Knowledge & Library Services (KLS) Evidence Briefing

What approaches are there for monitoring and evaluation of global public health programmes?

Nicola Pearce-Smith
26th July 2018
Question
This briefing summarises the evidence on the options for monitoring and evaluation of global public health programmes, including types of indicators and qualitative vs quantitative measures, from January 1st 1980 to July 17th 2018.

Key messages
- 3 areas to improve future evaluation of complex global health initiatives are: the importance of theory of change; the need to use multiple methods to address complexity; and the need to focus more on triangulation and synthesis of findings
- the BetterEvaluation Rainbow Framework can help evaluators navigate the choices available at each stage of an evaluation
- two commonly used programme frameworks are the logical framework and the results framework
- there are several types of theory-driven evaluation approaches - logic models, logical frameworks, outcomes hierarchies, realist evaluation and theory of change
- a successful evaluation involves developing an adequate theory of change, and identifying the critical links in program planning, implementation, and delivery
- the limitation of quantitative indicators are that they often need to be interpreted through qualitative enquiry; the limitation of qualitative indicators is that they often apply to a small number of people and may not be representative of a larger group
- there are four categories of performance indicators (impacts, outcomes, outputs, and inputs) each with their strengths and weaknesses
What approaches are there for monitoring and evaluation of global public health programmes?

- selecting performance indicators can be done by using SMART (for quantitative indicators) or SPICED (mainly for qualitative indicators)

- setting indicators with key stakeholders and communities is important because it results in more realistic, meaningful and achievable indicators

- there are alternatives to indicators, that rely on the generation of qualitative data, such as beneficiary assessment, focus groups, key informant interviews, rapid assessment and special studies

- mixed methods (MM) seek to draw on the strengths of quantitative and qualitative methods, and limit their weaknesses

- there are five main reasons for using a MM - triangulation of evaluation findings, development, complementarity, initiation and value diversity

- mixed methods may involve a trade-off in terms of rigour, scope, time available to conduct the research and the usefulness of findings

- studies have found that there is a gap between ‘best practice’ in program evaluation and the evaluative strategies actually being used - some impact and performance evaluations did not use relevant data, did not follow accepted social science methods or did not have high analytical validity and reliability; also some had too many indicators, which lacked validity, and did not fit into a logical framework
What approaches are there for monitoring and evaluation of global public health programmes?

Background

An editorial in the Lancet in 2010 stated that ‘evaluation must now become the top priority in global health’, as global health investment has not been matched by an equal commitment to evaluation (1). Evaluations of large-scale public health programmes should not only assess whether an intervention works, as randomised designs do, but also why and how an intervention works (2).

Monitoring and Evaluating (M&E)

“The wide variety of methods available for conducting project and programme monitoring and evaluation (M&E) can be overwhelming” (3).

An overview of a sample of M&E tools, methods, and approaches are outlined by the World Bank, including their purpose and use; advantages and disadvantages; costs, skills, and time required; and key references (4). This document discusses:

- Performance indicators
- The logical framework approach
- Theory-based evaluation
- Formal surveys
- Rapid appraisal methods
- Participatory methods
- Public expenditure tracking surveys
- Cost-benefit and cost-effectiveness analysis
- Impact evaluation

Traditional implementation-focused M&E systems are designed to address compliance, or “did they do it” questions (5). This approach focuses on monitoring and assessing how well a programme is being executed and often links the implementation to a particular unit of responsibility, but does not provide policymakers and stakeholders with an understanding of the success or failure of that programme. Results-based M&E systems are designed to address the “so what” - they provides feedback on the actual outcomes (5). Keys steps are:

- Formulate outcomes and goals
- Select outcome indicators to monitor
- Gather baseline information on the current condition
- Set specific targets to reach and dates for reaching them
- Regularly collect data to assess whether the targets are being met
- Analyse and report the results

In 2014, a workshop entitled Evaluation Design for Complex Global Initiatives looked at transferable insights gained across the spectrum of approaches incuding choosing the evaluator, framing the evaluation, designing the evaluation, gathering and
analysing data, synthesising findings and recommendations, and communicating key messages (6) (7). Appendix D Evaluation Information Summary for Core Example Initiatives provides a comparison of the evaluations for four multi-national health initiatives: Global Fund to Fight AIDS, Tuberculosis, and Malaria, U.S. President’s Malaria Initiative, Affordable Medicines Facility–malaria and U.S President’s Emergency Plan for AIDS Relief (PEPFAR). One key message from the workshop was: "Evaluations of complex initiatives are well served by the use of a logic model, theory of change, results chain, impact pathway, or other framework to describe how the program is intended to create change and to identify potential unintended results" (6) Chapter 4.

