autopsy standards: The 2012 WHO verbal autopsy instrument release candidate 1.** Geneva, Switzerland: WHO, Health Metrics Network, and INDEPTH Network; 2012.

60. World Health Organization (WHO): **Manual for the training of interviewers on the use of the 2016 WHO VA instrument.** Geneva, Switzerland: WHO; 2017.

61. Herrera S, Enuameh Y, Adjei G, Ae-Ngibise KA, Asante KP, Sankoh O, Owusu-Agyei S, Ye Y: **A systematic review and synthesis of the strengths and limitations of measuring malaria mortality through verbal autopsy.** *Malar J* 2017, **16:**421.

62. Mpimbaza A, Filler S, Katureebe A, Kinara SO, Nzabandora E, Quick L, Ratcliffe A, Wabwire-Mangen F, Chandramohan D, Staedke SG: **Validity of verbal autopsy procedures for determining malaria deaths in different epidemiological settings in Uganda.** *PLoS One* 2011, **6:**e26892.

63. Baiden F, Bawah A, Biai S, Binka F, Boerma T, Byass P, Chandramohan D, Chatterji S, Engmann C, Greet D, et al: **Setting international standards for verbal autopsy.** *Bull World Health Organ* 2007, **85:**570-571.

64. Ye Y, Wamukoya M, Ezeh A, Emina JB, Sankoh O: **Health and demographic surveillance systems: A step towards full civil registration and vital statistics system in sub-Sahara Africa?** *BMC Public Health* 2012, **12:**741.

65. World Health Organization (WHO): **WHO Evidence Review Group on malaria diagnostics in low transmission settings.** Geneva, Switzerland: WHO; 2014.

66. Tambo M, Auala JR, Sturrock HJ, Kleinschmidt I, Bock R, Smith JL, Gosling R, Mumbengegwi DR: **Evaluation of loop-mediated isothermal amplification as a surveillance tool for malaria in reactive case detection moving towards elimination.** *Malaria Journal* 2018, **17:**255.

67. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WMM, et al: **Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**5108-5113.

68. Victora CG, Schellenberg JA, Huicho L, Amaral J, El Arifeen S, Pariyo G, Manzi F, Scherpbier RW, Bryce J, Habicht JP: **Context matters: Interpreting impact findings in child survival evaluations.** *Health Policy and Planning* 2005, **20 Suppl 1:**i18-i31.

69. Pindolia DK, Garcia AJ, Wesolowski A, Smith DL, Buckee CO, Noor AM, Snow RW, Tatem AJ: **Human movement data for malaria control and elimination strategic planning.** *Malaria Journal* 2012, **11:**205.

70. Ruktanonchai NW, Bhavnani D, Sorichetta A, Bengtsson L, Carter KH, Córdoba RC, Le Menach A, Lu X, Wetter E, zu Erbach-Schoenberg E, Tatem AJ: **Census-derived migration data as a tool for informing malaria elimination policy.** *Malaria Journal* 2016, **15:**273.

71. Thomson M, Ukawuba I, Hershey C, Bennett A, Ceccato P, Lyon B, Dinku T: **Using rainfall and temperature data in the evaluation of national malaria control programs in Africa.** *Am J Trop Med Hyg* 2017, **97:**32-45.

72. Penfold RB, Zhang F: **Use of interrupted time series analysis in evaluating health care quality improvements.** *Academic Pediatrics* 2013, **13:**S38-44.

73. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D: **Segmented regression analysis of interrupted time series studies in medication use research.** *J Clin Pharm Ther* 2002, **27:**299-309.

74. Shadish WR, Cook TD, Campbell DT: **Experimental and quasi-experimental designs for generalized causal inference**. Belmont, CA: Wadsworth Cengage Learning; 2002.

75. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D: **Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis.** *BMJ* 2015, **350:**h2750.

76. Lopez Bernal J, Cummins S, Gasparrini A: **The use of controls in interrupted time series studies of public health interventions.** *Int J Epidemiol* 2018, **47:**2082-2093.

77. Lopez Bernal J, Soumerai S, Gasparrini A: **A methodological framework for model selection in interrupted time series studies.** *J Clin Epidemiol* 2018, **103:**82-91.

78. Ashton R, Bennett A, Al-Mafazy A, Abass A, Msellem M, McElroy P, Kachur S, Ali A, Yukich J, Eisele T, Bhattarai A: **Use of routine health information system data to evaluate impact of malaria control interventions in Zanzibar, Tanzania from 2000 to 2015.** *EClinicalMedicine.* 2019 Jul; **12**: 11–19. Published online 2019 Jun 21. doi: 10.1016/j.eclinm.2019.05.011.

79. Lal S, Ndyomugenyi R, Alexander ND, Lagarde M, Paintain L, Magnussen P, Chandramohan D, Clarke SE: **Health facility utilisation changes during the introduction of community case management of malaria in south western Uganda: An interrupted time series approach.** *PLoS One* 2015, **10:**e0137448.

80. Bryce J, Gilroy K, Jones G, Hazel E, Black RE, Victora CG: **The Accelerated Child Survival and Development programme in West Africa: A retrospective evaluation.** *Lancet* 2010, **375:**572-582.

81. Florey LS, Bennett A, Hershey CL, Bhattarai A, Nielsen CF, Ali D, Luhanga M, Taylor C, Eisele TP, Ye Y: **Impact of insecticide-treated net ownership on all-cause child mortality in Malawi, 2006-2010.** *Am J Trop Med Hyg* 2017, **97:**65-75.

82. Rhoda DA, Murray DM, Andridge RR, Pennell ML, Hade EM: **Studies with staggered starts: Multiple baseline designs and group-randomized trials.** *Am J Public Health* 2011, **101:**2164-2169.

83. Cisse B, Ba EH, Sokhna C, JL ND, Gomis JF, Dial Y, Pitt C, M ND, Cairns M, Faye E, et al: **Effectiveness of seasonal malaria chemoprevention in children under ten years of age in Senegal: A stepped-wedge cluster-randomised trial.** *PLoS Med* 2016, **13:**e1002175.

84. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ: **Impact evaluation in practice**. The World Bank; 2011.

85. Wagman J, Gogue C, Tynuv K, Mihigo J, Bankineza E, Bah M, Diallo D, Saibu A, Richardson JH, Kone D, et al: **An observational analysis of the impact of indoor residual spraying with non-**

pyrethroid insecticides on the incidence of malaria in Segou Region, Mali: 2012-2015. *Malar J* 2018, **17:**19.

86. Jakubowski A, Stearns SC, Kruk ME, Angeles G, Thirumurthy H: **The US President's Malaria Initiative and under-5 child mortality in sub-Saharan Africa: A difference-in-differences analysis.** *PLoS Med* 2017, **14:**e1002319.

87. O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon J: **Estimating causal effects: Considering three alternatives to difference-in-differences estimation.** *Health Services and Outcomes Research Methodology* 2016, **16:**1-21.

88. Abadie A, Diamond A, Hainmueller J: **Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program.** *Journal of the American Statistical Association* 2010, **105:**493-505.

89. Khandker SR, Koolwal GB, Samad HA: **Handbook on impact evaluation: Quantitative methods and practices**. The World Bank; 2010. 10.1596/978-0-8213-8028-4.

90. Over M, Bakote'e B, Velayudhan R, Wilikai P, Graves PM: **Impregnated nets or DDT residual spraying? Field effectiveness of malaria prevention techniques in Solomon Islands, 1993-1999.** *The American Journal of Tropical Medicine and Hygiene* 2004, **71:**214-223.

91. Lim S, Fullman N, Stokes A, Ravishankar N, Masiye F, Murray C, Gakidou E: **Net benefits: A multicountry analysis of observational data examining associations between insecticide-treated mosquito nets and health outcomes.** *PLoS Medicine* 2011, **8**.

92. Boulay M, Lynch M, Koenker H: **Comparing two approaches for estimating the causal effect of behaviour-change communication messages promoting insecticide-treated bed nets: An analysis of the 2010 Zambia malaria indicator survey.** *Malar J* 2014, **13:**342.

93. Druetz T, Corneau-Tremblay N, Millogo T, Kouanda S, Ly A, Bicaba A, Haddad S: **Impact Evaluation of Seasonal Malaria Chemoprevention under Routine Program Implementation: A Quasi-Experimental Study in Burkina Faso.** *Am J Trop Med Hyg* 2018, **98:**524-533.

94. Moscoe E, Bor J, Barnighausen T: **Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice.** *J Clin Epidemiol* 2015, **68:**122-133.

95. VanderWeele T: **Mediation analysis: A practitioner's guide.** *Annu Rev Public Health* 2016, **37:**17-32.

96. Lange T, Hansen K, Sorensen R, Galatius S: **Applied mediation analyses: A review and tutorial.** *Epidemiol Health* 2017, **39**.

97. Imai K, Keele L, Tingley D: **A general approach to causal mediation analysis.** *Psychol Methods* 2010, **15:**309-334.

98. Broadbent A, Vandenbroucke JP, Pearce N: **Response: Formalism or pluralism? A reply to commentaries on 'Causality and causal inference in epidemiology'.** *Int J Epidemiol* 2016, **45:**1841-1851.

99. Vandenbroucke JP, Broadbent A, Pearce N: **Causality and causal inference in epidemiology: The need for a pluralistic approach.** *Int J Epidemiol* 2016, **45:**1776-1786.

100. Benmarhnia T, Huang J, Basu R, Wu J, Bruckner T: **Decomposition analysis of black-white disparities in birth outcomes: The relative contribution of air pollution and social factors in California.** *Environ Health Perspect* 2017, **125**.

101. Eckert E, Florey LS, Tongren JE, Salgado SR, Rukundo A, Habimana JP, Hakizimana E, Munguti K, Umulisa N, Mulindahabi M, Karema C: **Impact evaluation of malaria control interventions on morbidity and all-cause child mortality in Rwanda, 2000-2010.** *Am J Trop Med Hyg* 2017, **97:**99-110.

102. Perrin B: **Linking monitoring and evaluation to impact evaluation.** In *Impact evaluation notes*: InterAction and The Rockefeller Foundation; 2012.

103. **Elimination 8 (E8)** [www.malariaelimination8.org]

104. **Asia Pacific Leaders Malaria Alliance (APLMA)** [http://aplma.org/]

105. United Nations: **Principles and recommendations for a vital statistics system.** In *Series M No 19/ Rev3*. New York: Department of Economic and Social Affairs; 2014.

