

Guidance note: Version 3.4

Epidemiological and Impact Analysis

31 December 2014
Geneva, Switzerland

Contents

Purpose and objective	3
Scope and scale	4
Timing and process for epidemiological and impact analysis	5
Partners	5
Resources for epidemiological and impact analyses	5
Abridged guidance	6
A. Planning and Preparation Phase	7
B. Analysis and synthesis phase.....	8
C. Interpretation of findings phase.....	8
D. Presentation and use of findings phase	9
Annex 1: Global Fund Evaluation Approach and Definitions	10
Annex 2: Adequacy, Plausibility, and Probability	13
Adequacy and Probability	13
Plausibility.....	15
Annex 3: Disease-specific generic Terms of Reference for epidemiological and impact analysis	18

Purpose and objective

1. Program Reviews are important management tools. They provide program managers and other stakeholders with an opportunity to assess program performance, impact and lessons learnt during a specific time period. Program Reviews inform improvement and subsequent strategy cycles and should also be linked to Health Sector Reviews. A substantial amount of planning and resources go into a Health Sector or Disease Program Review.
2. As part of Program Review, it is important to periodically analyze epidemiological trends by place, person and over time and assess the causal pathways between investments and impact. In doing so, it is also important to assess other possible explanations for impact aside from those strictly related to programs. Such an assessment, called Epidemiological and Impact Analysis (“epidemiological and impact analysis”), can significantly strengthen a program and inform prioritization and implementation decisions. It also supports the Global Fund’s focus on investing for impact by means of the New Funding Model. In the New Funding Model, countries able to demonstrate impact will receive incentives at the time of funding allocation.
3. The purpose of the epidemiological and impact analysis is to provide a more robust assessment of whether impact and outcomes than would be obtained from an exclusive focus on national program results. Specifically, an epidemiological and impact analysis seeks to establish whether changes in impact are either plausibly resulting from national program input and activities, or are likely due to some other cause. It is about further assessing contribution and causation along the results chain, i.e. answering the question “Have the interventions and other competing explanations or hypotheses contributed to and resulted in these impacts, whether positive or negative?” (See Annex 1: Global Fund Evaluation Approach and Definitions.) This evidence will form a foundation for important funding recommendations and decisions.
4. The epidemiological and impact analysis is not a stand-alone process, but should be integrated into either program or health sector reviews. This is accomplished by capitalizing on existing review processes to conduct a plausibility assessment of impact: to examine whether disease-specific programs are likely having the intended impact, after taking other factors into account. See Annex 2: Adequacy, Plausibility, and Probability, for a more detailed explanation of “plausibility” in the context of an epidemiological and impact analysis.
5. Key to this analysis and to program reviews generally is availability of good quality and disaggregated impact data, which most countries still do not have. In practical terms, then, the data needed on comparison groups for a plausibility assessment will rarely permit more than a qualitative judgment to be made about the credibility of alternative explanations for impact. However, in situations where data do allow it, the quantification of program and non-program impact could be pursued under an epidemiological and impact analysis. The Evaluation of the Impact of Malaria Interventions on Mortality in Children in Mainland Tanzania is an example of one such analysis.*
6. An epidemiological and impact analysis should state not only what was achieved, but also show clearly show causal mechanisms (e.g. the impact of ARV treatment on adult mortality). An epidemiological and impact analysis focuses on how and why change has occurred due to

* http://www.pmi.gov/docs/default-source/default-document-library/tools-curricula/tanzania_ie_report.pdf?sfvrsn=4

program components and mechanisms in relation to potentially confounding factors. Equipped with such insights into impact, a review will become more strategic and informed about program decisions and funding allocation.

7. An epidemiological and impact analysis will not, however, be able to (i) state categorically why those changes have occurred; (ii) attribute change exclusively to specific causal factors, donors, or program components; or (iii) quantify the contribution that various causal factors have had on observed impact.
8. In the context of the Global Fund's New Funding Model (NFM) developing funding applications, known as "Concept Notes," should incorporate knowledge from epidemiological and impact analysis about the likely impact of programs.

Scope and scale

1. It is important to emphasize that an epidemiological and impact analysis is a component of a review that complements other the aspects disease-specific or joint Program Review conducted using updated WHO partner guidelines. It is therefore not a substitute for a comprehensive program review. Health Sector Reviews present an excellent opportunity to implement an epidemiological and impact analysis, considering various health interventions and disease trends all together. The new WHO Program Review guidelines allow up to three months or so from planning stages through report finalization. This timeline should provide adequate time to address the epidemiological and impact analysis component as part of the process.
2. A Program Review conducted using updated the WHO guidelines alone would constitute an adequacy assessment (see Annex 2) of whether observed trends in disease burden can be plausibly related to programmatic efforts (e.g., service delivery and coverage) and behavioral trends where relevant. The guidelines call for: a period of assembling data; an attempt to assess data quality and availability; and a recommended set of investments needed to improve measurement of trends in each of the three diseases. Following the Program Review, a joint review of available financing information and trends in epidemiological data (disease incidence, prevalence, mortality) nationally and, where feasible, for subnational areas and subpopulations. (See Annex 3 for a summary of disease-specific guidance on implementing epidemiological and impact analysis.)
3. In sum, epidemiological and impact analysis links disease trends to program efforts and to non-program factors. It is focused on strengthening the explanations for change or trends in disease burden that appear most valid.

Timing and process for epidemiological and impact analysis

1. In the NFM, an epidemiological and impact analysis is an input into the Concept Note. It should precede Concept Note development. The epidemiological and impact analysis should coincide with an AIDS, TB, or malaria program review, or a National Health Sector review, which in turn, should be aligned to national cycles. An epidemiological and impact analysis should be considered for use in all HIV, TB, malaria and joint program reviews. Depending on the availability of data at the national level, the epidemiological and impact analysis can also precede the program review. Sequencing the epidemiological and impact analysis in this way with other phases of a review may be particularly useful if capacity of national resource persons is limited.
2. In some instances, especially during the transition to the NFM and the first year of its full rollout, it may be difficult to align the timing of an epidemiological and impact analysis with the existing national cycle. In this case, the national program and major stakeholders may choose to conduct an off-cycle epidemiological and impact analysis to help update the national strategic plan, develop a sound investment case, and inform the Country Dialogue that accompanies the development and finalization of the Concept Note. If epidemiological and impact analysis work occurs at another time point, however, it could be used to feed into a subsequent, larger program review, to follow up on the findings of a previous review, or for updating an existing NSP.
3. Country Teams, disease advisors, country team members and the Technical Review Panel (TRP) will thoroughly review the epidemiological analysis in the Concept Note to make evidence-informed recommendations to the Grant Approval Committee (GAC).

Partners

1. As with other aspects of Program Review, the respective national disease program should lead and coordinate the epidemiological and impact analysis. Typical partners in a Program Review include the epidemiologists, evaluation experts and data analysts of a disease program, the data manager of the country's health information system, technical partners with analytical and epidemiological capacities, such as WHO and UNAIDS, and major donors in country, e.g. World Bank, PEPFAR and PMI. Other partners, e.g. other bilateral programs, local universities and technical NGOs, should also be engaged, in particular in consultation meetings at the start and end of the process.
2. Developing local analytical capacities should also be a longer-term aim. As alternative causal pathways are developed as part of the epidemiological and impact analysis (see below), care should be taken to include technical stakeholders who may have insight into these pathways as well. Sub-national level health managers may be involved where feasible and important.

Resources for epidemiological and impact analyses

1. Financial resources for the epidemiological and impact analysis component of a program review should normally be planned for as part of a single, comprehensive Program Review activity. If required, additional funds for the epidemiological and impact analysis may be sought from

alternative sources or reallocated from existing Global Fund grants in a way that does not significantly affect service delivery.

2. Ideally an external consultant epidemiologist with no potential conflict of interest will participate in the epidemiological and impact analysis and be responsible for it (see Annex 3: Disease-Specific Terms of Reference). When necessary, a member of the Global Fund's Secretariat with the suitable skill set will participate. These experts will be focused on the development of the plausibility case. Modeling and modeling experts are not necessary to conducting an epidemiological and impact analysis.

Abridged guidance

The following steps are standard elements included in the epidemiological and impact analysis. They should be carried out in conjunction with the relevant or corresponding stage of the Program Review. At a high level, the Program Review process is comprised of four steps or phases:

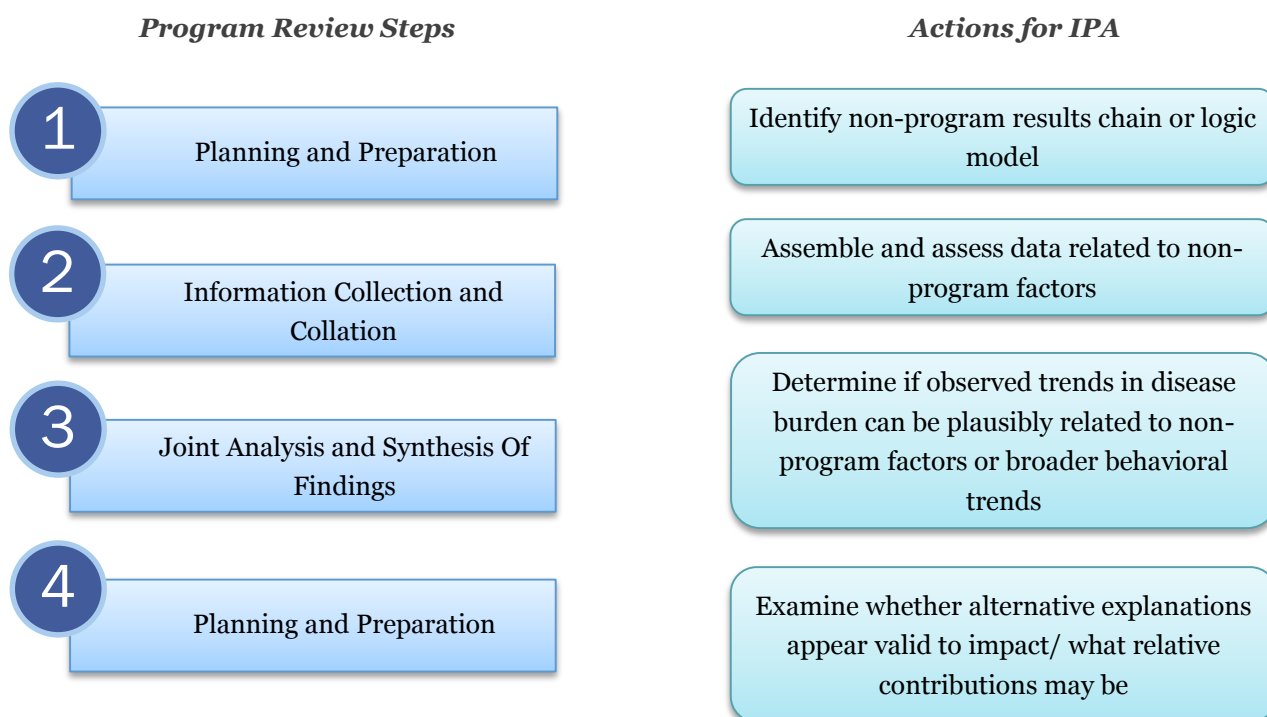
- Planning and preparation
- Information collection and collation
- Analysis, synthesis and interpretation
- Reporting

At each of these phases, the following supplemental activities are needed to develop the epidemiological and impact analysis. These are to be carried out by the expert engaged. They are:

- Selection of one or more alternative hypotheses
- Acquisition, assessment and mapping of data
- Analysis, synthesis and interpretation
- Incorporation of the findings and interpretation of the alternative hypothesis into the epidemiological and impact analysis report

The Figure below summarizes the Program Review steps and the additional epidemiological and impact analysis related activities.

Relationship between Steps in National Program Reviews and IPA



The epidemiological and impact analysis component activities, by Program Review phase, are broadly as follows:

A. Planning and Preparation Phase

1. **Hypothesis Formulation:** Select and explicitly state one or more reasonable and epidemiologically sound alternative hypotheses to be tested. For each hypothesis developed, consider (i) whether data are likely to be available for a given country, and (ii) whether the hypotheses explain trends in disease burden due to factors not related to disease-specific funding and associated interventions. For example:

- “Malaria declined because of significant climatological trends affecting mosquito populations and risk of transmission”
- “Child mortality has been rapidly declining due to secular trends and significant gains in other, frequently fatal childhood diseases”
- “TB is in decline due to significant improvements in HIV treatment and prevention.”

If more than one hypothesis is to be assessed, care should be taken to keep the epidemiological and impact analysis feasible and within reasonable scope. Develop a simple logic model or causal pathway diagram for each hypothesis, following the results chain.

2. **Data mapping:** Indicate likely or preferred sources of data. Determine availability and sources of data across the results chain that are informative of each alternative hypothesis, identifying any gaps. The identification of weaknesses and gaps can be an input into plans to strengthen the collection and use of necessary data.
3. **Data Acquisition:** Assemble and assess the relevant data that are available and that supplement information being collected and collated for the other components of the Program Review (e.g. rainfall data, entomological/vector studies). Pay close attention to any similar analyses that have been done before or are ongoing to look at how these can feed into the epidemiological and impact analysis. Hotspot mapping, and the World Bank multi-country study on HIV incidence decline are examples of work that involves gathering much of the needed data. The epidemiological and impact analysis should harness such resources and update as needed.
4. **Quality check and validation:** Assess the degree to which available data provide complete, reliable and consistent information on the alternative causal pathways.

B. Analysis and synthesis phase

1. **Analysis:** In parallel with other Program Review activities, conduct an analysis of the data that may support the alternative explanation. This should be done for the national level, and where relevant and feasible for: 1) person (e.g., by age, sex, wealth quintiles, or vulnerability and key populations); 2) and place (subnational geographic and ecological zones); 3) time (years or seasonal variations). Disaggregation of data sets is important where this can be achieved, even if it increases the preparation time before analysis can start.
2. **Synthesis:** Assemble the findings of the analysis in a narrative or case related to each alternative hypothesis and its causal pathway or logic model. The analysis and synthesis should put more weight on data generated with high quality and provide commentary on potential sources of bias or error in the source data.

C. Interpretation of findings phase

1. **Plausibility argument:** Use the results of the analyses and synthesis phase to assess the extent to which non-programmatic factors may have had an impact on trends in disease burden. This is a critical stage and requires adequate time and focus.
2. As a group, the entire Program Review team together with the expert engaged for the epidemiological and impact analysis should compare analyses of programmatic and non-programmatic impact. Normally this would entail a qualitative expert judgment of (i) where, on a fixed scale ranging from “no-effect” to “primary causal factor”, the preponderance of evidence points for the program-related results chain as the likely source of change in disease burden; and (ii) where, using the same scale, the evidence falls in support of the alternative explanation as either causing or contributing to any observed changes in disease-specific burden indicators.

3. It is about excluding hypotheses that cannot explain the trends. In some cases more rigorous, quantified plausibility arguments can be constructed if necessary data and statistical requirements can be met (see Annex 2⁺).

Examples

- The evidence suggests that there were significant reductions in annual rainfall and that populations of anopheles mosquitoes declined significantly over the last three to five years across the country. Additionally, evidence suggests that the NMP has stalled. Therefore it is possible that reductions in malaria transmission, prevalence, and the observed declines in parasite prevalence and anemia among children under five are due in significant part to changes in rainfall and climate during the review period. This could be further assessed by disaggregating outcomes by region.
- The HIV epidemic in the country appears stable, and there are no discernable trends in terms of either prevalence or the increase in the numbers of HIV+ individuals on treatment. Therefore it is unlikely that HIV is affecting the trends in the TB disease burden indicators.

D. Presentation and use of findings phase

1. Use the agreed results chain to assess contribution and causation – i.e. have the interventions and other competing explanations or hypotheses contributed to and resulted in these impacts, positive or negative? (Note that the TERG has stressed the importance of assessing both positive and negative impacts and outcomes.)
2. Compile all epidemiological and impact analysis materials into a summary of where the country stands in terms of disease burden and trends; the coverage and effectiveness of programmatic interventions and their impact on disease burden; possible alternative explanations for epidemiologic conditions; and recommendations for actions needed in the future.
3. Provide 3-4 key recommendations with supporting text that should inform relevant policies, e.g. the Concept note and program strategy. For example, identify whether nonprogrammatic conditions or factors plausibly account for a significant share of change in disease burden and/or are changing the nature of the epidemic. Should the epidemiological and impact analysis reach the conclusion that a plausible, evidence-based case can be made for an impact pathway other than that specifically related to program investments and results, the implications of the findings should be addressed specifically in the Concept Note.

⁺ Also see JP Habicht, CG Victora, JP Vaughn, "Evaluation designs for adequacy, plausibility, and probability of public health programme performance and impact." *Int J Epi* 1999;28:10-18

Annex 1: Global Fund Evaluation Approach and Definitions

The TERG reviewed the “country led platform for health and accountability” with the agreed results chain shown below to evaluate health progress and performance. They also stressed evaluation needs to be practical, build on country reviews and input to grant decisions.

The TERG agreed that impact evaluations would:

1. **Focus on the impact and outcome** questions:
 - a) Has there been a change in disease mortality/morbidity and/or incidence and prevalence, positive or negative?
 - b) Has there been a change in outcomes and behaviours, positive or negative?
2. **Use the agreed results chain to assess contribution and causation**, i.e. have the **interventions** and other **competing explanations** or hypotheses contributed to and resulted in these impacts, positive or negative?
3. The TERG stressed the importance of assessing **positive and negative impacts and outcomes**. Following this, contributing factors related to interventions, but also other hypotheses and explanations should be assessed.
4. The TERG stressed the need for **investment in rigorous analysis**, disaggregation of data by time, person and place including comparison groups where feasible, and using mixed methods approaches. Analysis should assess explicitly competing explanations and sources of bias.
5. The TERG agreed on its position on the importance of **contribution and assessing causation and competing explanations** rather than narrow attribution to just one source of financing or single intervention:

“The impact evaluation sets out to assess overall impact on the burden of cases and deaths due to the three diseases. The evaluation will describe the contribution of the Global Fund without direct attribution to any individual agency or effort”

The definition of impact evaluation combines **two components**, assessing final **disease outcomes and impact**, and assessing **contribution and causation** along the results chain:

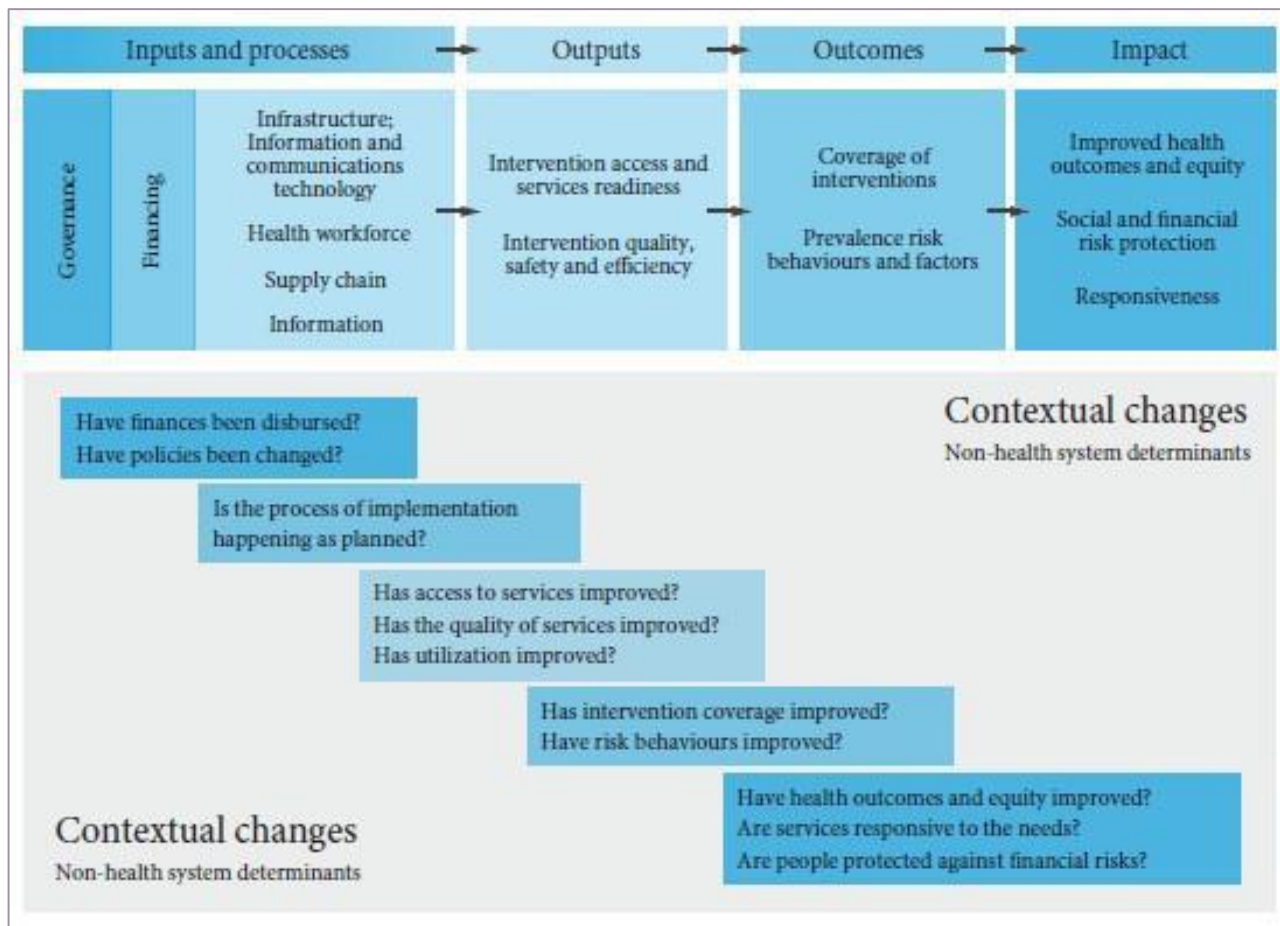
“Impact evaluation assesses the overall impact on the burden of cases and deaths due to the three diseases. It will assess causation and the contribution of the Global Fund and other explanations along the results chain from inputs to outcomes”.

This definition draws on definitions from the World Bank, the OECD and other sources, while applying it to the context of the three diseases and the Global Fund. These evaluations can and should build on **program reviews** which include impact and outcome data, and are supplemented by investments in analysis and assessment of competing explanations.

The TERG distinguished impact evaluations from **thematic reviews**, on which it will commission studies in order to provide analysis for strategic options. They were defined as:

“Reviews of specific issues or themes which assess past evidence and selected cases, with a primary focus on providing forward looking strategic options”

The evaluation framework



The following primary evaluation questions will be assessed:

1. Has there been a change in disease mortality/morbidity and/or incidence and prevalence, positive or negative?
2. Has there been a change in outcomes and behaviours, positive or negative?

In addition the following questions will be included, to assess contribution and causation along the results chain, wherever feasible:

1. Has there been an increase in **coverage** of key intervention services, and has these reached groups at risk?
2. Has **access by age, sex, equity and quality of key intervention services** improved?
3. Have **finances** been made available for key services and contributors?
4. Were there sufficient **quality data** to detect the effect of increase in service coverage and quality on disease burden? What were sources of bias?
5. What was **contribution of various sources of funding** in scale up of resources, increase of coverage of key intervention services, improvement of service quality and outcome? What were the other competing explanations and hypotheses of changes in outcomes and impacts, positive and negative?
6. How can contributions of the Global Fund be improved to better contribute to outcomes and impact? What are the management recommendations?

The process will involve:

1. Baseline analysis of secondary data and data systems;
2. In-country review and analysis building on program reviews;
3. Production of evaluation report, analysis, with TERG and country review.

Annex 2: Adequacy, Plausibility, and Probability

1. This Annex provides an explanation and operational definition of “plausibility assessment” for Global Fund evaluations, including for Epidemiological and impact analysis for Impact Analysis (epidemiological and impact analysis).
2. In the context of health program evaluation, the concepts of adequacy, plausibility, and probability are often used to describe evaluation designs that represent increasing levels of rigor and explanatory power.[#] In order to establish a common understanding of what a plausibility assessment is, it is necessary to (i) review what the terms ‘adequacy,’ ‘plausibility,’ and ‘probability’ mean in the context of public health program evaluation; and (ii) to develop a working or operational definition of a plausibility assessment. (Annex 1, “The Global Fund Evaluation Approach and Definitions” provides the rationale or basis for adopting contribution and plausibility as the standard of evidence sought in Global Fund evaluations and in the epidemiological and impact analysis.)

Adequacy and Probability

3. An adequacy assessment or evaluation answers the question: did the expected changes occur across the results chain – including changes in impact? Ideally the ‘expected’ change will be measured against specific targets. Even where no targets have been set, however, adequacy may be assessed in terms of time trends in impact indicators, such as reductions in total or age-specific mortality, anemia, or HIV or TB prevalence.
4. Adequacy evaluations are limited to describing whether or not change has occurred. Changes in the provision or utilization of a specific service may reasonably be ascribed to a program merely by looking at these changes. However, it is difficult to infer that a given program is responsible for causing changes in outcomes and impact by looking at trends alone – even across the results chain. In order to make a stronger inference about causes, it would be necessary to have a comparison group. Comparison groups are used to provide evidence about whether impact might have occurred anyway – even in the absence of the program. They permit the investigator to assess whether some alternative factor not related to the program has caused or contributed substantially to any observed impact. In an adequacy evaluation, no comparison groups are used. The following table from Habicht et al summarizes characteristics of an adequacy evaluation:

[#] Habicht et al coined these terms for alternative evaluation designs for public health programs in their seminal paper on the topic. This annex draws heavily on their work. (JP Habicht, CG Victora, JP Vaughn, “Evaluation designs for adequacy, plausibility, and probability of public health programme performance and impact.” *Int J Epi* 1999;**28**:10-18)

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Adequacy			Predefined adequacy criteria	Objectives met
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients		Activities being performed as planned in the initial implementation schedule
Cross-sectional	Once		Absolute value	
Longitudinal	Change		Absolute and incremental value	
Impact	Health and behavioural indicators	Programme recipients or target population		Observed change in health or behaviour is of expected direction and magnitude
Cross-sectional	Once		Absolute value	
Longitudinal	Change		Absolute and incremental value	

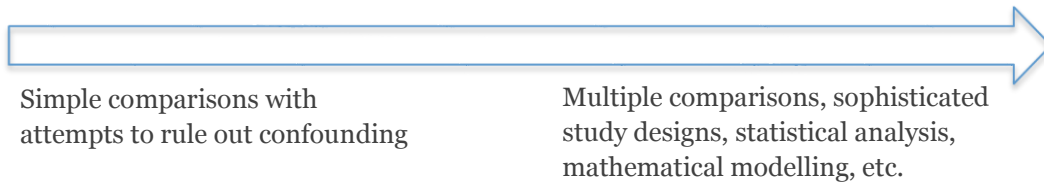
5. The partner guidelines for Program Reviews are largely geared to produce adequacy assessments. While they are comparatively inexpensive and may suffice for reviews of program outputs, Habicht et al point out that adequacy accounts provide less convincing evidence for interpreting how program results relate to outcomes and impact than do plausibility and probability designs.
6. Probability evaluation designs are at the other end of the spectrum of rigor and explanatory power. A probability assessment is often equated with attribution, and requires an experimental or quasi-experimental trial design that answers whether the program had a statistically attributable effect. Probability evaluations are able to confirm that there is only a small, known probability (for example, $P < 0.05$) that the difference between program and control areas were due to confounding (i.e. factors related to impact but not related to the program), bias or chance.
7. Probability designs are sometimes referred to as a gold standard in efficacy research. They require stringent methods such as random assignment to program and control groups, and require sample power considerations to drive the scale of the effort. They are the most expensive kinds of evaluation to undertake. The following table from Habicht et al. summarizes characteristics of a probability evaluation:

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Probability			Randomized control group(s)	The programme has an effect ($P < 0.05$)
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients		Intervention group has better performance than control
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Impact	Health and behavioural indicators	Programme recipients		Changes in health or behaviour are more beneficial in intervention than control group
Longitudinal-control	Relative change	Target population	Comparing before-after between intervention and control	

8. In the context of a conducting a retrospective Program Review, probability assessments of impact will almost never be undertaken. If, however, the results of previous studies of this kind are available for review, they would serve as a superior part of the evidence base for establishing an account of program performance and effectiveness.