Other key messages from this workshop were summarised (7):

- to improve future evaluation of complex global health initiatives, 3 areas were identified - the importance of theory of change for grounding complex evaluations; the need to use multiple methods to address complexity; and the need to focus more on triangulation and synthesis of findings.
- the critical role of theory of change (ToC) that depicts the series of expected causal steps between activities and impacts for optimising complex evaluations was confirmed - ToC has various other names e.g. logic models, results chain, causal chain pathway, program impact pathway, program theory, and program impact theory.
- for complex global health initiatives, evaluations need to use a methodological approach that includes multiple data collection and analysis methods. “Multiple” and “mixed” methods described how the 4 large-scale evaluation examples used a number of complementary methods to arrive at evaluation findings and conclusions.
- Multidisciplinary teams of evaluators are needed to implement the multiple research methods needed for large-scale evaluations.
- Using multiple methods to address complexity demands triangulation to consolidate multiple sources of qualitative and quantitative information. A theory of change supports robust triangulation and synthesis processes. Systematic triangulation and synthesis is important for increasing confidence in the reliability, credibility, and applicability of evaluation findings.

The authors conclude their summary by stating “We need to improve how we commission, conduct, consume and convert evaluation findings from complex global health initiatives to improvements in implementation and impact. Without sufficient and immediate attention to these 3 areas from all parties, we risk continued low returns on our evaluation investments and minimal progress in building the evidence base for improved global health” (7).
What approaches are there for monitoring and evaluation of global public health programmes?

There have been no systematic efforts to develop an approach to monitor and evaluate transitions of global health programs (8). No set of indicators will be applicable across all types of health program transitions and the use of both quantitative indicators and qualitative investigation are recommended. Quantitative indicators identify changes that have occurred, demonstrate trends, and track whether transition goals, objectives, and milestones are being met. Qualitative methods describe transition experiences, explain why changes have occurred and their repercussions, and indicate what feedback and adaptation are taking place (8).

**Frameworks**

A M&E framework outlines the plan for monitoring and evaluating across an entire program, or across different programs – it includes the monitoring strategies, any studies, reviews or evaluations to do, with details about data sources, timing, management processes, and an overall program theory/logic model (9).

When writing an M&E framework (or evaluation matrix) you need to decide which indicators you will use to measure the success of your program, and define them - if you don’t have definitions there is a risk that indicators might be calculated differently at different times, which means the results cannot be compared (10). Identify where the data will come from and decide how frequently it will be measured. There is no standard definition of a Monitoring and Evaluation (M&E) framework – it is often a table that that lists all the indicators for your program, including the data source, baseline, target, how often it will be measured, and who is responsible for measuring it.

Programme frameworks are management tools used to design, monitor and evaluate interventions (11). Two commonly used examples are the logical framework (or log frame) and the results framework (11) (12). The logical framework helps to translate strategic plans into resources (inputs), activities, outputs, and outcomes—what the program is trying to achieve and what it does to get results (13). The results framework comprises four major indicator domains: system inputs and processes, outputs, outcomes and impact. Data sources are recommended and the framework also outlines what is needed across the results chain in terms of tools for data quality assurance, synthesis and analysis, and addresses the importance of dissemination, communication and use of the monitoring and evaluation results to inform policy-making (14).

Another suggested framework is the conceptual framework, useful for identifying and illustrating the factors and relationships that influence the outcome of a programme, and are typically shown as diagrams illustrating causal linkages between the key components of a programme and the outcomes of interest (12). Conceptual frameworks are usually used for programme design rather than programme M&E.
What approaches are there for monitoring and evaluation of global public health programmes?

In 2018 a framework of Critical Evaluation with 3 levels was proposed: upstream evaluation assessing the “who” and “how” of programming decisions; midstream evaluation focusing on the “who” and “how” of selecting program objectives; and downstream evaluation, the focus of current mainstream evaluation, which assesses whether the program achieved its stated objectives (15). This framework aimed to address communities' high priority concerns.

An evaluation tool that aims to capture contextual differences may help programme implementers account for different outcomes for the same intervention in diverse settings – this tool is known as the “driver diagram” (2). Beginning with the outcome or aim, an implementation team works backward to identify the primary levers or “drivers” and the secondary activities needed to lead to that outcome.