106. World Health Organization (WHO): **WHO malaria terminology.** Geneva, Switzerland: WHO; 2016.

107. MEASURE Evaluation: Surveillance, Monitoring, and Evaluation of Malaria Programs: Online Course. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. 2020. [https://www.measuremalaria.org/publications/surveillance-monitoring-and-evaluation-of-malaria-programs-online-course/]

108. World Health Organization (WHO): **The role of mass drug administration, mass screening and treatment, and focal screening and treatment for malaria.** WHO Global Malaria Programme; 2015.

ANNEX 1. ANNOTATED BIBLIOGRAPHY

Introduction

This annotated bibliography provides a review of the available guidance documents and tools, from 2001 to 2017, that are relevant to the evaluation of malaria programs. The review includes guidance documents, manuals, framework documents, peer-reviewed literature, reports, and PowerPoint presentations. The documents were identified through online database and website searches. Each document was reviewed to assess whether any guidance or framework was provided for national malaria programs (NMPs) or development partners to evaluate the impact of the program at a national or subnational level in moderate- to low-transmission settings.

The documents contained in this annotated bibliography encompass malaria strategies, monitoring and evaluation (M&E), impact evaluations, surveillance, and other guidance for NMPs. Global malaria strategies were reviewed for strategic approaches, frameworks, and indicators to monitor progress as programs move toward elimination. Additional frameworks and indicators were reviewed among the M&E documents currently available, which cover indicator guidance (malaria household surveys, integrated community case management, and behavior change communication) to frameworks used by development partners (the Global Fund, President's Malaria Initiative, and Roll Back Malaria). In addition to M&E, guidance on conducting malaria surveillance and evaluating these systems was reviewed for any useful tools. The peer-reviewed literature provided relevant analytical methods and novel metrics for evaluating programs and surveillance systems. Other additional documents include service availability and readiness assessment of health facilities and guidance for NMPs shifting from malaria control to elimination. Finally, impact evaluation guidance for NMPs, their malaria control activities, and interventions were reviewed for relevant frameworks, indicators, and data sources.

The documents reviewed show that there is comprehensive information for the M&E of malaria control programs and impact evaluations; however, most of the documents focused on high-transmission settings. None of the documents available provided a framework for malaria program evaluation or impact evaluation to assess malaria morbidity or mortality reduction in low- or moderate-transmission settings.

Guidance documents and tools

1. United States Centers for Disease Control and Prevention. (2001). Updated guidelines for evaluating public health surveillance systems: Recommendations from the guidelines working group. *Morbidity and Mortality Weekly Report:30 (No. RR-13)*. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm

   This manual provides guidance on how to conduct an evaluation of a public health surveillance system. It describes the specific tasks involved in carrying out an evaluation of a surveillance system, which includes the following: (1) engaging stakeholders in the evaluation, (2) describing the surveillance system to be evaluated, (3) focusing the evaluation design through defining the purpose of the evaluation and how the findings will be used, (4) gathering evidence regarding the performance of the surveillance system, (5) justifying and making conclusions and recommendations, and (6) ensuring use of the findings and sharing lessons learned. The guidelines discuss the following attributes of a surveillance system that should be assessed: simplicity, flexibility, data quality, acceptability, sensitivity, predictive value positive, representativeness, timeliness, and stability.

2. Maternal and Child Health Integrated Program (MCHIP). (2013). *Indicator guide: Monitoring and evaluating integrated community case management*. Washington, DC, USA: MCHIP. Retrieved from http://1rqxbs47ujl4rdy6q3nzf554.wpengine.netdna-cdn.com/wp-content/uploads/2016/07/iCCM-Indicator-Guide.pdf

   This document serves as a reference guide, providing a compilation of indicators for measuring key components of integrated community case management (iCCM) programs. These components include the following: (1) coordination and policy setting, (2) costing and financing, (3) human resources, (4) supply chain management, (5) service delivery and referral, (6) communication and social mobilization, (7) supervision and performance quality assurance, and (8) monitoring and evaluation and health management information systems. The document details the rationale for and purpose of the guide, discusses the process for developing the guide, reviews the categories of indicators included in the guide, provides detailed indicator reference sheets, and maps the indicators to an iCCM results framework.

3. MEASURE Evaluation, Demographic and Health Surveys, President's Malaria Initiative, Roll Back Malaria Partnership, UNICEF, and World Health Organization. (2018). *Household survey indicators for malaria control*. Chapel Hill, NC, USA: MEASURE Evaluation, University of North Carolina. Retrieved from http://www.malariasurveys.org/documents/Household%20Survey%20Indicators%20for%20Malaria%20Control_FINAL.pdf

   This document serves as a reference guide, providing detailed information on how to measure and interpret household survey indicators on malaria intervention coverage (prevention and case management), malaria morbidity, and all-cause child mortality. The current version (2013) includes updates to some of the previous indicators and discusses some of the main issues related to measurement of the indicators.

4. Mortality Task Force of the Roll Back Malaria's Monitoring and Evaluation Reference Group. (2014). *Guidance for evaluating the impact of national malaria control programs in highly endemic countries*. Rockville, MD, USA: MEASURE Evaluation. Retrieved from https://www.measureevaluation.org/resources/publications/ms-15-100/

   This document reviews and updates the evaluation framework that was proposed by the Roll Back Malaria Monitoring and Evaluation Reference Group in 2007, provides recommendations for how to evaluate the scale-up of malaria control interventions in highly malaria-endemic countries, and summarizes experience and data on malaria morbidity and mortality measurement gathered through various methods and data sources. The document provides guidance on processes for implementing an impact evaluation; discusses evaluation design options and a conceptual framework for assessing the impact of malaria control programs on malaria and all-cause child mortality; discusses what to cover in a program description and options for measuring malaria intervention coverage, malaria transmission intensity, malaria morbidity and mortality, and other key contextual factors; and provides recommendations for data synthesis, triangulation, and interpretation.

5. Noor, A. (2017). *Malaria surveillance, monitoring and evaluation manual*. Presented at the Malaria Policy Advisory Committee Meeting, Geneva, Switzerland.

The presentation provides an overview of the updated malaria surveillance, monitoring, and evaluation manual, which combines the 2012 control and elimination operational manuals into one document and is aligned to the Global Technical Strategy 2016-2030 and the 2017 Elimination Framework. The manual includes new sections providing guidance on surveillance in the private and community health sectors and for migrant and mobile populations. It covers monitoring and evaluation guidance for national programs, the Global Technical Strategy for Malaria 2016-2030, and surveillance systems. The manual discusses the adaptations of the surveillance system through the continuum of malaria transmission; key concepts and practice of malaria surveillance systems; the establishment of surveillance systems for malaria elimination; and use of surveillance, surveys, and other data for monitoring and evaluating national programs and the Global Technical Strategy.

6. President's Malaria Initiative (PMI). (2015). *President's Malaria Initiative strategy: 2015–2020*. Washington, DC: PMI. Retrieved from https://www.pmi.gov/docs/default-source/default-document-library/pmi-reports/pmi_strategy_2015-2020.pdf?sfvrsn=24

   This document outlines the President's Malaria Initiative's (PMI) vision, guiding principles, goal, objectives, and strategic approach for 2015-2020. It provides an overview of the core areas of strategic focus for PMI, which include the following: (1) achieving and sustaining scale of proven malaria control and prevention interventions, (2) adapting to changing epidemiology and incorporating new tools, (3) improving country capacity to collect and use information, (4) mitigating risk against the current malaria control gains through monitoring the development and spread of insecticide and drug resistance, and (5) building capacity and health systems of countries to effectively implement their national malaria programs. The document also describes the critical assumptions for achieving PMI's goal and objectives, PMI's core operating principles, and an overview of PMI governance and management.

7. President's Malaria Initiative (PMI). (2017). *President's Malaria Initiative technical guidance*. Washington, DC: PMI and United States Agency for International Development. Retrieved from https://www.pmi.gov/docs/default-source/default-document-library/tools-curricula/pmi-technical-guidance-(march-2016).pdf

   This guidance document is intended to be a resource for President's Malaria Initiative (PMI) staff in helping draft annual Malaria Operational Plans and for PMI country teams as a technical reference document. It provides technical guidance on key malaria control and prevention interventions, which include vector monitoring and control, malaria in pregnancy, vaccines and other preventive approaches including seasonal malaria chemoprevention and intermittent preventive treatment in infants, and case management. It also provides guidance on malaria surveillance and monitoring and evaluation, operational research, commodity procurement and supply chain management, and programming guidance for countries moving toward pre-elimination. The guidance reflects the latest global policies and most recent up-to-date guidance for malaria programming.

8. Roll Back Malaria Monitoring Evaluation Reference Group (RBM MERG). (2005). *Building capacity in monitoring and evaluating Roll Back Malaria in Africa: A conceptual framework for the Roll Back Malaria*

Partnership. Geneva, Switzerland: RBM MERG. Retrieved from http://www.rollbackmalaria.org/wp-content/uploads/2017/08/merg_ConceptualFramework.pdf

This document was developed out of a recognition of the need for improved country capacity to monitor and evaluate the progress and impact of malaria control investments and programs. The document identifies the key functions of a national monitoring and evaluation (M&E) system, reviews current issues and opportunities that exist at the country level, and provides recommendations on the capacities that should be built to carry out M&E at the country level for different malaria epidemiological settings. The document describes the key malaria data sources and products of an M&E system, which include properly managed malaria data, monthly monitoring reports, quarterly review reports, national malaria meetings to review progress and program planning, an annual malaria review and report, and periodic evaluation reports. It also describes the recommended positions and components of an M&E unit in the national malaria program and the different roles and responsibilities and other needs for a functioning M&E system.

9. Roll Back Malaria (RBM) Partnership. (2014). *Malaria behavior change communication (BCC) indicator reference guide*. Geneva, Switzerland: RBM. Retrieved from https://www.rollbackmalaria.org/wp-content/uploads/2017/08/Malaria-BCC-Indicators-Reference-Guide.pdf

   This document serves as a reference guide, providing detailed information on how to measure and interpret household survey indicators on malaria targeted behaviors; reach and exposure to malaria messages; knowledge and awareness of the cause, symptoms, treatment, and preventive measures for malaria; risk and efficacy of malaria and malaria preventive behaviors; norms and attitudes related to malaria practices and behaviors; and other experimental indicators.