Plausibility

9. For the purposes of epidemiological and impact analysis, a plausibility assessment is defined as one that answers the question: did the program seem to have an effect above and beyond other, external factors? In this context, “plausible” means that there is (i) an epidemiologically reasonable hypothesis for impact; (ii) that the evidence assembled is believable, is worthy of acceptance; and (iii) credibly establishes whether external, confounding factors are related to any impact observed. Plausibility is often equated to contribution.
10. TERG has established that a plausibility assessment provides the right standard of credible evidence for the purposes of an epidemiological and impact analysis. This option balances considerations of rigor; cost; and the practicalities of conducting program reviews in situations where epidemiological and impact data are rarely sufficient to do a more extensive evaluation.
11. Plausibility assessments go beyond adequacy assessments by providing an analysis that attempts to rule out external factors that might have caused or contributed to the observed effect. Conducting a plausibility assessment is accomplished by selecting a control or comparison group either before or after the start of an evaluation. When conducted in the context of an epidemiological and impact analysis, the selection of controls will almost always be retrospective.
12. The selection of the control or comparison group is often done opportunistically, based on what each situation will allow. There are a variety of methods that fall under the category of plausibility assessment. They can range from historical comparisons within one group, to case-control studies. The former have comparatively lower level of explanatory power and more modest data requirements, and the latter have comparatively high explanatory power and more extensive data and analytical requirements.
13. As the figure below indicates, plausibility statements encompass a continuum, from a weak to a strong statement. Both the context and requirements of the investigators for certain standards of credible evidence would dictate what can be attempted in a given situation. In most cases, the retrospective nature of the exercise and common data limitations will rarely permit more than a qualitative judgment to be made about the credibility of alternative explanations for impact. However, in situations where data do allow it, the quantification of program and non-program impact could be pursued under an epidemiological and impact analysis.
14. The epidemiological and impact analysis guidelines and accompanying disease-specific TORs are written to permit the most commonly feasible and simplest way to approach a plausibility assessment. This approach uses historical controls or comparison groups in a manner that tries to account for factors external to the program and will involve a qualitative expert judgment to be made about the plausible level of contribution of programmatic and non-programmatic factors to impact. In most cases data should be available to permit this kind of plausibility assessment based on simple comparisons and attempts to rule out confounding.



15. Stronger types of plausibility assessments can be used where data are more widely available. For example, control groups can come from within a programmatic coverage area. Such control groups can be used if (i) there have been differing degrees of program intensity in the program area, with differing effects on the units of analysis; (ii) outcome and impact measures are available at the unit of analysis level; and (iii) measures of external/confounding factors are also available at this level. Comparisons can also be made to areas without a program. These comparisons can be either at one time point (using the most recent epidemiological impact data available) or longitudinal (using data from early and late in the review period), though measurement of potentially available confounding factors at all time points.
16. The most rigorous form of a plausibility design is a case-control study. A case-control study can compare previous exposure to the program among those with and without disease or, conversely, to compare disease status among those with and without exposure to the program. In many cases, the data requirements for retrospective case-control studies may exceed the data availability and quality. In addition to the measurement requirements, confounders must be treated using the correct statistical procedures for matching, standardization, and stratification, or multivariate analytical techniques. While beyond the scope of a routine epidemiological and impact analysis, if the necessary data and expertise exist and data are correctly analyzed, case control studies can offer quite definitive results.
17. Nevertheless, even in the most rigorous plausibility assessment one cannot completely ‘rule out’ that the observed trends were due to alternative explanations. To reach this standard of causal inference, a probability evaluation design would be needed. This being said, less stringent plausible statements are usually sufficient in a programmatic context. The following table from Habicht et al. summarizes the characteristics of a plausibility evaluation:

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Plausibility			‘Opportunistic’ or non-randomized control group	The programme appears to have an effect above and beyond the impact of non-programme influences
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients (dichotomous or dose-response)		Intervention group appears to have better performance than control
Cross-sectional	Once		Control group	
Longitudinal	Change		Before-after	
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Impact	Health and behavioural indicator	Programme recipients or target population (dichotomous or dose-response)		Changes in health or behaviour appear to be more beneficial in intervention than control group
Cross-sectional	Once		Control group	
Longitudinal	Change		Before-after	
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Case-control	Once	Target population	Comparing exposure to programme in diseased (cases) and non-diseased (controls)	

Examples of Plausibility Assessments

Hypothesis: Over the past four years and accounting for other factors, TB cases have declined due to increased coverage of symptomatic cases and a reduction of time between diagnosis and starting treatment.

Design: Historical control

Data requirements: Trends in case detection and potential non-TB program factors such as HIV prevalence and numbers of PLWHA on treatment.

Hypothesis: In locations with high risk of malaria transmission, anaemia in children under five will be lower in areas where bed net use is at or above program targets, compared to areas where targets are not being reached.

Design: Internal comparison group

Data requirements: Measures of anaemia in under-fives; consistent and comparable measures of: bed net use and potential confounders such as the presence of other programs in each area (e.g. indoor spraying).

Hypothesis: HIV prevalence in key populations will be lower in areas receiving targeted prevention interventions compared to similar populations not receiving these services.

Design: External comparison group or case-control method

Data requirements: Population denominators/estimates, information about sexual and other risk-related behaviour (e.g. injecting drug use) for each program and comparison group.

Annex 3: Disease-specific generic Terms of Reference for epidemiological and impact analysis

[Disease-specific ToRs begin on the next page]

National health sector and national HIV
program reviews:

Terms of reference for Epidemiological and Impact Analysis for Global Fund concept notes

1. BACKGROUND

Program Reviews are important management tools that provide managers and other stakeholders with an opportunity to assess program performance, impact and lessons learnt during a specific time period. Program Reviews also inform program improvement and the subsequent strategy cycles, and should be linked to health sector reviews. A substantial amount of planning and resources go into a national disease control Program Review.

As part of Program Review, it is also important to assess periodically the causal pathways between investments and impact, and assess other possible explanations for impact aside from those related to programs. Such analysis, called an “Epidemiological and impact analysis”, can significantly strengthen a program and its management as well as an application for funding from the Global Fund and other donors by focusing on investing for impact.

The purpose of an epidemiological and impact analysis is to provide a more robust assessment of whether impact and outcomes are plausibly resulting from programmatic input and activities, or might be due to other factors. It is about further assessing contribution and causation along the results chain, i.e. answering the question “Have the interventions and other competing explanations or hypotheses contributed to and resulted in these impacts, whether positive or negative?” This evidence will form a key basis for important funding recommendations and decisions.

The services of a consultant epidemiologist, or an evaluator with epidemiological analysis expertise, are required to carry out the tasks and provide deliverables for the epidemiological and impact analysis.

This consultancy, associated tasks, and deliverables should be considered part of a larger HIV AIDS Program Review effort (See *WHO Guide to Conducting Programme Reviews for the Health Sector Response to HIV/AIDS*; see Guidance Note for Epidemiological and impact analysis for Impact Analysis (epidemiological and impact analysis) for the New Funding Model)

2. OBJECTIVES

Objective 1. Describe and assess current national HIV surveillance and vital statistics, with particular attention to their capacity to measure the level of and trends in HIV disease burden (incidence, prevalence, morbidity and mortality).

Objective 2. Assess the level of, and trends in, HIV disease burden using available surveillance, survey, programmatic and other data.

Objective 3. Assess causal pathways leading to impact on HIV/AIDS due to programmatic explanations and factors aside from those related to HIV programs.

The work undertaken through this consultancy will complement and feed into tasks and outputs produced by other components of the Program Review of the Health Sector Response to HIV/AIDS.

3. DETAILED TASKS AND DELIVERABLES BY OBJECTIVE

Objective 1. Describe and assess current national HIV surveillance and vital statistics, with particular attention to their capacity to measure the level of and trends in HIV disease burden (incidence, prevalence, morbidity and mortality).

1. The assessment of HIV-related data consists of the following sub-tasks:

1.1. Provide a written description and explanation of the main features of the current national HIV data and information system across the results chain (See Annex 1 for a diagram of the results chain). The sources should include routine program reports on inputs (resource tracking documents; national health accounts; control program budgets); service delivery outputs (including facility assessments and clinical reporting); and outcome and impact (population-based surveys; HIV surveillance; and vital registration sources). See “Q11. Strategic Information” in WHO Guidelines document. The description should include:

- a) Definition of the agencies/individuals responsible for data collection, analysis and reporting;
- b) Mechanisms/processes used to assure data quality;
- c) Timing and timeliness of reporting including lag times that hamper reporting;
- d) The type of data available at the national level and sub-national levels (e.g. aggregated/dis-aggregated reports, case-based data);
- e) Approach to analysis and reporting of data;
- f) How HIV data are related to/linked with any other health information systems (e.g. TB Program Information Systems)

1.2. Assess the current capacity of national systems, surveys and the vital registration systems to provide direct measures of HIV disease burden.

1.3. Summarize the main strengths of the current surveillance system and the weaknesses/gaps that need to be addressed, based on the findings from 1.1 and 1.2.

The output/deliverable from this task is:

A narrative description of the main features of the current national HIV data and information system touching on all the review topics mentioned above, and mentioning areas in need of improvement

Objective 2. Assess the level of, and trends in, HIV disease burden using available surveillance, survey, programmatic and other data.

1. This assessment includes review and compilation of published estimates of HIV incidence, prevalence, morbidity and mortality that are already available from existing sources to assess the level of, and trends in, HIV disease burden (at least nationally and when feasible sub-nationally and among sub-populations); and interpretation of available data.

1.1. Analysis of the level of, and trends in, HIV mortality.

- a) This is best done using data from a national or sample civil registration system of vital statistics with cause of death data.
- b) If local data are not available estimates can be obtained from WHO/UNAIDS data or from Global Burden of Diseases data.

1.2. Analysis of trends in the distribution AIDS deaths associated with TB, if data are available

- a) From 2012, estimates of TB mortality among HIV-positive people are being produced using the TB component of Spectrum, and published on an annual basis by WHO and UNAIDS

1.3. Analysis of the level of, and trends in, AIDS morbidity, HIV prevalence and HIV incidence.

1.4. Other miscellaneous analyses may be called for in specific settings (to be determined by the epidemiologist(s) undertaking the assessment in consultation with the Review Team and key stakeholders).