The BetterEvaluation Rainbow Framework can help evaluators navigate the choices available at each stage of an evaluation (6) Chapter 2:

1. Define what is to be evaluated. Develop an initial description of the program being evaluated and develop a program theory or logic model.
2. Frame what is to be evaluated. This involves identifying the primary intended users, the purposes of the evaluation, the key questions and determining what “success” would look like
3. Describe what happened. Involves the use of samples, measures, or indicators; the collection and management of data and the combination of qualitative and quantitative data
4. Understand the causes of outcomes and impacts. What caused particular impacts, and did an intervention contribute to those outcomes? Do the results support causal attributions etc
5. Synthesise data to make overall judgments about the worth of an intervention. Was it good? Did it work? Was it effective? For whom did it work? In what ways did it work? Did it provide value for money etc
6. Report results and support use

There are several overlapping types of theory-driven evaluation approaches - logic models, logical frameworks, outcomes hierarchies, realist evaluation and theory of change. Logic models outline the inputs, processes, outputs and outcomes of a programme in a similar manner to ToC but are usually presented in a linear form - they can be rigid and do not make explicit the causal pathways through which change happens in the way that ToC does (16). Log frames were initially developed to summarise discussions with stakeholders, but funder-driven formats have largely reduced log frames to a results-based management tool (16).

Theory of Change (ToC) is a participatory theory driven approach to programme design and evaluation whose underlying principle is to improve our understanding of how and why a programme works. The approach was developed from theory driven evaluation approaches which include the logical frameworks and logic models (17).
What approaches are there for monitoring and evaluation of global public health programmes?

A theory of change provides an analytical framework for triangulation, but the theory may need to change as data are analysed. A successful evaluation involves developing an adequate theory of change, which requires identifying the critical links in program planning, implementation, and delivery and identifying critical conditions, assumptions, and supporting factors (6) Chapter 4 and 8.

A systematic review showed how ToCs have been developed and used to evaluate public health interventions (16). A range of methods were used to develop ToCs - participatory methods which encourage stakeholder participation and ownership of the ToC, to more evaluator focused approaches such as programme observation and review of programme documentation. The papers in the review used a variety of qualitative and quantitative data collection and analysis methods - this flexibility can be an advantage if researchers can design evaluations which seek to understand and evaluate both the outcomes and causal mechanisms, but flexibility may also result in evaluations being poorly formulated, less rigorous or incorrectly interpreted. This review provides a Checklist for reporting ToC in Public Health Intervention (16) table 5 pg14.

**Indicators**

“*The best technical set of measurable, ambitious, and achievable indicators with targets is the ideal, but in the end, it is not the indicators or targets, rather the political commitment to achieve them, that is the main determinant of outcomes*” p1457 (18).

The OECD definition of an indicator is “quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development factor” (11).

There are two different types of indicator – quantitative (numerical, such as units, prices, proportions etc), and qualitative (words, measures of changes in attitudes, knowledge and behaviour/ skills). The limitation of quantitative indicators are that they often need to be interpreted through qualitative enquiry; the limitation of qualitative indicators is that they often apply to a small number of people/situations, and may not be representative of a larger group (11).

A set of indicators for a programme should include at least one indicator for each significant aspect of the program’s activities (12). Ideally use a small number of meaningful and useful indicators (19). As the World Bank states: “*What is the ideal number of indicators for any one outcome? The minimum number that answers the question: “Has the outcome been achieved?”*” (5) p67.
What approaches are there for monitoring and evaluation of global public health programmes?

In many countries lists of well-tested indicators are currently available but are skewed towards particular elements of the results chain - the challenge is to ensure an appropriate balance across the full range (14). Baseline data are not always available, making monitoring efforts difficult. The added value of the results framework is that it brings together indicators and data sources across the results chain in its entirety.

Indicators should be tested - one way is to look at data generated from different hypothetical situations and consider how each would influence decision making. If vastly different data does not influence your decisions, the indicator is not useful and will need to be revised (11).