10. Roll Back Malaria (RBM) Partnership & United Nations Development Programme (UNDP). (2013). *Multisectoral action framework for malaria*. Geneva, Switzerland: RBM and UNDP. Retrieved from https://endmalaria.org/sites/default/files/9_Multisectoral-Action-Framework-for-Malaria.pdf

    This framework document describes the major determinants of malaria at the societal, environmental, population, and household and individual levels; and calls for a multisectoral response (e.g., adding a development dimension to the response) to address the key social and environmental determinants of malaria. It discusses which sectors should be involved in the response and maps these sectors to the key determinants. It also provides project and country-specific examples of multisectoral responses to malaria. The document notes the current limitations with global monitoring and evaluation guidance for malaria, which is based on a biomedical response and does not capture efforts outside the health sector. It also discusses the financing of the type of response, key knowledge gaps and research needs to inform the operationalization of the framework, and key immediate next steps for beginning to operationalize and implement the framework.

11. Roll Back Malaria Partnership Monitoring and Evaluation Reference Group (RBM MERG). (n.d.) *Assessing the impact of malaria control activities on mortality among African children under 5 years of age*. Geneva, Switzerland: RBM MERG. Retrieved from https://www.rollbackmalaria.org/wp-content/uploads/2017/08/MERGGuidanceNote_MalariaImpactAssessment.pdf

This technical note provides guidance on how best to assess the impact of malaria control interventions on mortality among African children under five. The recommendations put forth include at a minimum monitoring coverage of malaria intervention coverage and all-cause child mortality from population-based household surveys (Multiple Indicator Cluster Surveys and Demographic and Health Surveys), and using the Child Health Epidemiology Reference Group child survival impact model to assess malaria-specific mortality burden. The technical note also states that, where available, data from local research project or sentinel surveillance sites should be assessed and malaria data from health information and vital registration systems should be reviewed.

12. Rowe, A. K., Steketee, R. W., Arnold, F., Wardlaw, T., Basu, S., Bakyaita, N., . . . Roll Back Malaria Monitoring Evaluation Reference Group. (2007). Viewpoint: Evaluating the impact of malaria control efforts on mortality in sub-Saharan Africa. *Tropical Medicine & International Health, 12*(12), 1524–1539. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18076561

    This paper describes a framework for evaluating the impact of malaria control efforts on all-cause childhood mortality. The approach uses an ecologic study design with a plausibility argument and examines trends in coverage of malaria control interventions, other factors that influence childhood mortality, malaria-associated morbidity, and all-cause childhood mortality. The paper describes potential sources for mortality data (population-based household surveys, demographic surveillance systems with verbal autopsy, sample or sentinel vital registration systems, mortality surveys with verbal autopsy, routine health facility data, and mathematical models) and the key attributes and limitations of each source. It also discusses the limitations of the evaluation framework/approach.

13. The Global Fund to Fight AIDS, Tuberculosis and Malaria (Global Fund). (n.d.) *The Global Fund's approach to monitoring and evaluation*. Geneva, Switzerland: Global Fund. Note: This document is no longer available online.

    This document describes how the Global Fund uses results and its guiding principles for monitoring and evaluation (M&E). It gives a broad overview of the Global Fund's approach for M&E and discusses measurement of impact of its programs and guidance and tools available for its program recipients. It discusses the Global Fund's strategy for measuring quality of data and quality of services provided, its framework for strengthening M&E systems, the reporting of results, and its involvement and support of M&E at the country level with its principal recipients. The document also provides a list of core indicators used by the Global Fund to report on its programs for HIV, malaria, tuberculosis, and health systems and community systems strengthening.

14. The Global Fund to Fight AIDS, Tuberculosis and Malaria (Global Fund). (2017). *Indicator guidance sheets: Malaria*. Geneva, Switzerland: Global Fund. Retrieved from https://www.theglobalfund.org/media/5195/me_indicatorguidancesheets-annexc-malaria_sheet_en.xlsx?u=636637835290000000

    The Global Fund indicator guidance sheets discuss how different indicators (coverage and output, and impact and outcome level) will be used for programmatic decision-making. For each indicator, the definition, disaggregation, geographical coverage, data source, data collection and reporting frequency, and analysis and interpretation are provided.

15. The Global Fund to Fight AIDS, Tuberculosis and Malaria (Global Fund). (2017). *Modular framework handbook*. Geneva, Switzerland: The Global Fund. Retrieved from https://www.theglobalfund.org/media/4309/fundingmodel_modularframework_handbook_en.pdf

    This handbook provides an overview of Global Fund's modular approach, which is used to organize and track programmatic and financial information for each grant recipient. The framework consists of several modules across the three disease areas and health and community system strengthening, and it reviews the broad program intervention areas of the four areas (HIV, tuberculosis, malaria, and health and community system strengthening). The framework includes indicators for monitoring progress in each module.

16. University of California, San Francisco Malaria Elimination Initiative. (n.d.) *Reactive Case Detection (RACD) Monitoring & Evaluation Tool*. Retrieved from http://www.shrinkingthemalariamap.org/tools/reactive-case-detection-monitoring-evaluation-tool

    The Reactive Case Detection Monitoring and Evaluation tool assesses the completeness and timeliness of active surveillance and response. It also examines the key components of active surveillance activities, assesses the strengths and gaps of an active surveillance program, and evaluates the cost of carrying out reactive case detection.

17. United States Agency for International Development & World Health Organization (WHO). (2012). *Measuring service availability and readiness: A health facility assessment methodology for monitoring health system strengthening: Service readiness indicators*. Geneva, Switzerland: WHO. Retrieved from https://www.who.int/healthinfo/systems/SARA_Reference_Manual_Chapter3.pdf?ua=1

    This guide provides a comprehensive set of indicators for measuring health facility service readiness. It includes indicators for general service readiness across the following domains: basic amenities, basic equipment, standard precautions for infection prevention, diagnostic capacity, and essential medicines. It also includes service-specific readiness indicators across 19 service areas, including malaria, and indicators for measuring the availability of maternal and child health priority medicines.

18. World Health Organization (WHO). (2017). *Data requirements and methods to support the evaluation of new vector control products*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/requirements-vector-control-products/en/

    This document describes the updated World Health Organization guidance on data requirements and methods for evaluating new vector control products. It gives general recommendations for efficacy testing criteria for new indoor residual spraying, space spray, and larvicide products; and further evaluation requirements when new products do not meet the laid out criteria. The document also outlines important areas for further research to inform the evidence base on new vector control products both in terms of testing their entomological and epidemiological efficacy.

19. World Health Organization (WHO). (2007). *Malaria elimination: A field manual for low and moderate endemic countries*. Geneva, Switzerland: WHO. Retrieved from https://www.scribd.com/document/

54924826/WHO-Malaria-Elimination-A-Field-Manual

This manual describes what is required for malaria elimination; the major program reorientations and approaches when going from malaria control to elimination and to prevention of reintroduction; and interventions, milestones, indicators, and programmatic issues for each of these three phases. It discusses the feasibility of malaria elimination and the tools and approaches specific to elimination programs. It also provides a monitoring and evaluation framework for assessing progress toward malaria elimination for pre-elimination and elimination programs, with guidance on potential indicators and data sources or methods. It also discusses what is needed to prevent the re-establishment of malaria and discusses the requirements for World Health Organization certification of malaria elimination.

20. World Health Organization (WHO). (2010). *Malaria programme reviews: A manual for reviewing the performance of malaria control and elimination programmes*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/whomprmalariaprogramperformancemanual/en/

    This manual provides guidance on how to conduct a Malaria Program Review (MPR), which is a joint collaborative evaluation of the national malaria control or elimination program. The manual reviews the main objectives of the MPR, which are as follows: review the epidemiology of malaria in the country; review the structure, organization, management framework for malaria policy and program development; assess progress toward achievement of targets; review current program performance by intervention thematic areas and service delivery levels; and define steps for improving program performance. The manual describes the timing, scope, and structure of MPR; the methods used; and provides guidance for planning an MPR, conducting the thematic desk review and the field review, and writing and disseminating the report.

21. World Health Organization (WHO). (2011). *Monitoring, evaluation and review of national health strategies: A country-led platform for information and accountability*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/healthinfo/country_monitoring_evaluation/1085_IER_131011_web.pdf?ua=1

    This document provides guidance to countries and development partners on how to strengthen the monitoring, evaluation, and review of national health plans and strategies. It outlines key attributes and characteristics across four key areas for a strong country-led platform for monitoring, evaluation, and review of the performance of the health sector. The four key areas outlined are as follows: (1) the national health strategy, which serves as the basis for information and accountability; (2) institutional capacity to support regular monitoring, review, and action; (3) the monitoring and evaluation system; and (4) the establishment of country mechanisms for review and action. The intended uses of the guidance document are to assess, improve, or develop the monitoring and evaluation component of the national health plan or strategy or a specific health sector program, or to evaluate health system strengthening interventions.

22. World Health Organization (WHO). (2012). *Community-based reduction of malaria transmission. Consultation report*. Geneva, Switzerland: WHO. Retrieved from https://www.who.int/malaria/publications/atoz/9789241502719/en/

This document describes an approach for implementing a comprehensive package of community-based malaria interventions for transmission reduction and pre-elimination settings. It discusses assessing transmission as an initial preparatory step for this type of approach, appropriate intervention packages for different transmission settings, factors for a supportive environment for community-based interventions, how to engage communities in this work, and monitoring and evaluating community-based approaches for malaria transmission reduction. It also discusses areas for additional research to help expand and inform these approaches, including research on refining the elements of a community-based approach; diagnostic tests to be used in community-based interventions; community health worker performance and retention; and issues related to the use of drugs for mass drug administration, mass screening and treatment, focal screening and treatment, and high focal screening and treatment interventions.

23. World Health Organization (WHO). (2012). *Disease surveillance for malaria control: Operational manual.* Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/9789241503341/en/

    This manual describes the objectives and features of a surveillance system in high/moderate, low, and very low malaria transmission settings. It provides definitions of malaria surveillance concepts, reviews surveillance indicators, describes limitations of surveillance data specific to settings that are in the malaria control phase, and reviews the key objectives of the gathering surveillance data. It details how surveillance data should be recorded and reported at the different levels of the health system. It provides operational guidance for establishing a surveillance system in the control phase, covering the different tools, procedures, and human resources and structures needed for a surveillance system.

24. World Health Organization (WHO). (2012). *Disease surveillance for malaria elimination: Operational manual.* Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/9789241503334/en/

    This manual describes the objectives and features of a surveillance system in high/moderate, low, and very low malaria transmission settings. It provides definitions of malaria surveillance concepts specific to settings that are in the malaria elimination phase. It details how surveillance data should be recorded, reported, and analyzed at the different levels of the health system for the elimination phase. It provides operational guidance for establishing a surveillance system specific to the pre-elimination phase and reviews the certification of elimination and surveillance in the prevention of reintroduction phase.