The outputs/deliverables from this task are:

- Time series plots of HIV prevalence, incidence, morbidity and mortality with specific attention to dates/years of scale-up of ART services
- Analysis of the geographic distribution of AIDS and HIV incidence and prevalence, whether this has changed over time, and exploration of reasons for observed trends and geographical heterogeneity
- Trends in age- and sex-specific HIV incidence, prevalence, morbidity, mortality whether this has changed over time, and exploration of reasons for observed trends
- Any data available on HIV in key populations; numbers, denominators; and if available proportions and trends
- A set of other relevant analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to disease burden

Objective 3. Assess causal pathways leading to impact on HIV/AIDS due to programmatic explanations and factors aside from those related to HIV programs.

1. The assessment of causal pathways consists of several sub-tasks:

1.1. Hypothesis formulation.

- a) State reasonable and epidemiologically sound hypotheses to be tested. There should be one hypothesis centered on the results chain of the HIV program, and at least one hypotheses centered on factors external to HIV programs that may have influenced or results in impact on HIV (See Annex 1 for examples.)
- b) For each hypothesis developed, consider (i) whether data are likely to be available for a given country, and (ii) whether the non-program hypotheses explain trends in disease burden due to

factors not related to disease-specific funding and associated interventions. Each hypothesis should have a simple logic model/causal pathway following the results chain.

The outputs/deliverables from this sub-task are:

- A list of hypotheses and associated logic models or causal pathway diagrams.

1.2. Data Mapping.

- a) For both HIV and non-program hypotheses indicate likely or preferred sources of data needed to assess the hypotheses, and determine availability and sources of data across the results chain from inputs to outcomes (note that impact measures will have been addressed under Objectives 1 and 2). The data must be informative of each alternative hypothesis, and any gaps identified.
- b) For the HIV, or program, hypothesis of impact consult in-country stakeholders in addition to examining data sources related to:
 - Inputs (e.g. funding, policies)
 - Service delivery outputs (e.g. routine program and facility reports)

The outputs/deliverables from this sub-task are:

- A map or matrix of data and sources for each hypothesis to be tested, across the results chain, and identifying any gaps.

- Outcomes (e.g. program reports, population-base surveys)

- c) For the non-program hypothesis, see Annex 1 and consult additional in-country stakeholders as needed. In the case of the non-program hypotheses, it is understood that in some instances it

The outputs/deliverables from this sub-task are:

- Complete, analysis-ready data set

will not be possible to locate or obtain data sources for all components of the results chain.

1.3. Data Acquisition.

- a) Assemble and assess quality of available data pertaining to each hypothesis. Obtain data from relevant in country and online sources, as well as from international partners and donors where relevant.

1.4. Quality Check and Validation.

- a) Undertake standard checks on the quality of data. This would include a consideration of any known sources of measurement error and how these were addressed; completeness of the data set; and a consideration of non-measurement sources of error or bias (e.g. selection bias).

The outputs/deliverables from this sub-task are:

- A spreadsheet or table listing all data sources and assessment of quality

1.5. Analysis, Synthesis and interpretation.

- a) Analyze and interpret the evidence across the each results chain to relate the interpretation to the appropriate logic model/causal pathway and trends in HIV impact analyzed under Objective 2 of the epidemiological and impact analysis.

The outputs/deliverables from this sub-task are:

- A set of analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to the alternative hypotheses tested

1.6. Develop Plausibility Argument, Report writing, Presentation, and Use of Findings

- a) Address the objectives of the epidemiological and impact analysis outlined earlier into a report section, using output of previous tasks. This section should be included with the findings of other components of the review and should contain the following content:
- The extent to which changes in disease burden plausibly reflect programmatic efforts or appear due, in part or in whole, to other factors
 - A qualitative judgment about the relative contribution to impact of program and non-program factors should be included
 - Whether there are specific geographical areas or subpopulations in which the burden of disease is especially high and that warrant increased attention including greater investment of financial resources and/or reallocation of resources to focus on more effective, higher impact interventions
 - Potential areas of investment needed to improve evidence about impact (trends in disease burden) in future
 - Implications and recommendations to improve program management
- b) Work in collaboration with the other members of the review team under the guidance of the NACP or MOH to produce a single, complete review report incorporating the epi and impact analysis

- c) Participate in data use and dissemination workshops or presentations at the conclusion of the review period to share findings with the key in-country stakeholders.

The outputs/deliverables from this sub-task are:

- Complete epidemiological and impact analysis section of National AIDS Program Review Report, with a few, prioritized recommendations on how to improve the program for impact
- Potential participation in data use and dissemination activities

4. SUMMARY TABLE OF DELIVERABLES

Objective	Deliverable
Objective 1	<ul style="list-style-type: none"> • A narrative description of the main features of the current national HIV data and information system touching on all the review topics mentioned under Objective 1, and mentioning areas in need of improvement
Objective 2	<ul style="list-style-type: none"> • Time series plots of HIV prevalence, incidence, morbidity and mortality with specific attention to dates/years of scale-up of ART services • Analysis of the geographic distribution of HIV incidence and prevalence, whether this has changed over time, and exploration of reasons for observed trends and geographical heterogeneity • Trends in age- and sex-specific HIV incidence, prevalence, morbidity, mortality whether this has changed over time, and exploration of reasons for observed trends • Any data available on HIV in key populations; numbers, denominators; and if available proportions and trends • A set of other relevant analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to disease burden
Objective 3	<ul style="list-style-type: none"> • A list of hypotheses and logic models or causal pathway diagrams. • A map or matrix of data and sources for each hypothesis to be tested, across the results chain, and identifying any gaps. • Complete, analysis-ready data set. • Spreadsheet or table listing all data sources and assessment of quality

	<ul style="list-style-type: none"> • A set of analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to the alternative hypotheses tested • Complete epidemiological and impact analysis analysis section of National AIDS Program Review Report and potential participation in data use and dissemination activities
--	--

5. PROFILE REQUIRED

A senior epidemiologist or statistician with extensive quantitative analytical skills and a proven track record of producing results and communicating them well

Excellent HIV understanding of epidemiology

Extensive experience in working with national health programs and offering technical assistance, preferably on HIV

Familiarity with country data systems beyond HIV

6. TIME REQUIRED

This depends in part on the extent to which the person(s) conducting the analysis are already familiar with the country where the assessment is being done and the associated data, and their previous experience of conducting such analyses. For someone familiar with the country and the data and with previous experience of such work, it is estimated that 2-3 weeks is required. For other people that meet the profile defined in section 5, it is estimated that 2 weeks of preparatory work are required to compile all necessary data and other information, plus an additional 1-2 weeks of in-country work. The work may be spread out over a few months to coincide with other components of the Program Review.

Annex 1. Results Chain and List of Potential Alternative Hypotheses

Results Chain

Review questions	Level
Are the right things being done?	Inputs
Are they being done the right way?	Activities
Are they being done on a large enough scale?	Outputs
Are the right people being reached?	Outcomes
Is the programme making a difference?	Impact

Source: WHO Guide for Conducting Programme Reviews for the Health Sector Response to HIV/AIDS

Alternative explanatory factors	Mechanism of influence	Suggested data sources
Other Health Programs		
RMNCH/STD: improving diagnosis and treatment	<ul style="list-style-type: none"> - Reduce vulnerability for HIV transmission - Also may lead to reporting bias 	Program data/HMIS, surveys
TB programs (Within HIV program)	<ul style="list-style-type: none"> - Improved case findings/reporting of cases - (HIV prevention and treatment interactions should be considered) 	Program data
Non-program factors		
Secular trend, natural dynamics	<ul style="list-style-type: none"> - Population at risk can be saturated - ART needs/AIDS mortality can be reduced due to incidence decline a decade ago - HIV prevalence can be increased due to increased survival 	HIV epi and behaviour data, modeling
Socio-economic development	<ul style="list-style-type: none"> - Economic level may affect indirectly through other factors - Increased sex work, tourism, etc. - Other job opportunities - Improved literacy - New tools/opportunities for networking/finding sexual partners 	UNDP
Population dynamics (including IDP, refugee)	<ul style="list-style-type: none"> - Population increase will increase population at risk (sexually active) - Urbanization may change risk behaviour, access to health - Migration (e.g., mining, between border) can change population at risk - IDP camps may provide better access to health 	Ministry of Planning, Statistical Office, UNFPA, UN Population division, Provincial health authority (for development projects)
Legal, political changes; changes in stigma, discrimination, human rights situation	<ul style="list-style-type: none"> - Influence program implementations (including access to KAP) - Influence populations (behaviour, access, etc) - Raid, arrest, detention of people with high risk behaviour 	Legal documents, programs, media, civil society data
Measurement issues/bias		
Better access to health facilities	<ul style="list-style-type: none"> - Incentive/motivation to go to health facilities (e.g., for treatment) - Better transportation, etc. 	Surveys
Changes in surveillance systems and reporting	<ul style="list-style-type: none"> - Change in eligibility criteria for ART - Changes in case definitions - Better diagnosis (sensitivity /specificity) - Sentinel sites usually start in higher burden areas, and then expand to lower burden areas 	

National health sector and national TB programme reviews, and “Epidemiological stage” for Global Fund concept notes:

Terms of reference for TB epidemiological and impact analysis

1. BACKGROUND

An excellent understanding of the level of, and trends¹ in, disease burden and how these have been (and can be) influenced by the implementation of prevention and treatment interventions is of considerable importance to national health programmes, as well as international donor agencies. It can help to ensure the appropriate allocation of funding and ultimately help to save more lives in the future. Epidemiological and impact analysis should be included systematically as part of National Health Sector Reviews and disease-specific programme reviews. Such analyses are also now required as part of the development of “concept notes” that provide the basis for funding applications to the Global Fund in the new funding model introduced in 2013; in this context, the analyses are called the “Epidemiological stage”, and should precede the development of the concept note. These terms of reference cover the objectives and associated tasks and expected deliverables for TB epidemiological and impact analyses conducted as part of national TB programme reviews, as inputs to health sector reviews and for the “epidemiological stage” of the Global Fund’s new funding model.