Performance indicators are “measures of project impacts, outcomes, outputs, and inputs that are monitored during project implementation to assess progress toward project objective” (20). These four categories of performance indicators are outlined in a table, with their strengths and weaknesses (13) pg 19. Selecting performance indicators can be done using the acronyms “SMART' or SPICED (13) (21):

SMART indicators:
Specific (to the change being measured)
Measurable (and unambiguous)
Attainable (and sensitive)
Relevant (and easy to collect)
Time bound (with term dates for measurement)

SPICED indicators:
Subjective
Participatory
Interpreted (and communicable)
Cross-checked
Empowering
Diverse and disaggregated

SMART describes the properties of the indicators themselves, while SPICED relates more to how indicators should be used. SMART is suitable for quantitative indicators in particular, it is easier to set SMART indicators than SPICED indicators, so SMART indicators are more practical (21). The SPICED approach uses a bottom-up participatory approach, it is qualitative; it appreciates local understandings of change and is a good tool for thinking about why it is important to work with communities, but it does require more involvement of community members, so more resources are needed than for the SMART approach (21).
What approaches are there for monitoring and evaluation of global public health programmes?

Another method of selecting good performance indicators is “CREAM” – the indicators should be clear, relevant, economic, adequate, and monitorable (5). The more precise the indicators, the better focused the measurement strategies will be.

Every indicator has cost and work implications, so indicators should be chosen carefully. There are a number of pros and cons associated with using predesigned indicators: (5)

- **Pros** - they can be aggregated across similar projects, programs and policies; they reduce costs of building multiple unique measurement systems; they make possible greater harmonisation of donor requirements.
- **Cons**: they often do not address country specific goals; they are often viewed as imposed, as coming from the top down; they do not promote key stakeholder participation and ownership; they can lead to the adoption of multiple competing indicators.

The concerns of interested stakeholders must be considered when selecting indicators (5). Setting indicators with your key stakeholders and communities is important because it results in more realistic, meaningful and achievable indicators than those set by top-down methods, it highlights the different information needs and ideas of change of stakeholders and community groups, and it helps to increase community ownership of and involvement in projects, learning and empowerment (21).

Does monitoring without relying just on indicators make us worry that if there are no numbers, nothing has happened? As Gegenheimer states: “Qualitative monitoring then faces a challenge: If it is not rigorous enough to be considered evaluation and not numeric enough to be an indicator, what’s the point?” (22). As a supplement to indicators, qualitative monitoring can help by using tools and techniques usually associated with formal evaluations but on a smaller scale. While quantitative indicators are emphasised in mainstream M&E approaches, they often need to be qualitative to be most effective and appropriate. The most important indicators may not be quantifiable – e.g. the number of people participating in a social network is relatively unimportant compared to the quality of relationships and dialogue within that network (21).

Indicators are a very effective way to collect and analyse data, but are not always the best methodology for collecting information for evaluation. There are some alternatives to indicators, that rely on the generation of qualitative data (23):

- **Beneficiary assessment** – a qualitative approach relying on the knowledge of the beneficiaries of the intended programme
- **Focus group discussion**
- **Key informant interviews**
What approaches are there for monitoring and evaluation of global public health programmes?

- Rapid assessment – a method used to quickly collect practical information on a situation
- Special study – to collect information on a specific behaviour, situation or population

Global monitoring involves five iterative steps: development of standard indicator definitions and measurement approaches to ensure comparability across countries; collection of high-quality data at the country level; compilation of country data at the global level; organisation of global databases; and rounds of data quality checking (24). Tensions in the selection of global monitoring indicators include the desire to have comprehensive information about the program and the need to keep the number of indicators small to minimise the reporting burden; knowing what is important for improving public health and what can be measured reasonably well given available data sources and methods; and the need for timely data to guide decision making and the cost and resources required to conduct frequent surveys (24).

Some examples of indicators used for M&E existing global health programmes are available:

- National Anti-Malaria Programme (25)
- National Leprosy Elimination programme (26)
- Revised National TB control programme (27)
- National AIDS control programme (28) (29) (30)
- National Diabetes control programme (31)

The document ‘Common M&E Framework for Global Health Initiatives’ contains a table outlining current M&E initiatives, including the main types of indicators and data sources (32).

A paper outlining practical experiences from the MEASURE Evaluation Project to illustrate real world challenges for large-scale impact evaluations, showcased some of the solutions employed to design and conduct impact evaluations for public health (33). A series of case studies highlight design and implementation challenges that required creative solutions to move forward; cross-cutting themes included challenges with identification and selection of program beneficiaries, random assignment in complex environments, identification of a robust comparison or control group, heterogeneity of program impacts, timing of baseline data collection, and absence of baseline data.
What approaches are there for monitoring and evaluation of global public health programmes?

**Mixed methods**

Mixed methods (MM) can be applied to the evaluation of a project operating in a single village or a multicomponent national development initiative involving many different international and national agencies (34). “MM approaches seek to integrate social science disciplines with predominantly QUANT and predominantly QUAL approaches to theory, data collection and data analysis and interpretation” (34) p3.