25. World Health Organization (WHO). (2013). *Service Availability and Readiness Assessment: An annual monitoring system for service delivery: Reference manual.* Geneva, Switzerland: WHO. Retrieved from https://www.who.int/healthinfo/systems/sara_reference_manual/en/

    This manual describes the Service Availability and Readiness Assessment (SARA), which is aimed at generating information on service availability, general service readiness, and service-specific readiness at the health facility level. It reviews the background on SARA, the objectives of the assessment, and discusses the key focus areas assessed and the methodology used for the assessment. It also provides guidance for implementing the SARA, including the steps for preparing for the survey, planning and

implementing the survey, and guidance for data collection, entry, and analysis. The manual includes the core instrument and provides an indicator index for all indicators collected in the SARA.

26. World Health Organization (WHO). (2014). *From malaria control to malaria elimination: A manual for elimination scenario planning*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/9789241507028/en/

    This manual reviews key concepts related to malaria elimination and discusses the technical, operational, and financial feasibility to reduce transmission and then ultimately to achieve elimination. It describes data sources and methods for estimating a baseline for malaria prevalence, which can be used to analyze the feasibility of elimination and for planning for moving toward elimination. It also describes different scenarios through which malaria transmission can be reduced, from baseline levels to a low level of transmission and then to elimination.

27. World Health Organization (WHO). (2015). *Control and elimination of Plasmodium vivax malaria: A technical brief*. Geneva, Switzerland: WHO. Retrieved from https://www.who.int/malaria/publications/atoz/9789241509244/en/

    This technical brief reviews the geographical distribution and burden of *P. vivax* malaria and biological characteristics and challenges for control and elimination. It describes the strategies currently in use to control and eliminate *P. vivax*, specifically focusing on vector control, chemoprevention, diagnosis of *P. vivax* and G6PD deficiency, treatment of uncomplicated and severe *P. vivax*, drug resistance, and surveillance. It describes challenges in the area of *P. vivax* control and elimination, highlights areas where further research is needed in understanding the biology and epidemiology of *P. vivax* malaria, and provides suggestions of innovations and tools needed to be able to successfully control and eliminate *P. vivax* malaria.

28. World Health Organization (WHO). (2015). *Strategy for malaria elimination in the Greater Mekong Subregion (2015–2030)*. Geneva, Switzerland: WHO. Retrieved from http://iris.wpro.who.int/bitstream/handle/10665.1/10945/9789290617181_eng.pdf

    This strategy document outlines the vision, goals, principles, objectives, and milestones and targets for malaria elimination in the Greater Mekong Subregion for 2015-2030. It details the three key intervention areas and the two supporting elements required to achieve the goal and targets, which are aligned with the Global Technical Strategy and include the following: (1) case detection and management, (2) disease prevention in transmission areas, (3) malaria case and entomological surveillance, (4) expansion of research for innovation and improved delivery of services (supporting element #1), and (5) strengthening of the enabling environment (supporting element #2). Furthermore, it proposes key activities to undertake to achieve the three objectives outlined in the strategy. The strategy describes in broad terms the focus of monitoring and evaluation efforts for programs in the region, highlights key issues to monitor and required information to be fed into a malaria elimination database, and discusses steps needed for strengthening monitoring and reporting. It also discusses guiding principles for governance and coordination for malaria efforts in the region.

29. World Health Organization (WHO). (2016). *Global technical strategy for malaria 2016–2030*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/areas/global_technical_strategy/en/

This document provides a framework for the development of malaria programs and strategies for the accelerated progress toward malaria elimination. Global goals and targets for malaria burden reduction are outlined for 2020, 2025, and 2030. The framework is based on three pillars with two supporting elements, which include the following: ensuring universal access to malaria prevention, diagnosis, and treatment (pillar 1); accelerating efforts toward elimination and attainment of malaria-free status (pillar 2); transforming malaria surveillance into a core intervention (pillar 3); harnessing innovation and expanding research (supporting element 1); and strengthening the enabling environment (supporting element 2). The document provides guidance on the minimum set of outcome and impact indicators that should be assessed to track progress toward the set goal and targets.

30. World Health Organization (WHO). (2016). *WHO malaria terminology*. Geneva, Switzerland: WHO. Retrieved from https://www.who.int/malaria/publications/atoz/malaria-terminology/en/

This document provides a comprehensive glossary of malaria terminology. It also describes the process used by the World Health Organization (WHO) Global Malaria Programme to update the WHO terminology of malaria.

31. World Health Organization (WHO). (2017). *Data Quality Review (DQR) Toolkit Module 1: Framework and metrics*. Geneva, Switzerland: WHO. Retrieved from http://apps.who.int/iris/bitstream/10665/259224/1/9789241512725-eng.pdf?ua=1

This toolkit provides guidance and tools for carrying out routine, annual, and periodic data quality reviews (DQRs) to assess the quality of health facility data. The toolkit consists of three modules: (1) framework and metrics, (2) desk review of data quality, and (3) data verification and system assessment. The methodology for the DQR includes a desk review of the data reported to the national level and health facility assessment to conduct a data verification and system assessment exercise. The first module discusses the framework for the DQR and recommended indicators that should be assessed in the DQR. It also discusses the dimensions of data quality and metrics for assessing the different dimensions of data quality. It also describes a process for implementing the DQR and disseminating and using the results from the DQR.

32. World Health Organization (WHO). (2017). *Data Quality Review (DQR) Toolkit Module 2: Desk review of data quality*. Geneva, Switzerland: WHO. Retrieved from http://apps.who.int/iris/bitstream/10665/259225/1/9789241512732-eng.pdf?ua=1

This toolkit provides guidance and tools for carrying out routine, annual, and periodic data quality reviews (DQRs) to assess the quality of health facility data. The toolkit consists of three modules: (1) framework and metrics, (2) desk review of data quality, and (3) data verification and system assessment. The second module covers how to prepare for and implement the desk review of data quality, how to review the data requirements, and methods for data collection, formatting, and compiling for the review. It also describes the analysis, output, and interpretation of desk review data.

33. World Health Organization (WHO). (2017). *Data Quality Review (DQR) Toolkit Module 3: Data verification and system assessment*. Geneva, Switzerland: WHO. Retrieved from http://apps.who.int/iris/bitstream/10665/259226/1/9789241512749-eng.pdf?ua=1

This toolkit provides guidance and tools for carrying out routine, annual and periodic data quality reviews (DQRs) to assess the quality of health facility data. The toolkit consists of three modules: (1) framework and metrics, (2) desk review of data quality, and (3) data verification and system assessment. The third module provides an overview of measuring data quality through health facility assessments and reviews the recommended core indicators for the assessment and the dimensions of data quality to be assessed. It provides guidance on preparing for and implementing the data verification and system assessment at the sampled health facilities and reviews the analysis to be completed and how to interpret the results from the assessment.

34. World Health Organization (WHO). (2017). *A framework for malaria elimination*. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/malaria/publications/atoz/9789241511988/en/

This framework discusses the key principles of malaria elimination and key strategies and interventions for malaria elimination, which are mapped to the Global Technical Strategy for Malaria 2016-2030. It provides guidance and considerations for managing and planning an elimination program, which includes the following: conducting an assessment of the malaria program; developing several costed plans (strategic, elimination, operational, and monitoring and evaluation); continual monitoring and use of data for programmatic decision-making; establishing an independent national malaria elimination advisory committee; and ensuring a supportive enabling environment for elimination. It discusses what is needed for preventing re-establishment of malaria and the requirements for certification and verification of malaria elimination. It also discusses a research agenda for malaria elimination and key areas for operational research. It provides a list of monitoring and evaluation indicators for an elimination program, with guidance on targets/norms and data sources.

Relevant manuscripts

1. Ashton, R. A., Bennett, A., Yukich, J., Bhattarai, A., Keating, J., & Eisele, T. P. (2017). Methodological considerations for use of routine health information system data to evaluate malaria program impact in an era of declining malaria transmission. *American Journal of Tropical Medicine and Hygiene, 97*(3_Suppl), 46–57. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28990915

In this study, the authors conducted a literature review to identify how routine health information system (HMIS) data have been used in malaria impact evaluations. They describe the different methods identified in the studies for malaria impact and outcome evaluations, including pre-post intervention comparisons, descriptive analyses of trend, interrupted time series design, and subnational dose-response. The authors argue that interrupted time series design and dose-response analyses are the strongest quasi-experimental design options for impact and outcome evaluations using HMIS data. They also present methods that can help maximize the internal validity of HMIS data and provide recommendations for reducing bias in impact estimates.

2. Calba, C., Goutard, F. L., Hoinville, L., Hendrikx, P., Lindberg, A., Saegerman, C., & Peyre, M. (2015). Surveillance systems evaluation: A systematic review of the existing approaches. *BMC Public Health, 15*(1), 448. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25928645

   In this study, the authors conducted a systematic review to identify and analyze existing health surveillance system evaluation approaches, and specifically to assess the advantages, limitations, and existing gaps in current approaches. Four common steps emerged in the evaluation process across the different approaches, including describing the context, describing the evaluation process, implementing the evaluation, and providing recommendations. The focus of the approaches varied; some focused on the evaluation of the structure of the system, but the majority included an evaluation of the quality of the data generated and the system's performance. A key limitation noted by the authors was the lack of detail provided to the evaluators on how to practically implement the evaluation; in other words, the guidance was too generic and often lacked information on methods and tools for the implementation of the evaluation. Another key limitation observed was the lack of a comprehensive list of attributes to assess as well as guidance on which attributes should be assessed based on the surveillance objectives.

3. Churcher, T. S., Cohen, J. M., Novotny, J., Ntshalintshali, N., Kunene, S., & Cauchemez, S. (2014). Public health. Measuring the path toward malaria elimination. *Science, 344*(6189), 1230–1232. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24926005

   In this article, the authors propose new methods for assessing progress toward elimination, arguing that evaluation of programs working toward elimination should not rely on an approach whereby success is defined only as achieving no locally acquired malaria cases. Rather, a meaningful evaluation should take into account the local and regional epidemiological circumstances. The authors propose monitoring and evaluating the status of controlled non-endemic malaria in an area, rather than just looking at the number of malaria cases. To do this, they propose measuring the proportion of imported cases among detected cases and whether it is above a certain threshold (e.g., reproductive rate [R_0] is greater than or equal to 1). They also note that it is important to assess R_0 by season in areas that experience seasonal transmission. The authors note important limitations to these metrics, mainly that more work is needed to understand at which spatial resolution (subnational area) this type of analysis should be carried out, given the large spatial heterogeneity in malaria transmission. Furthermore, they also note that this metric was developed for *falciparum* malaria and not for *vivax* malaria, noting that the threshold metric may not have sufficient power.