2. OBJECTIVES

1. Describe and assess current national TB surveillance and vital registration systems, with particular attention to their capacity to measure the level of and trends in TB disease burden (incidence and mortality).
2. Assess the level of, and trends in, TB disease burden (incidence, prevalence, mortality) using available surveillance, survey, programmatic and other data.
3. Assess whether recent trends in TB disease burden indicators are plausibly related to changes in TB-specific interventions taking into account external factors including economic or demographic trends.
4. Define the investments needed to directly measure trends in TB disease burden in future.

3. TASKS BY OBJECTIVE

Objective 1: Assessment of current national TB surveillance and vital registration systems with particular attention to their capacity to measure the level of and trends in TB disease burden

a) Provide a written description and explanation of the main features of the current national TB surveillance and vital registration systems. These should include the data being captured (e.g. notified cases, treatment outcomes, causes of death); definition of the agencies/individuals responsible for data collection, analysis and reporting and how they interact; mechanisms/processes used to capture and transmit data between different administrative levels and agencies (e.g. standardized forms; paper-based and/or electronic systems) and to assure data quality; timing and timeliness of reporting including lag times that hamper capacity to detect, investigate and contain events such as local epidemics (including events related to the emergence of drug resistance); the type of data available at the national level (e.g. aggregated reports, case-based data); approach to analysis and reporting of data; staffing levels; how TB data are related to/linked with other health information systems (e.g. health insurance, hospital reporting systems, district health information systems). To help characterize the TB surveillance system, Part A of the WHO TB surveillance checklist (18 questions) should be completed.²

b) Assess the current capacity of national TB notification and vital registration systems to provide a direct measure of TB disease burden using the WHO TB surveillance checklist (Part B). *The ultimate goal is to measure TB incidence and mortality directly from notification and vital registration data, respectively; Part*

¹ Analyses of time trends should be attempted as far back in time as possible before the health sector or programme review.

² http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/en/

B of the checklist consists of a set of 13 standards and associated benchmarks that allow assessment of the extent to which existing surveillance systems (notification and vital registration) meet these standards. (NB the first standard in the checklist relates to case definitions. In this context, there should be an assessment of whether the 2013 WHO revised case definitions and reporting framework have been adopted and implemented, and at what scale, and any actions needed to introduce or fully implement them).

c) Summarize the main strengths of the current surveillance system and the weaknesses/gaps that need to be addressed, based on the findings from a) and b).

(Suggested data sources³: Interviews with relevant staff; national and sub-national case-based or aggregated TB notification data, national or sample vital registration data, results from facility audits (e.g. Service Availability and Readiness Assessment, SARA) or reviews of the quality of recorded data, results from drug resistance surveillance including drug resistance surveys, research literature). A comprehensive list of data sources is provided in the user guide that accompanies the checklist).

Objective 2: Assessment of the level of, and trends in, TB disease burden

This assessment includes review and compilation of published estimates of TB morbidity and mortality that are already available to assess the level of, and trends in, TB disease burden (at least nationally and when feasible sub-nationally and among sub-populations); analysis of TB notification data; and interpretation of available data.

a) Analysis of the level of, and trends in, TB mortality.

- i. Analysis of trends in TB mortality among HIV-negative individuals. This is best done using data from a national or sample civil registration system of vital statistics with cause of death data that meet the standards defined in the WHO TB surveillance checklist. Each year, WHO publishes estimates of TB mortality among HIV-negative people from 1990 onwards for all countries in the annual global TB report (the global TB report also identifies the countries for which mortality among HIV-negative individuals has been estimated from vital registration data and mortality surveys, and the countries for which estimates rely on other methods).
- ii. Analysis of trends in the distribution of contributory causes of AIDS deaths (with particular emphasis on TB), if data are available. From 2012, estimates of TB mortality among HIV-positive people are being produced using the TB component of Spectrum, and published on an annual basis by WHO and UNAIDS.

(Suggested data sources: WHO TB database, AIDSinfo database, records from national or sample civil registration of vital statistics with cause of death data from NTP/MoH databases, results from mortality surveys, research literature).

b) Analysis of the level of, and trends in, TB prevalence. If data are available from a baseline and at least one repeat survey, then there is strong evidence about trends in disease burden. If results from two surveys conducted about 10 years apart are not available, estimates of trends are available from WHO but uncertainty intervals are wide. The results from a recent survey can be used to assess the current level of TB disease burden and may also provide important evidence about the effectiveness of current TB programmatic efforts and actions needed to improve TB care and control.

(Suggested data sources: results from surveys of the prevalence of TB disease, WHO TB database, research literature)

³ It is likely that some of the suggested data are not yet available. The identification of these data gaps is important and they should be identified in a specific section of the final report, along with clearly defined next steps for addressing these gaps.

- c) Analysis and interpretation of the level of, and trends in, TB case notifications (e.g. for the last 5-10 years).
- i. Plot time series of case notifications and analyse results, including to assess trends and to identify if there is any evidence of reporting problems (e.g. missing data or sudden changes in time-series of reported new episodes of TB at national and first subnational level e.g. state or province). Analysis of results should take into consideration any changes in reporting policies and practices, and case definitions.
 - ii. Analysis of the geographic distribution of case notification rates among subnational areas and how this has changed over time, and exploration of reasons for observed trends and geographical heterogeneity. These include, but are not limited to, the availability of TB diagnostic services, case finding activities, changes in the ratio of TB cases to the number of people investigated for “presumptive” TB (note that data on the number of people investigated for TB are often not quality-assured and duplicate entries from multiple visits by the same person may exist), health systems characteristics, determinants of/risk factors for TB (e.g. overall levels of income and poverty, HIV prevalence).
 - iii. Analysis of trends in the proportions of notified cases: (a) by type of TB disease - bacteriologically confirmed and extra-pulmonary TB; (b) by age group, including the proportion of cases among children (0-4, 5-14); (c) by category (retreatment out of the sum of new and retreatment cases).
 - iv. Trends in age- and sex-specific case notification rates, the average age of newly notified cases, and the extent to which these can be explained by demographic or other factors.
 - v. Analysis of the level of (and ideally trends in) under-reporting from national inventory studies if these are available before the assessment.
 - vi. Any data available on TB in high risk groups such as people living with HIV, the elderly, people with diabetes, people with compromised immune systems, prisoners, miners, etc.; numbers, denominators; and if available proportions and trends.
 - vii. Other miscellaneous analyses that may be relevant in specific settings (to be determined by the epidemiologist(s) undertaking the assessment).

(Suggested data sources: National and sub-national case-based or aggregated TB notifications, laboratory data, results from inventory studies to measure TB under-reporting (and under certain circumstances estimate incidence), laboratory data, research literature, national databases with information about overall health system characteristics and determinants/risk factors related to TB).

Objective 3: Are recent trends in TB disease burden plausibly related to changes in TB-specific interventions accounting for other external factors?

Funding for and implementation of high-quality TB-specific interventions should result in detection of people with TB and curative treatment; in turn, this should have a direct impact on TB mortality (cutting case fatality rates compared with no treatment or substandard treatment). Shortening the duration of disease through detection and treatment of cases will also reduce the prevalence of TB disease, and therefore, transmission. There will be an impact on TB incidence if transmission can be reduced sufficiently and/or if preventive treatment of people with latent TB infection is effectively implemented on a large scale. At the same time, a range of factors besides TB-specific interventions influence levels of TB disease burden, by affecting population susceptibility to both TB infection and the risk of developing TB disease once infected. These include overall levels of wealth and the distribution of wealth (measured e.g. as GNI per capita, the proportion of people living in poverty), the overall coverage and quality of health services and the prevalence of HIV and other risk factors

for TB. Having considered trends in disease burden in Objective 2, it is important to assess whether these trends can partly be related to changes in TB-specific interventions (and associated funding).

a) Define and compile data that are relevant to assessment of the extent to which changes in TB disease burden in recent years (e.g. for the last 5–10 years) can be explained by TB-specific interventions/programmatic efforts. This should include, at a minimum:

- i. Government and international donor funding for TB care and control;
- ii. Number of health facilities providing TB diagnostic services per 100,000 population;
- iii. Number of health facilities providing TB treatment services per 100,000 population;
- iv. Number of people investigated for presumptive TB (if available data are reliable) and the ratio of presumptive TB to notified TB cases;
- v. Performance of community/active case finding (number of cases screened and detected by each mechanism);
- vi. Performance and coverage of public-private mix activities in the country. Coverage should be expressed where possible both as % of the country (geographic) and type, the % of providers covered (e.g., 30% of estimated pharmacies and 50% of estimated private pulmonologists);
- vii. Any quantitative data on diagnostic delays (due to patient, private sector, or public sector delays);
- viii. Number of people successfully treated for TB out of all notified;
- ix. MDR-TB treatment coverage (comparing numbers detected and treated with the estimated number of cases among notified TB patients and describing the size of waiting lists), and treatment outcomes among MDR-TB patients. This is especially relevant in countries in which MDR-TB cases account for a relatively large share of the total number of TB cases;
- x. HIV testing, ART and CPT coverage of TB patients, treatment outcomes among PLHIV. This is especially relevant in countries with a high TB/HIV burden.