Quantitative and qualitative evaluations have a number of limitations.

Quantitative methods are:

- Costly to organise
- Do not provide contextual information
- Offer limited insights on what is happening

Qualitative methods:

- Information collected cannot be generalised
- Information harder to analyse (3)

MM seeks to draw on their strengths and limit their weaknesses. There are five main reasons for using a MM evaluation design (34):

- Triangulation of evaluation findings
- Development- using results of one method to help develop the sample for another
- Complementarity - extending the comprehensiveness of evaluation findings through results from different methods
- Initiation - generating new insights into evaluation findings through results from the different methods
- Value diversity - incorporating a wider diversity of values through the use of different methods

The World Bank does suggest that for results-based M&E systems, a simple and quantitatively measurable system is preferable rather than inserting qualitatively measured indicators upfront (5).

Many evaluations use mixed methods (MM), which combine qualitative and quantitative methods to maximise complementary strengths and compensate for limitations. Although mixed methods can bolster the validity and reliability of an evaluation’s process and findings, it is important to note that there will likely be trade-offs in rigour, scope, time available to conduct the research and the usefulness of findings (35) (34) (36).
What approaches are there for monitoring and evaluation of global public health programmes?

A number of challenges face evaluators wishing to apply mixed methods in the monitoring and evaluation of development projects: (36)

- Considerable time and effort is usually required to identify and build a team of local researchers who can work well together on a mixed methods approach
- There is a lack of local expertise in mixed method research in many countries
- There are logistical challenges in applying mixed methods in many development contexts.

Evaluating the evaluations

An assessment of how well aid agencies are evaluating programs looked at the methodological quality of 37 randomly selected programme evaluations from 5 major global health funders (37) (38). They found that less than 40% of impact evaluations and less than 10% of performance evaluations used relevant data, followed accepted social science methods or had high analytical validity and reliability. The study provides 10 recommendations to improve evaluations, including that early planning of evaluations is associated with better quality, the need for better sampling approaches in data collection, and the disclosure of potential conflicts of interest and data (38).

There is also a gap between ‘best practice’ in program evaluation and the evaluative strategies actually being used (39).

A qualitative study investigated users' experiences of M&E systems in the context of HIV interventions in western India (40). NGO employees described a major gap between what they considered their “real work” and the indicators used to monitor it and believed that inappropriate and unethical means were being used to meet targets, and criticised indicators for being misleading and inflexible. Monitoring appeared to constitute an instrument of performance management rather than as a means of rational programme improvement.

A systematic analysis of diverse international health research funders’ evaluations uncovered a broad set of indicators including metrics available to measure return on investment in health research (41). Comments on the quality of indicators were rare though, and only a few individual indicators met most of the SMART criteria. Each stakeholder may be interested in different indicators on account of their differing roles and should therefore be involved in early planning regarding the selection and quality of indicators to be used.

A comparison of the M&E systems of the Global Fund to Fight AIDS, Malaria and Tuberculosis and the World Bank discovered that the investment in M&E tended to be heavy on developing M&E frameworks and selecting indicators (there were too many, they lacked validity, and did not fit into a logical framework of inputs, outputs,
What approaches are there for monitoring and evaluation of global public health programmes?

outcomes, and impacts) with insufficient attention to strengthening underlying data systems and analytical processes (42). Several key lessons learned were:

- Avoid complicated, burdensome M&E systems
- Emphasise analytical processes rather than indicators and targets
- Work toward integrating and strengthening national M&E structures
- Find a balance between demonstrating agency-specific effectiveness and combined impact
What approaches are there for monitoring and evaluation of global public health programmes?

Endnote database matrix showing the highly relevant papers with key information

This database can be obtained on request.

KLS Evidence Briefing

26th July 2018
What approaches are there for monitoring and evaluation of global public health programmes?

References


KLS Evidence Briefing
What approaches are there for monitoring and evaluation of global public health programmes?

What approaches are there for monitoring and evaluation of global public health programmes?

https://openknowledge.worldbank.org/bitstream/handle/10986/3732/WPS5245.pdf?sequence=1&isAllowed=y


42. IEG. Comparison of the Monitoring and Evaluation Systems of the World Bank and the Global Fund. 2012. https://openknowledge.worldbank.org/bitstream/handle/10986/26678/726060NWP0Box30IC00GF0WP0april2012.pdf?sequence=1&isAllowed=y