4. Florey, L. S., Bennett, A., Hershey, C. L., Bhattarai, A., Nielsen, C. F., Ali, D., . . . Yé, Y. (2017). Impact of insecticide-treated net ownership on all-cause child mortality in Malawi, 2006-2010. *American Journal of Tropical Medicine and Hygiene, 97*(3_Suppl), 65–75. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28990922

   This study presents two analytical approaches for assessing the association between increasing insecticide-treated net (ITN) coverage and all-cause child mortality (ACCM) over time, using data from the 2006 Malawi Multiple Indicator Cluster Survey and the 2010 Malawi Demographic and Health Survey. The first

approach used a retrospective cohort analysis of individual children, modeled through a Cox proportional hazards framework that controlled for various environmental, household, and individual confounders. The second approach assessed the population-level association between insecticide-treated net (ITN) ownership and ACCM, using a district-level ecologic analysis using negative binomial regression. The findings showed a significant association between ITN ownership and ACCM and suggest that increasing ITN ownership may have contributed to the observed decrease in ACCM between the two survey period (2006 and 2010).

5. Stresman, G., Cameron, A., & Drakeley, C. (2017). Freedom from infection: Confirming interruption of malaria transmission. *Trends in Parasitology, 33*(5), 345–352. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28108199

   The authors apply a method developed in veterinary epidemiology—freedom from infection—to malaria. This method may be used to produce reliable estimates for the probability of detecting the disease if present at a defined low level and also to inform malaria program decision-making.

6. Yé, Y., Eisele, T. P., Eckert, E., Korenromp, E., Shah, J. A., Hershey, C. L., . . . Bhattarai, A. (2017). Framework for evaluating the health impact of the scale-up of malaria control interventions on all-cause child mortality in sub-Saharan Africa. *American Journal of Tropical Medicine and Hygiene, 97*(3 Suppl), 9–19. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28990923

   This article describes an updated framework for evaluating the impact of the scale-up of malaria control interventions on all-cause child mortality in highly endemic countries in sub-Saharan Africa. It builds on and expands the plausibility design evaluation framework that was proposed by the Roll Back Malaria Monitoring and Evaluation Reference Group in 2007, by incorporating further experience from the implementation of the framework in several countries. Specifically, it expands the framework through the inclusion of risk stratification to examine subgroups that are more likely to be greater impacted by the expansion of malaria interventions, use of a national platform framework for carrying out evaluations, and analysis of complete birth histories to assess child survival over the evaluation period. The article describes the evaluation questions, design, indicators and data sources, appropriate timing for the evaluation, and analytic methods. It also discusses the strengths and limitations of the framework.

ANNEX 2. INDICATOR REFERENCE GUIDE

Table A2.1. Monitoring and evaluation core indicator reference guide

Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Input					
Expenditure per capita for malaria control or elimination [9]	Malaria expenditure (domestic and international)	Population at risk of malaria	Routine administrative and financial systems	Source of funds (e.g., domestic, private sector, household, international), program area, geographical area	Direct malaria expenditure can be reported if expenditures shared with other programs cannot be readily apportioned to malaria
Human resources: Number of health workers per 10,000 population [22]	Number of health workers X 10,000	Population size	Census, routine administrative systems	Geographical area	
Annual number of malaria commodities procured by type	Number of malaria commodities procured during one year		Routine program reporting	Type of commodity	
Output					
Social and behavior change communication					
Number and proportion of population at risk who recall hearing or seeing malaria messages within the past six months	Number of people at risk of malaria who recall hearing or seeing any malaria message during the last six months	Total number of survey respondents	Population-based household survey	Geographical area, age, sex	
Vector control					
Number of insecticide-treated nets (ITNs) distributed	Number of ITNs distributed during one year		Routine program reporting	Geographical area	
Number and proportion of households targeted for	Number of households targeted for IRS that	Total number of households targeted for IRS	Routine program reporting	Geographical area	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	indoor residual spraying (IRS) that received IRS	received IRS during the reporting period	during the reporting period			
	Number of areas targeted for larviciding that are covered	Number of areas targeted for larviciding that are covered during the reporting period		Routine program reporting	Geographical area	
	Number of entomological monitoring sites	Number of entomological monitoring sites		Routine program reporting		
Chemoprevention						
	Number of sulfadoxine-pyrimethamine doses delivered for IPTp	Number of sulfadoxine-pyrimethamine doses delivered during the reporting period		Logistics management information system	Geographical area	
	Number of children ages 3–59 months who received the full number of courses of seasonal malaria chemoprevention (SMC) per transmission season [9]	Number of children ages 3–59 months who received the full number of courses of SMC in a transmission season		Routine health information system, census	Geographical area, type of facility, sex	
Diagnostic reporting						
	Number and proportion of health facilities with microscopy or rapid diagnostic test (RDT) capability	Number of health facilities with microscopy or RDT capability	Total number of health facilities	Routine administrative reporting	Geographical area, type of facility	
	Number of blood slides taken and read	Number of blood slides taken and read		Routine health information system	Geographical area, type of facility	
	Number of RDTs done and read	Number of RDTs done and read		Routine health information system	Geographical area, type of facility	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Number of microscopy slides cross-checked by national reference laboratory	Number of microscopy slides cross-checked by national reference laboratory		Routine health information system, routine program reporting	Geographical area, type of facility	
Treatment						
	Number of first-line antimalarial treatment courses administered	Number of first-line antimalarial treatment courses administered during the reporting period		Logistics management information system	Age, sex	
	Number of pre-referral treatment courses administered	Number of pre-referral treatment courses administered during the reporting period		Logistics management information system	Age, sex	
	Number of radical cure treatment courses (primaquine or tafenoquine) administered (*P. vivax* settings)	Number of radical cure treatment course (primaquine or tafenoquine) administered during the reporting period		Logistics management information system	Age, sex	
	Number of single, low-dose primaquine treatment courses administered for *P. falciparum* transmission blocking	Number of single, low-dose primaquine treatment courses administered		Logistics management information system	Age, sex	
	Number of severe malaria cases referred	Number of severe malaria cases referred		Logistics management information system	Age, sex	
	Number of antimalarial treatment courses for severe malaria cases administered	Number of antimalarial treatment courses for severe malaria cases administered during the reporting period		Logistics management information system	Age, sex	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Commodities						
	Number of health facilities with stockouts of key commodities for diagnostic testing	Number of health facility reports received on time during the month		Logistics management information system, health facility survey	Geographical area, type of facility	
	Number of health facilities with stockouts of key malaria drugs	Number of health facility reports received that are complete during the month		Logistics management information system, health facility survey	Geographical area, type of facility	
Surveillance						
	Number and proportion of expected health facilities reports received on time	Number of health facility reports received on time during the month	Number of health facilities	Routine health information system	Geographical area, type of facility	
	Number and proportion of expected health facility reports received that are complete	Number of health facility reports received that are complete during the month	Number of health facilities	Routine health information system, data quality audit	Geographical area, type of facility	May require a data quality audit to assess the completeness of forms
Training and supervision						
	Number and proportion of health facilities with a trained clinician in case management	Number of health facilities with a trained clinician in case management	Number of health facilities	Routine administrative reporting	Geographical area, type of facility	
	Number and proportion of health facilities with staff trained in surveillance, monitoring, and evaluation	Number of health facilities with staff trained in surveillance, monitoring, and evaluation	Number of health facilities	Routine administrative reporting	Geographical area, type of facility	
	Number and proportion of health facilities that received supervisory visits in the reporting period	Number of health facilities that received a supervisory visit during the reporting period	Number of health facilities	Routine program reporting	Geographical area, type of facility	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Drug and insecticide efficacy monitoring						
	Number of drug efficacy studies completed	Number of drug efficacy studies completed during the reporting period		Routine program reporting	Geographical area	
	Number of insecticide efficacy studies completed	Number of insecticide efficacy studies completed during the reporting period		Routine program reporting	Geographical area	
Outcome						
Malaria knowledge						
	Proportion of population at risk who know the main symptom of malaria	Number of people who know that the main sign/symptom of malaria is fever	Number of people surveyed	Population-based household survey	Geographical area, age, sex	
	Proportion of population at risk who know the treatment for malaria	Number of people who know that the appropriate treatment for malaria is artemisinin-based combination therapy	Number of people surveyed	Population-based household survey	Geographical area, age, sex	
	Proportion of population at risk who know preventive measures for malaria	Proportion of people who know that the primary preventive measures for malaria include using bed nets, taking preventive medication during pregnancy, taking seasonal prophylaxis, or having house sprayed with insecticide	Number of people surveyed	Population-based household survey	Geographical area, age, sex	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Vector control						
	Proportion of population at risk with access to an ITN in their household	Total number of individuals at risk for malaria who could sleep under an ITN if each ITN in the household is used by two people	Total number of individuals at risk of malaria who spent the previous night in surveyed households	Population-based household survey	Geographical area, age, sex, urban or rural, wealth index, household size	
	Proportion of population at risk that slept under an ITN the previous night	Number of individuals sleeping under an ITN the previous night	Total number of individuals who spent the previous night in surveyed households	Population-based household survey	Geographical area, urban or rural, wealth index, educational status, pregnancy status, age, sex, household size	
	Proportion of population at risk protected by IRS during previous 12 months	Number of people protected by IRS in the previous 12 months	Population at risk of malaria	National malaria program records, census	Geographical area, age, sex	
	Proportion of population at risk with access to an ITN in their household	Number of people at risk with access to an ITN in their household	Population at risk of malaria	Routine program reporting	Geographical area, urban or rural, wealth index, household size	
	Proportion of adult female vectors alive after exposure to insecticide (resistance frequency)	Number of dead or incapacitated *Anopheles* malaria vector	Total number of *Anopheles* malaria vectors exposed to a discriminating concentration of insecticide in standard bioassays	Special study		