(Suggested data sources: WHO TB database, NTP database and reports, Service Availability and Readiness Assessments (SARAs), results from inventory studies that show the level of TB under-reporting, research literature, grey literature, national TB prevalence surveys, WHO HIV/AIDS data and statistics, AIDSinfo database, MOH and NGO databases, <http://www.foreignassistance.gov> for USAID funding data).

b) Define and compile data that are relevant to assessment of the extent to which changes in TB disease burden in recent years can be explained by factors that are not specifically related to TB-specific funding and associated interventions. This should include, at a minimum:

- i. Prevalence of HIV among the general population, and ART coverage. *(Suggested data sources: WHO HIV/AIDS data and statistics, AIDSinfo database);*
- ii. Prevalence of diabetes, tobacco use and under-nutrition. *(Suggested data sources: WHO HIV/AIDS data and statistics, AIDSinfo database, WHO Global Health Observatory);*
- iii. GNI per capita and the % of the population under the poverty line, and the impact of economic crises. *(Suggested data sources: World Bank Indicators);*
- iv. Coverage of financial protection for health care costs (by government health budget or health insurance etc.) and social protection programmes (overall, and for DS-TB and MDR-TB specifically where available) and the percentage of health-care expenditures accounted for by out-of-pocket payments *(Suggested data sources: Research literature, national health accounts, social protection/welfare programme information on coverage of target groups, as relevant and available from WHO at <http://www.who.int/nha>; research literature) ;*

- v. Demographic changes; percentage of population who are less than 15, and those more than 65, years (*Suggested data sources: UNPD database*)
- vi. Under-5 mortality rate (as an indicator of the overall performance of the health-care system). (*Suggested data sources: WHO Global Health Observatory*)

Objective 4: Assessment of investments needed to directly measure trends in disease burden in the future

- a) From the implementation of the WHO TB surveillance checklist: for standards defined in the checklist that are not yet met due to data gaps or data quality problems, identification of the investments required to improve surveillance (including estimated budget). (*Suggested data sources: same as in 1.b, NTP reports*)
- b) Assessment of whether a baseline or repeat survey (e.g. prevalence survey, inventory study, cause of death survey) is needed and if so what timing would be appropriate. An appropriate amount of time should be ensured between repeat surveys (for example, a repeat TB prevalence survey should normally be done about 10 years after the previous one). Guidance on countries where prevalence surveys are recommended is available from the Global Task Force on TB Impact Measurement.

4. DELIVERABLES

A comprehensive report addressing all tasks under the three objectives of the epidemiological and impact analysis outlined in this document with a conclusion section on:

- a) The robustness of estimates of TB incidence, prevalence and mortality and their sources of uncertainty.
- b) Whether it is plausible that TB control interventions have contributed to changing the course of the TB epidemic, accounting for other external factors.
- c) Whether there are specific geographical areas or subpopulations (vulnerable/those with poor access) or sectors (e.g. mining, prisons/detention, etc.) in which the burden of disease is especially high and that warrant increased attention including greater investment of financial resources and/or reallocation of resources to focus on more effective, higher impact interventions.
- d) Investments needed to improve evidence about trends in disease burden in future.


5. PROFILE REQUIRED

- A senior epidemiologist or statistician with extensive quantitative skills and a proven track record of producing results and communicating them well (including in scientific publications in peer reviewed journals);
- Excellent understanding of epidemiology, preferably with emphasis on TB epidemiology, TB policies and interventions, and health systems;
- Extensive experience in working with national health programmes and offering technical assistance, preferably on TB.

6. TIME REQUIRED

This depends in part on the extent to which the person(s) conducting the analysis are already familiar with the country where the assessment is being done and the associated data, their previous experience of conducting such analyses, but also the availability and expertise of national M&E counterparts who will participate in this exercise. For someone familiar with the country and the data and with previous experience of such work, it is estimated that 2 weeks of in-country work are required. An additional 2 weeks of preparatory work might be necessary depending on the country context.

Guidance on and related examples of schedules for previous missions that covered the Terms of Reference described are available from WHO and KNCV on request.



National health sector and national Malaria
programme reviews:

Terms of reference for Epidemiological Impact Analysis for Global Fund concept notes

1. BACKGROUND

Program Reviews are important management tools that provide managers and other stakeholders with an opportunity to assess program performance, impact and lessons learnt during a specific time period. Program Reviews also inform program improvement and the subsequent strategy cycles, and should be linked to health sector reviews. A substantial amount of planning and resources go into a national disease control Program Review.

As part of Program Review, it is also important to assess periodically the causal pathways between investments and impact, and assess other possible explanations for impact aside from those related to programs. Such analysis, called an “Epidemiological and impact analysis” can significantly strengthen a program and its management as well as an application for funding from the Global Fund and other donors by focusing on investing for impact.

The purpose of an epidemiological and impact analysis is to provide a more robust assessment of whether impact and outcomes are plausibly resulting from program input and activities, or might be due to other factors. It is about further assessing contribution and causation along the results chain, i.e. answering the question “Have the interventions and other competing explanations or hypotheses contributed to and resulted in these impacts, whether positive or negative?” This evidence will form a key basis for important funding recommendations and decisions.

The services of a consultant epidemiologist, or an evaluator with epidemiological analysis expertise, are required to carry out the tasks and provide deliverables for the epidemiological and impact analysis.

This consultancy, associated tasks, and deliverables should be considered part of a larger Malaria Program Review effort (see Guidance Note for Epidemiological and impact analysis for the New Funding Model)

2. OBJECTIVES

- **Objective 1.** Describe and assess current national malaria surveillance and vital statistics, with particular attention to their capacity to measure the level of and trends in disease burden (incidence, prevalence, morbidity and mortality).
- **Objective 2.** Assess the level of, and trends in, malaria burden using available surveillance, survey, programmatic and other data.
- **Objective 3.** Assess causal pathways leading to impact on malaria due to programmatic explanations and factors aside from those related to malaria programs.

The work undertaken through this consultancy will complement and feed into tasks and outputs produced by other components of the Program Review of the Malaria Control Program

3 DETAILED TASKS AND DELIVERABLES

The detailed tasks associated with this consultancy are as follows:

Objective 1. Describe and assess current national malaria surveillance and vital statistics, with particular attention to their capacity to measure the level of and trends in disease burden (incidence, prevalence, morbidity and mortality).

1. The assessment of malaria-related data consists of the following sub-tasks:

1.1. Provide a written description and explanation of the main features of the current national malaria data and information system across the results chain (See Annex 1 for a diagram of the results chain). The sources should include routine program reports on inputs (resource tracking documents; national health accounts; control program budgets); service delivery outputs (including facility assessments and clinical reporting); and outcome and impact (population-based surveys; malaria surveillance; and vital registration sources). The description should include:

- a) Definition of the agencies/individuals responsible for data collection, analysis and reporting;
- b) Mechanisms/processes used to assure data quality;
- c) Timing and timeliness of reporting including lag times that hamper reporting;
- d) The type of data available at the national level and sub-national levels (e.g. aggregated/dis-aggregated reports, case-based data);
- e) Approach to analysis and reporting of data;

1.2. Assess the current capacity of national systems, surveys and the vital registration systems to provide direct measures of malaria disease burden.

1.3. Summarize the main strengths of the current surveillance system and the weaknesses/gaps that need to be addressed, based on the findings from 1.1 and 1.2.

The output/deliverable from this task is:

- A narrative description of the main features of the current national malaria data and information system touching on all the review topics mentioned above, and mentioning areas in need of improvement

Objective 2. Assess the level of, and trends in, malaria disease burden using available surveillance, survey, programmatic and other data.

1. This assessment includes review and compilation of published estimates of malaria incidence, prevalence, morbidity and mortality that are already available from existing sources to assess the level of, and trends in, disease burden (at least nationally and when feasible sub-nationally and among sub-populations); and interpretation of available data. See Annex 2 for suggested outcome and impact indicators.

1.1. Analysis of the level of, and trends in, malaria-specific mortality.

- a) This is best done using data from a national or sample civil registration system of vital statistics with cause of death data.
- b) If local data are not available estimates can be obtained from CHERG data (for children) or from Global Burden of Diseases data (for all ages).

1.2. Analysis of the level of, and trends in, malaria morbidity, prevalence and incidence as well as prevalence of severe anemia children.

1.3. Other miscellaneous analyses may be called for in specific settings (to be determined by the epidemiologist(s) undertaking the assessment in consultation with the Review Team and key stakeholders).

The outputs/deliverables from this task are:

- Time series plots of malaria and anemia prevalence, incidence, morbidity and mortality with specific attention to dates/years of scale-up of key interventions
- Analysis of the geographic distribution of malaria incidence and prevalence, whether this has changed over time, and exploration of reasons for observed trends and geographical heterogeneity
- Trends in age- and sex-specific malaria incidence, prevalence, morbidity, mortality whether this has changed over time, and exploration of reasons for observed trends
- A set of other relevant analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to disease burden

Objective 3. Assess causal pathways leading to impact on malaria due to programmatic explanations and factors aside from those related to malaria control and elimination programs.

1. The assessment of causal pathways consists of several sub-tasks:

1.1. Hypothesis formulation.

- a) State reasonable and epidemiologically sound hypotheses to be tested. There should be one hypothesis centered on the results chain of the malaria control and elimination program, and at least one hypotheses centered on factors external to malaria programs that may have influenced or results in impact on malaria (See Annex 1 for examples.)
- b) For each hypothesis, consider (i) whether data are likely to be available for a given country, and (ii) whether the non-program hypotheses explain trends in disease burden due to factors not related to disease-specific funding and associated interventions. Each hypothesis should have a simple logic model/causal pathway following the results chain.

The outputs/deliverables from this sub-task are:

- A list of hypotheses and associated logic models or causal pathway diagrams.