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Resistance to insecticide status	Number of *Anopheles* malaria vectors confirmed resistant, possibly resistant, or susceptible	Total number of *Anopheles* malaria vectors exposed to a discriminating concentration of insecticide in standard bioassays	Special study		
Chemoprevention						
	Proportion of pregnant women who received three or more doses of intermittent preventive treatment in pregnancy	Number of women who received three or more doses of intermittent preventive treatment in pregnancy	Number of expected pregnancies (routine health information system) Number of women ages 15–49 surveyed who had a live birth in the last two years (population-based household survey)	Routine health information system, population-based household survey	Geographical area, age	
	Proportion of eligible children ages 3–59 months who received the full number of courses of SMC per transmission season	Number of children ages 3–59 months who received the full number of courses of SMC in a transmission season	Number of children ages 3–59 months requiring SMC	Routine health information system, census	Geographical area, age, sex	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Diagnostic testing						
	Proportion of patients tested among all febrile patients	Number of febrile patients tested for malaria	Number of all febrile patients	Routine health information system, population-based survey, health facility survey	Geographical area, type of facility, age, sex	
	Proportion of cases confirmed by a parasitological test of all reported cases	Number of cases confirmed by a parasitological test	Number of reported cases	Routine health information system, health facility survey	Geographical area, type of facility, age, sex	
	Proportion of health facilities without stockouts of key commodities for diagnostic testing	Number of health facility months with no stockouts of key commodities for diagnostic testing	Number of health facility months	Routine health information system, health facility survey	Geographical area, type of facility	
	Proportion of microscopy results cross-checked by national reference laboratory	Number of microscopy results cross-checked by national reference laboratory	Total number of microscopy results	Routine health information system, health facility survey	Geographical area, type of facility	Disaggregate by positive and negative results
	Proportion of microscopists achieving both sensitivity and specificity greater than 90 percent during proficiency tests	Number of microscopists achieving both sensitivity and specificity greater than 90 percent during proficiency tests	Total number of microscopists assessed through proficiency tests		Geographical area, type of facility	
Treatment						
	Proportion of children under five with fever in the past two weeks for whom advice or treatment was sought from a health provider	Number of children under five with fever in the past two weeks for whom advice or treatment was sought from a health provider	Total number of children under five with fever in the past two weeks	Population-based household survey	Geographical area, urban or rural, wealth index, educational status, sex	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Proportion of patients with confirmed malaria who received first-line antimalarial treatment according to national policy [9]	Number of patients with confirmed malaria who received first-line antimalarial treatment according to national policy	Total number of confirmed malaria case (includes cases found both passive and active surveillance)	Routine health information system, health facility survey or audit	Geographical area, type of facility, parasite species, age, sex	Will likely be more accurate from health facility survey or audit, because routine data will likely presume cases were treated according to national policy
	Proportion of patients with *P. vivax* or *P. ovale* infection who received radical cure treatment (primaquine or tafenoquine) [9]	Number of patients with a confirmed *P. vivax* or *P. ovale* infection who received radical cure treatment (primaquine or tafenoquine)	Number of patients with confirmed *P. vivax* or *P. ovale* infection	Routine health information system, health facility survey	Geographical area, type of facility, parasite species, age, sex	
	Proportion of confirmed *P. falciparum* cases who received single, low-dose primaquine	Number of confirmed *P. falciparum* cases who received single, low-dose primaquine	Total number of confirmed *P. falciparum* cases	Routine health information system, health facility survey	Geographical area, type of facility, age, sex	
	Proportion of severe malaria cases that were referred	Number of patients with severe malaria who were referred	Number of patients with severe malaria	Routine health information system, health facility survey	Geographical area, type of facility, age, sex	Denominator for this indicator may not be collected and the indicator may not be able to be calculated
	Proportion of referred patients with severe malaria that received pre-referral treatment	Number of referred patients with severe malaria that received pre-referral treatment	Number of patients with severe malaria that were referred	Routine health information system, health facility survey	Geographical area, type of facility, age, sex	
	Proportion of health facility months without stockouts of first-line treatments (includes treatment for severe anemia)	Number of health facility months without stockouts of first-line treatments	Number of health facility months	Routine health information system, health facility survey	Geographical area, type of facility	Disaggregated by type of treatment (malaria and anemia)

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Proportion of patients with confirmed malaria with adequate clinical and parasitological response	Number of patients with confirmed malaria with adequate clinical and parasitological response on day 28 (or 42)	Number of patients with confirmed malaria that were treated according to national policy and assessed on day 28 (or 42)	Therapeutic efficacy study	Geographical area/sentinel site, age, sex	
Surveillance						
	Proportion of malaria cases detected by surveillance systems	Number of confirmed malaria cases identified through active and passive surveillance over 1 year X 1,000	Estimated number of malaria cases over 1 year X 1,000	Routine health information system	Geographical area, age, sex	Estimated number of malaria cases (denominator) should include the proportion of patients who seek care, proportion who receive a diagnostic test, and proportion of health facility reports received
	Annual blood examination rate	Number of patients receiving a parasitological test during one year	Mid-year number of people at risk for malaria	Routine health information system	Geographical area, type of facility	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Proportion of expected health facility reports received	Number of reports received from health facilities during the reporting period	Number of reports expected from health facilities during the reporting period (number of health facilities multiplied by the number of reports expected per health facility during the reporting period)	Routine health information system	Geographical area, type of facility	
	Number and proportion of malaria epidemics detected within two weeks [23, 24]	Number of malaria epidemics detected within two weeks	Number of malaria epidemics detected	Routine health information system	Geographical area	Indicator should be measured within a one-year time frame
	Number and proportion of suspected malaria outbreaks investigated	Number of malaria suspected outbreaks investigated	Total number of suspected malaria outbreaks			
	Number and proportion of malaria outbreaks responded to in a timely manner	Number of malaria outbreaks responded to in a timely manner	Total number of malaria outbreaks			
	Proportion of inpatient deaths due to malaria (e.g., case fatality rate)	Number of inpatient deaths due to malaria	Total number of inpatient deaths	Routine health information system, health and demographic surveillance system (HDSS)/sentinel sites	Geographical area, age, sex	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
Impact						
	Malaria case incidence: number and rate per 1,000 people per year (disaggregate by species and active and passive case detection for low-transmission settings)	Number of confirmed malaria cases identified by active and passive surveillance during 1 year X 1,000	Mid-year number of people at risk for malaria infection during reporting year	Routine health information system, community health information system, HDSS/sentinel sites	Geographical area or focus, risk group, active versus passive case detection, age, sex, and species	May report number of cases when incidence is low
	Malaria test positivity rate	Number of confirmed malaria cases	Number of patients who received a parasitological test	Routine health information system, community health information system, HDSS/sentinel sites	Geographical area, age, sex, parasite species	
	Proportion of admissions for malaria	Number of inpatient admissions for malaria	Total number of inpatient admissions	Routine health information system, HDSS/sentinel sites	Geographical area, age, sex	
	Malaria mortality: number and rate per 100,000 people per year	Number of malaria-specific deaths reported in the previous year X 10,000	Mid-year number of people at risk for infection during the reporting year	Routine health information system, HDSS/sentinel sites, civil registration and vital statistics	Geographical area, age, sex, risk group and parasite species	May report number of deaths when mortality rate is low
	All-cause child mortality (Number of deaths among children ages 0–59 months per 1,000 live births)	Number of deaths among ages children 0–59 months per 1,000 live births	1,000 live births	Population-based household surveys, census, civil registration and vital statistics, HDSS/sentinel sites	Age	

	Indicator	Numerator	Denominator	Source	Disaggregation	Comments
	Annual number of malaria outbreaks reported	Number of malaria outbreaks reported in the previous year		Routine health information system, routine program reporting		
	Parasite prevalence: proportion of population with infection with malaria parasites	Number of people with malaria infection detected by RDT or microscopy	Total number of people tested for malaria parasites by RDT or microscopy	Population-based household survey, special study	Geographical area, urban or rural, wealth index, educational level, sex	
	Seroprevalence	Number of people who tested positive for antimalarial antibodies	Total number of people tested for antimalarial antibodies	Population-based household survey, special study	Geographical area, age, sex	

Source: Drawn from the World Health Organization Malaria Surveillance, Monitoring and Evaluation Reference Manual [9] and the World Health Organization Malaria Manual for Elimination Scenario Planning [10]

ANNEX 3. CASE STUDIES

Haiti case study

A3.1. Background and rationale

The Malaria Zero consortium aims to support Ministries of Health in their efforts to eliminate malaria from the island of Hispaniola by 2020. To this end, Malaria Zero is supporting the piloting of a package of malaria elimination interventions in five communes of the Grand'Anse Department in 2018 that have the highest confirmed malaria case incidence in Haiti. The strategy will consist of improving surveillance using the DHIS 2 system and improving access to diagnosis and case management, including the use of community case management to expand treatment access. Within 12 malaria transmission foci that are suspected of significantly contributing to malaria transmission in the target area, one round of targeted mass drug administration (tMDA) with sulfadoxine-pyrimethamine and indoor residual spraying (IRS) with Actellic-300CS will be conducted. Community engagement will also be conducted throughout the pilot area to bolster intervention acceptance and uptake. In conjunction with improved surveillance and access to case management, the goal of the tMDA and IRS is to interrupt malaria transmission in the suspected transmission sources in the pilot area and thereby interrupt transmission across the entire pilot area. If successful, the intervention package is planned to be expanded to all 12 communes of the Grand'Anse Department in 2019.

As the pilot elimination package will be implemented at full scale across the entire target area, even with tMDA and IRS targeted to the highest transmission areas, an experimental study design with a randomized contemporaneous control group will not be possible for evaluating the impact of the overall intervention package. Because contiguous communes in the rest of the Grand'Anse Department have much lower transmission overall, and are therefore not comparable epidemiologically to the pilot communes, assigning them to serve as a control group in a non-randomized quasi-experimental study design is also not a viable option. Moreover, repeated cross-sectional household surveys are not an ideal data collection method for this area because of the overall low transmission level that has very high spatial heterogeneity.

Evaluation approach

The impact of the package of elimination interventions will be measured together with an integrated epidemiological evaluation (IEE) that will incorporate all available data sources in the analyses, as outlined below. The goal of the IEE will be to provide proof of principle of effectiveness of the aggressive elimination strategy in the Grand'Anse pilot area to achieve interruption of transmission (zero cases). The IEE primary objectives are as follows:

- Assess and document implementation of Malaria Zero-supported elimination activities in the target area and generate data for iteratively improving and expanding elimination efforts
- Quantify the impact of Malaria Zero-supported elimination activities in reducing malaria prevalence and cases in the pilot area
- Document interruption of malaria transmission in the pilot area

Because of the context outlined above, the evaluation of the elimination pilot will primarily rely on a combination of observational study designs. At the core, the effectiveness of the intervention package will be assessed using a pre-post study design with trends in confirmed malaria case incidence as the primary outcome, meaning that the evaluation assumes that without the intervention package scale-up, no observed decrease in malaria transmission would have occurred in the target area. Because this design provides relatively weak evidence of causality between the exposure to the intervention package and observed outcome, the evaluation will also rely on a set of quasi-experimental study designs with constructed control groups to better assess causal inference between program exposure and outcomes. Planned analyses are as follows:

- **Interrupted time series (ITS) using trends in confirmed malaria case incidence:** This study design will be used to bolster causal inference between the intervention package scale-up and changes in confirmed malaria case incidence over time. The approach will involve constructing a time series of confirmed malaria case incidence in the launch area before and after the implementation of the intervention package scale-up. Changes in confirmed malaria case incidence rates before and after implementation will then be tested statistically using a segmented regression method. To further strengthen causal inference, trends in confirmed malaria case incidence in the remaining seven communes in the Grand'Anse Department will be included in the ITS analysis in a difference-in-differences approach. Analyses will adjust for treatment-seeking, DHIS 2 reporting completeness, diagnostic testing rates, and environmental characteristics over time, to the extent possible.
- **Repeated easy access group (EAG) surveys of schools and health facilities:** A baseline EAG survey was conducted during the peak malaria transmission season in November 2017 to ascertain rapid diagnostic test (RDT) malaria parasite infection prevalence and seroprevalences in the five commune pilot area. The EAG survey will be repeated the following year during the peak transmission season (after the scale-up of the elimination intervention package) to serve as a follow-up for measuring changes in these outcomes. Assessing changes in infection prevalence will be used to supplement the ITS approach described above, as well as quantify where transmission is still occurring. The use of a set of short-term serological markers to detect recent exposure to malaria parasite infections will be included in the follow-up EAG survey to estimate the level of malaria transmission still occurring in the target area after the intervention package scale-up.
- **Freedom from disease (FFD) analysis:** Being able to confirm the absence of transmission will be critical in documenting that malaria transmission has been interrupted in the pilot area and meeting the objectives of the IEE. Confirming the absence of transmission is difficult because the population will not be sampled in its entirety. The probability that transmission would be detected if it exists in the pilot area after the scale-up of the elimination intervention package can be estimated using an FFD analysis. Data from all available sources, including routine surveillance of confirmed malaria cases and all actively collected data on infections from the repeated EAG surveys, will be used for the FFD analysis. At the point when negative reporting begins (e.g., all individuals tested are confirmed to be free of malaria), the probability of having achieved freedom from malaria transmission will be quantified. This estimate will enable the identification of areas where transmission is and is not likely to be persisting in the pilot area based on the available evidence.

Data from the ITS and EAG evaluation studies will provide independent estimates of the impact of the elimination intervention package scale-up on confirmed malaria case incidence and current and past infection prevalence. Data from the ITS and repeated EAG surveys can also be integrated to identify where transmission foci are still occurring and where they have been interrupted. If the elimination intervention

package is shown to have had a substantial impact on reducing malaria outcomes to the point where transmission interruption in the pilot area is possible, the package will be expanded to the remaining seven communes in the Grand'Anse Department. When the health facilities in the pilot area start reporting zero confirmed malaria cases, and there are no parasite infections identified from EAG surveys, the FFD analysis will incorporate all available data to estimate the probability that interruption of malaria transmission has been achieved in the target areas. If the elimination intervention package is not shown to have had a meaningful impact on malaria outcomes to the point where interruption of transmission is possible in the pilot area, a re-examination of the pilot strategy will be undertaken and adjusted to improve its effectiveness, and the evaluation will continue.

Cambodia case study

A3.2. Background

There is good evidence that malaria transmission is decreasing in much of Cambodia. As transmission decreases, it becomes harder to both stratify and assess the impact of control on stratified areas based on classical infection measures, such as malaria prevalence detected through microscopy or RDT, because infection events become much rarer. Serological surveillance (the detection of antibodies specific to malaria proteins) provides additional information on transmission dynamics because it reflects historical exposure to infection. When integrated with age, serological data can estimate a seroconversion rate, which is analogous to the force of infection for malaria. Serological analysis of four consecutive Malaria Indicator Surveys (MIS) in Cambodia was conducted to examine changes in transmission patterns and the utility of antibody responses in identifying remaining areas of transmission as compared to alternative tools (microscopy and polymerase-chain reaction [PCR]).

A3.3. Evaluation study design

Samples collected in large countrywide MIS surveys in 2004, 2007, 2020, and 2013 were measured for antibody responses to both P. falciparum and P. vivax antigens. All four surveys followed a standardized methodology, and all data collection took place at the end of the rainy season, between October and November (peak malaria transmission time). All surveys were cross-sectional, two-stage cluster household surveys stratified by geographical domain and risk zone where the first stage was risk villages sampled using probability proportional to size and the second stage was households sampled using simple random sampling. Surveys were conducted during peak malaria season using standardized questionnaires. In each survey, 80–100 clusters were sampled, with about 90 individuals per cluster. Serological assays used the same antigens throughout; AMA-1 and MSP-1 for both P. falciparum and P. vivax. Surveys varied in spatial coverage, including the geographical domains and number of provinces sampled, and the definition of risk zone varied (<2km or <5km from forest). For the purposes of this trend analysis, provinces were re-grouped according to those included in every survey round (12 provinces). No PCR was done in 2007 and no vivax serology was done in 2010.

A3.4. Evaluation outcomes and indicators

Parasite prevalence by microscopy and PCR. Serological data as seroprevalence, magnitude of antibody response and seroconversion rate.

A3.5. Data sources

The MIS surveys were conducted by Malaria Consortium and the National Center for Parasitology, Entomology and Malaria Control, and laboratory work was performed by the Institute Pasteur in Cambodia and the London School of Hygiene and Tropical Medicine.

A3.6. Data synthesis and analysis

Cluster-level prevalence was estimated for parasitological and serological endpoints. All analysis was performed in Stata using the SVY command, accounting for the original survey design. Seroprevalence was calculated separately for each antigen and each survey using a two component fixed mixture model of antibody responses, using the mean plus 3SD of the narrower distribution component as the cut-off for seropositivity. Species seroprevalence was defined as positive to one or both antigens. Seroconversion curves were fitted to age-seroprevalence data using maximum-likelihood methods. Optimal model fits for two forces of infection were assessed using profile likelihood plots, whereby log likelihood was plotted for each model with one-year increments. Note that from 2007 all surveys included clusters in the 2–5 km from forest, which was not used in the 2004 survey. This allowed more detailed risk strata analyses of serological trends and comparison to PCR.

A3.7. Key findings

There are clear decreasing tends in parasite and seroprevalence over time (see Table 1). Overall, PCR estimates were higher than microscopy by approximately 50 percent, which is consistent with the literature. Serological estimates were approximately ten-fold higher, likely reflecting historical exposure and antibodies from previous infections.

Table A3.1. Infection and exposure prevalence for *P. falciparum* and *P. vivax* in each of the four MIS surveys

	Fever	*P. falciparum* microscopy positive	*P. vivax* microscopy positive	*P. falciparum* PCR positive	*P. vivax* PCR positive	*P. falciparum* serology positive	*P. vivax* serology positive
2004	15.3 (1,244/8,116)	2.2 (181/8,116)	1.0 (84/8,116)	ND	ND	30.3 (2,276/7,501)	21.2 (1,622/7,650)
2007	13.1 (1,055/8,067)	1.8 (146/8,067)	1.1 (92/8,067)	4.4 (339/7,707)	1.9 (146/7,707)	31.5 (2,437/7,736)	13.8 (1,084/7,840)
2010	8.7 (939/10,853)	0.3 (34/10,853)	0.4 (48/10,853)	1.0 (99/10,250)	1.7 (178/10,250)	11.5 (1,159/10,091)	ND
2013	12.5 (1,059/8,443)	0.04 (3/8,440)	0.1 (11/8,440)	0.7 (59/8,443)	0.5 (46/8,443)	8.6 (712/8,261)	8.3 (671/8,110)

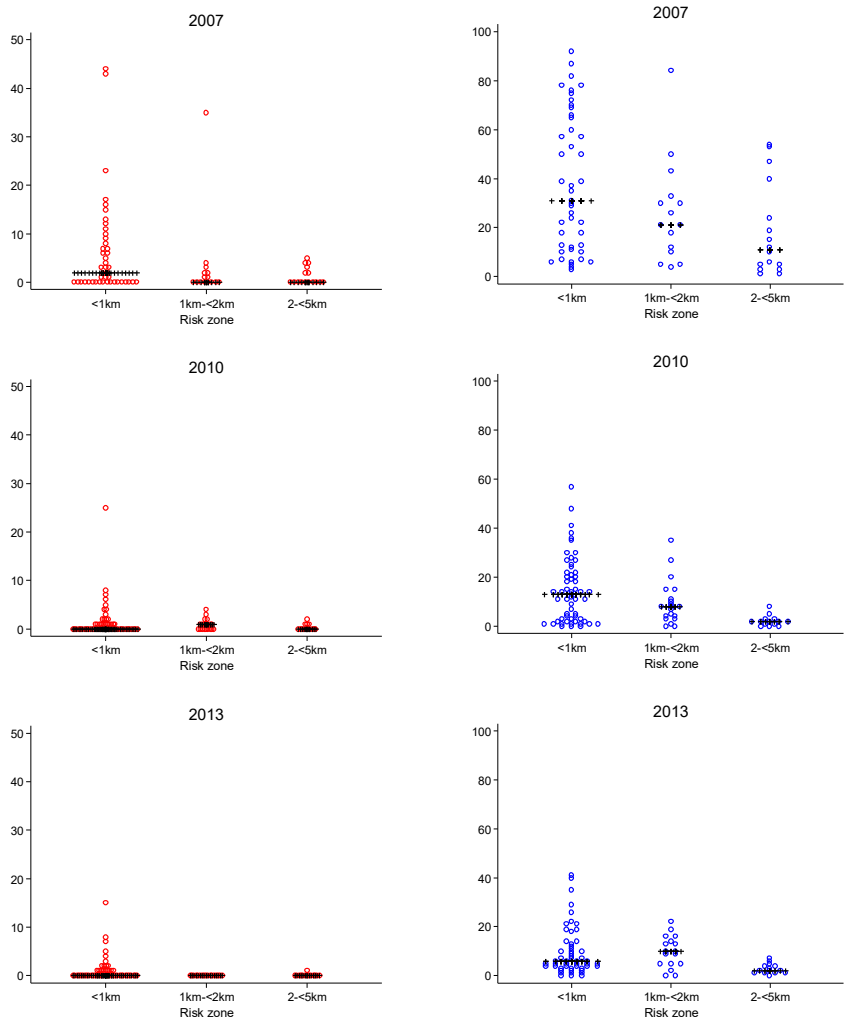
NOTE: The lack of difference between 2007 and 2004 is due to the difference in sampling area; the 2007 survey did not include the known very low prevalence area around Phnom Penh.