1.2. Data Mapping.

- a) For both malaria programmatic and non-programmatic hypotheses indicate likely or preferred sources of data needed to assess the hypotheses, and determine availability and sources of data across the results chain from inputs to outcomes (note that impact measures will have been addressed under Objectives 1 and 2). The data must be informative of each alternative hypothesis, and any gaps identified.
- b) For the malaria program hypotheses of impact consult in-country stakeholders in addition to examining data sources related to:
 - Inputs (e.g. funding, policies)
 - Service delivery outputs (e.g. routine program and facility reports)
 - Outcomes (e.g. program reports, population-base surveys)
- c) For the non-program hypotheses, see Annex 1 and consult additional in-country stakeholders as needed. In some of the non-programmatic hypotheses, it may not be possible to locate or obtain data sources for all components of the results chain.

1.3. Data Acquisition.

- a) Assemble and assess quality of available data pertaining to each hypothesis. Obtain data from relevant in country and online sources, as well as from international partners and donors where relevant.

The outputs/deliverables from this sub-task are:

- Complete, analysis-ready data set

1.4. Quality Check and Validation.

- a) Undertake standard checks on the quality of data. This include a consideration of any known sources of measurement error and how these were addressed; completeness of the data set; and a consideration of non-measurement sources of error or bias (e.g. selection bias).

The outputs/deliverables from this sub-task are:

- A spreadsheet or table listing all data sources and assessment of quality

1.5. Analysis, Synthesis and interpretation.

- a) Analyze and interpret the evidence across the each results chain to relate the interpretation to the appropriate logic model/causal pathway and trends in malaria impact analyzed under Objective 2 of the epidemiological and impact analysis.

The outputs/deliverables from this sub-task are:

- A set of analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to the alternative hypotheses tested

1.6. Develop Plausibility Argument, Report writing, Presentation, and Use of Findings.

- a) Address the objectives of the epidemiology and impact analysis outlined earlier into a comprehensive report section, using output of previous tasks. This section should be included with the findings of other components of the review and should contain the following content:
- The extent to which changes in disease burden plausibly reflect programmatic efforts or appear due, in part or in whole, to other factors
 - A qualitative, expert judgment about the relative contribution to impact of program and non-program factors should be included
 - Whether there are specific geographical areas or subpopulations in which the burden of disease is especially high and that warrant increased attention including greater investment of financial resources and/or reallocation of resources to focus on more effective, higher impact interventions
 - Potential areas of investment needed to improve evidence about impact (trends in disease burden) in future
 - Implications and recommendations to improve program management
- b) Work in collaboration with the other members of the review team under the guidance of the NMCP or MOH to produce a single review report incorporating the epidemiology and impact analysis
- c) Participate in data use and dissemination workshops or presentations at the conclusion of the
- d) review period to share findings with the key in-country stakeholders.

The outputs/deliverables from this task are:

- Complete epidemiological and impact analysis analysis section of Malaria Program Review Report, with a few, prioritized recommendations on how to improve the program for impact
- Potential participation in data use and dissemination activities

Objective	Deliverable
Objective 1	<ul style="list-style-type: none"> • A narrative description of the main features of the current national malaria data and information system touching on all the review topics mentioned under Objective 1, and mentioning areas in need of improvement
Objective 2	<ul style="list-style-type: none"> • Time series plots of malaria prevalence, incidence, morbidity and mortality with specific attention to dates/years of scale-up of key interventions • Analysis of the geographic distribution of malaria incidence and prevalence, whether this has changed over time, and exploration of reasons for observed trends and geographical heterogeneity • Trends in age- and sex-specific malaria incidence, prevalence, morbidity, mortality whether this has changed over time, and exploration of reasons for observed trends • A set of other relevant analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to disease burden

<p>Objective 3</p>	<ul style="list-style-type: none"> • A list of hypotheses and associated logic models or causal pathway diagrams. • A map or matrix of data and sources for each hypothesis to be tested, across the results chain, and identifying any gaps. • Complete, analysis-ready data set. • Spreadsheet or table listing all data sources and assessment of quality • A set of analytics including tables, charts, and/or graphs and accompanying narrative containing detailed, evidence-based interpretations and conclusions reached with respect to the alternative hypotheses tested • Complete epidemiological and impact analysis analysis section of National Malaria Program Review Report
---------------------------	--

4 SUMMARY TABLE OF DELIVERABLES

5. PROFILE REQUIRED

A senior epidemiologist or statistician with extensive quantitative skills and a proven track record of producing results and communicating them well;

Excellent understanding of malaria epidemiology, including awareness of strengths and limitations of data on all-cause; cause-specific; and verbal-autopsy-derived mortality measurements;

Extensive experience in working with national health programs and offering technical assistance, preferably in malaria.

6. TIME REQUIRED

This depends in part on the extent to which the person(s) conducting the analysis are already familiar with the country where the assessment is being done and the associated data, and their previous experience of conducting such analyses. For someone familiar with the country and the data and with previous experience of such work, it is estimated that 2-3 weeks is required. For other people that meet the profile defined in section 5, it is estimated that 2 weeks of preparatory work are required to compile all necessary data and other information, plus an additional 1-2 weeks of in-country work. The work may be spread out over a few months to coincide with other components of the Program Review.

Annex 1. Results Chain and List of Potential Alternative Hypotheses

Results Chain

Review questions	Level
Are the right things being done?	Inputs
Are they being done the right way?	Activities
Are they being done on a large enough scale?	Outputs
Are the right people being reached?	Outcomes
Is the programme making a difference?	Impact

Source: WHO Guide to Conducting Programme Reviews for the Health Sector Response to HIV/AIDS

Alternative explanatory factors	Mechanism of influence	Suggested data sources
<i>Other Health Programs</i>		
Other health programs, e.g. Better immunization coverage, nutrition, other infectious diseases (pneumonia, diarrhoea)	<ul style="list-style-type: none"> - Reducing anemia - Reducing child mortality 	Program data/HMIS; WHO/UNICEF Joint Reporting Forms for immunization; DHS/MICS
Better access to health facilities	<ul style="list-style-type: none"> - Incentive/motivation to go to health facilities (e.g., for treatment) - Better transportation, etc. 	HMIS data; DHS; LSMS or household budget surveys
Availability of over-the-counter drugs	<ul style="list-style-type: none"> - Self-medication/home-based treatment, lack of reporting 	Pharmacy or drug shop surveys; Malaria Indicator Surveys
<i>Secular trends</i>		
Economic development, e.g. increased SES; Improved water and sanitation; women's education level	<ul style="list-style-type: none"> - Economic level may affect indirectly through other factors 	UNDP; World Bank World Development Indicators; LSMS or household budget surveys (e.g. accessible through the World Bank micro-data library)
Population dynamics (including migrant workers, IDP, refugee, conflicts/war)	<ul style="list-style-type: none"> - Reductions in total, child, and maternal mortality - Population increase will increase population at risk - Urbanization may reduce population at risk - Internal or cross-border migration, including development projects (e.g., irrigation, dam, roads, etc.), can change population at risk - Sometimes country's territories change 	Ministry of Planning; Statistical Office; UNFPA; UN Population Division; Provincial health authority (for development projects); DHS; MICS
Climate changes/rainfall	<ul style="list-style-type: none"> - Changes in vector 	Meteorological institute, Earth Institute (Columbia Univ.)
Deforestation	<ul style="list-style-type: none"> - Changing entomology and exposure to risk 	Ministry of Planning, Global Forest Watch
Changes in ecology	<ul style="list-style-type: none"> - Changes in vector, parasite species, drug resistance, human host susceptibility (mainly migration) can also affect malaria incidence 	Program data, entomological

Annex 2: Malaria indicators and sources

Data should be:

- From at least two time points
- Measured using comparable methodologies
- Available for at least the national level and when feasible sub-nationally and among sub-populations.

Indicator/Data item	Suggested Data Sources
Inputs & processes	
• Finance	NMCP, MoH, National Health Accounts, Global Fund
• Significant process or activity milestones	NMCP, MoH
Outcomes	
• ITN and IRS coverage	DHS, NMP, MoH
• % suspected cases tested	HMIS, Malaria Indicator Survey, local research studies
• % cases treated with effective ACT	HMIS, Malaria Indicator Survey
Impact	
• Malaria parasite prevalence*	Results from surveys of malaria parasite prevalence/research literature** NMCP; biomarker surveys; Malaria Indicator Surveys; research literature
• Number of outpatient cases & malaria admissions	HMIS; NMCP
• Number of cases of severe malaria	HMIS; NMCP
• Anemia*	DHS; Malaria Indicator Survey
• All cause <5 mortality*	National or sample civil registration of vital statistics with cause of death data; DHS; results from mortality surveys; Inter-agency group on Mortality Estimation (IGME)
• Malaria case incidence	HMIS; NMCP
• Severe malaria	HMIS; NMCP
• Anemia admissions	HMIS
• Malaria-specific mortality*	National Vital Statistics (if available); WHO; GBD data;

* Particularly among children under five; population-based surveys should provide at least two data points using comparable survey approaches.

* If data are available from a baseline and at least one repeat survey, then there is strong evidence about trends in disease burden. If results from two surveys conducted about 10 years apart are not available, estimates of trends are available from WHO but uncertainty intervals are wide.

* for countries with high endemicity of falciparum malaria that accounts for >5% of estimated child deaths

To address missing data points in surveillance data, such as data from a particular facility-month, one may need to impute data or restrict analysis to facilities with consistent reporting, or do both.

* for malaria it is useful to note that reliable malaria-specific mortality data from routine sources such as civil registration will rarely be available. For high-burden counties especially, the plausible association between malaria program scale-up and reduction of all-cause under-five mortality will be adequate (Rowe et al., 2007. Predictions of the impact of malaria control efforts on all-cause child mortality in sub-Saharan Africa. Am J Trop Med Hyg. (77)6 suppl:48-55)