ND = not done

There was a decreasing prevalence of infection detected by PCR and seroprevalence over time with distance to forest (malaria risk zones), shown in Figure 1. Using serological markers improves the ability to examine

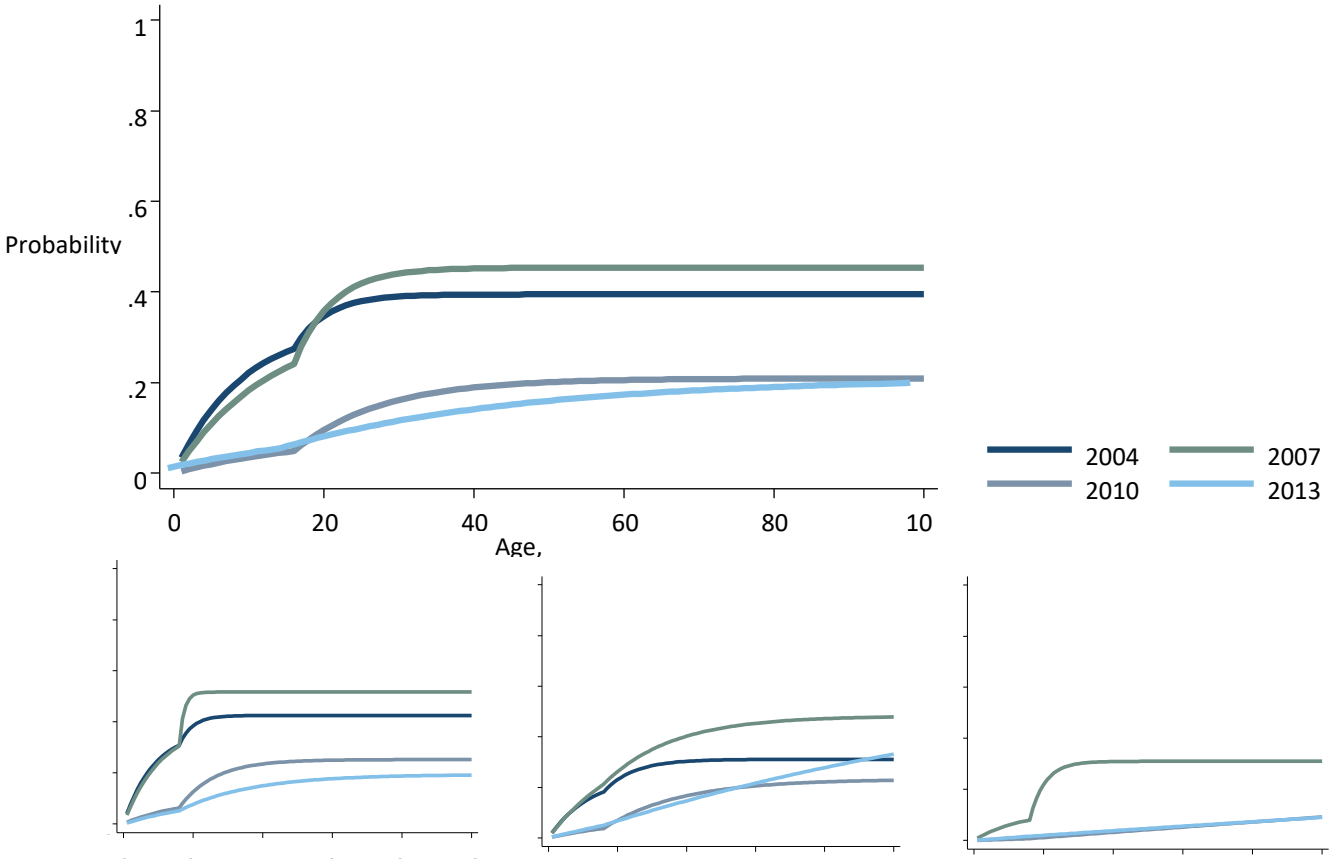
changes in transmission over time even when to very low levels (e.g., from 2010 to 2013) and also allows the potential to use the more granular metric for stratification, identifying areas of potential receptivity where no parasites can be detected using molecular methods.

Figure A3.1. Cluster level PCR prevalence (red) and seroprevalence (blue) by risk zone for *P. falciparum*



Age-seroprevalence curves shown in Figure 3 highlight that the models fits better with two forces of seroconversion with a change at approximately 14 years of age, consistent with increased exposure being associated with forest-going activities at this age. Data also show a consistent decrease in seroconversion rate in the oldest age group, from 0.156 in 2004 to 0.008 in 2013. Seroconversion rate is the number/percentage of individuals becoming seropositive each year, so in 2004 this was 15.6 percent and in 2013 it was 0.8 percent. The risk zone plots demonstrate both that transmission is higher closer to the forest and that transmission has reduced in all settings. Similar reductions were seen with P. vivax (0.071 in 2004 to 0.005 in 2013), and the effect was most pronounced in the highest risk zone, decreasing with distance from the forest. The effect was manifest both as seroprevalence and antibody level (data not shown).

Figure A3.2. Age seroprevalence for *P. falciparum* for each survey overall (a) and by each risk zone as distance to forest, (b) <1km, (c) 1–2km, and (d) 2–5km



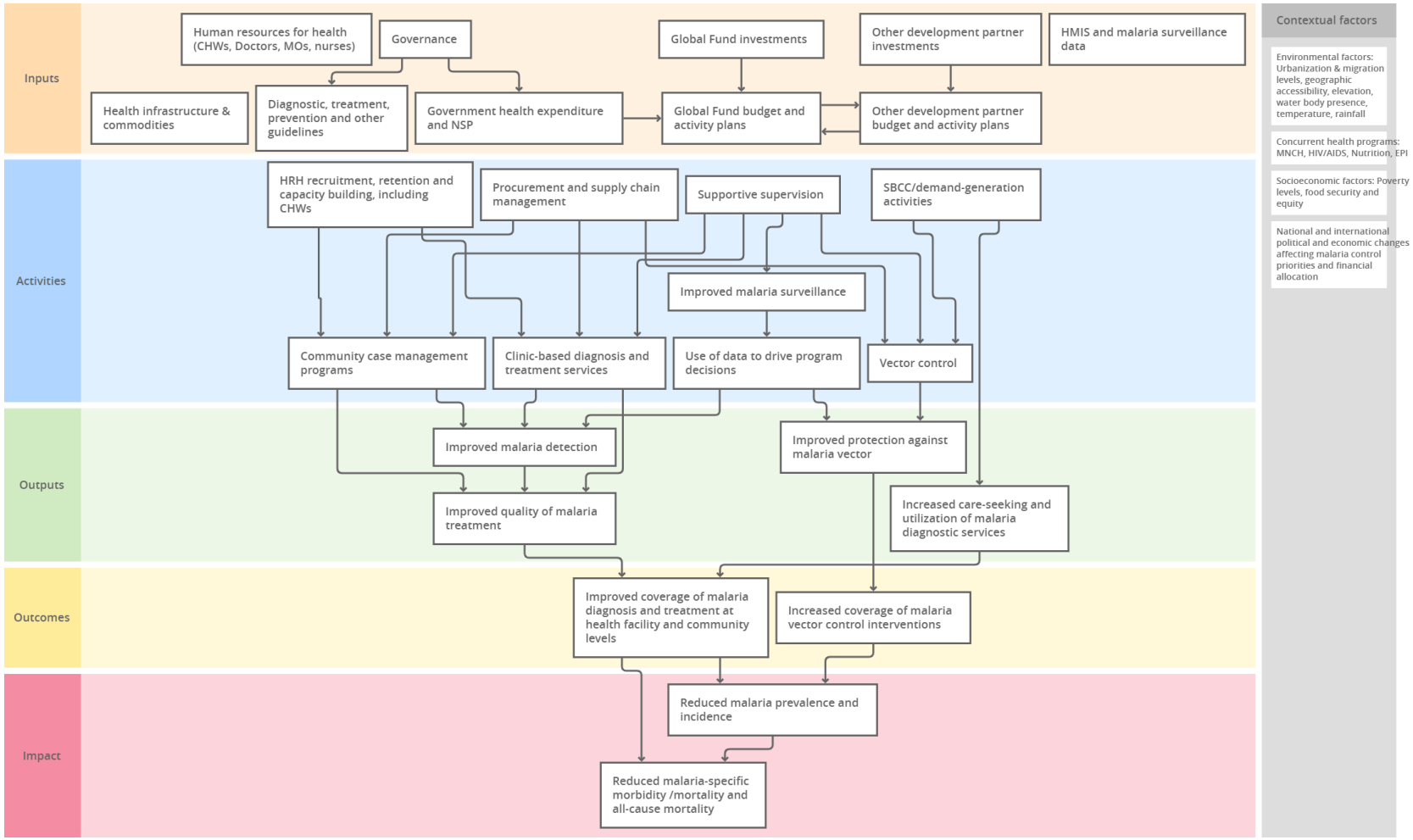
A3.8. Conclusions

The trend analyses presented pooled more than 33,000 samples over 4 consecutive large-scale surveys. High heterogeneity of transmission in Cambodia was described by both parasite prevalence and seroprevalence with more granularity evident with serology. Transmission associated with risk zone and higher seropositivity in older ages suggested differential risks for different age groups. Significant reductions in both P. falciparum and P. vivax over 10 years can be measured with serology even when malaria prevalence is very low (below 1 percent) and where other molecular methods (e.g., PCR) are not able to identify further statistically significant reductions. The use of hybrid approaches, including surveillance data, serology markers, and smaller-scale (or targeted) surveys, among others, could represent a suitable monitoring and evaluation package to assess the impact of interventions in very low transmission settings.

ANNEX 4. EXAMPLE OF AN IMPACT MODEL

The following is an example of an impact model was developed for the Global Fund Prospective Country Evaluations.[12]

Figure A4.1. Example of an impact model



[12] The model and its sources are available at https://evaluationplanningtool.org/model/mojdbjafakbgehabhdc.

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, NC 27516 USA
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This research publication has been supported by the President's Malaria Initiative (PMI) through the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AIDOAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of PMI, USAID, or the United States government. TR-19-334

ISBN: 978-1-64232-129-6

USAID FROM THE AMERICAN PEOPLE
CDC
U.S. President's Malaria Initiative
RBM Partnership To End Malaria
MEASURE Evaluation