MEASURE Evaluation

Working Paper Series

Outcome Monitoring for Global Health Programs

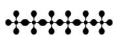
Janine Barden-O'Fallon
Mahua Mandal

August 2014 WP-14-153

This paper has been supported by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through the U.S. Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement GHA-A-00-08-00003-00, which is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, with Futures Group, ICF International, John Snow, Inc., Management Sciences for Health, and Tulane University. The views expressed in this publication do not necessarily reflect the views of PEPFAR, USAID or the United States government. .

Carolina Population Center
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: 919-445-9350
measure@unc.edu

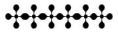
www.cpc.unc.edu/measure



This working paper series is produced by MEASURE Evaluation in order to speed the dissemination of information from research studies. Working papers may be under review or awaiting journal publication, and are distributed as received from the authors with minor editing and formatting adjustments.

The series and other publications are available from the MEASURE Evaluation at:

http://www.cpc.unc.edu/measure



Contents

Section 1. Executive Summary

While many global health programs are engaged in the practice of outcome monitoring, there are few available resources on the practice of outcome monitoring itself. This paper helps fill the gap by providing a general resource on the practice of outcome monitoring. Additionally, included in the paper are a number of useful references and links to resources specific to different aspects of outcome monitoring.

Outcome monitoring is the periodic measurement of the knowledge, behaviors and/or practices that a program or intervention intends to change. While the purpose of outcome monitoring is to provide routine, timely information for program management on whether the program is making progress towards meeting its health objectives, it cannot be used to determine if changes in outcomes are a direct result of the program. Outcome monitoring falls between monitoring outputs to answer the question: "Are we implementing the program as planned?" and evaluating impact to answer the question: "Is the program responsible for impact?" Outcome monitoring instead addresses the question: "Do we see expected changes in the target population?" There has been increased interest in outcome monitoring in the global health community, partly because it allows stakeholders to assess the progress of large, multi-national, or global initiatives, such as the Millennium Development Goals. Outcome monitoring also moves attention beyond the process of program implementation, while being more time and resource efficient than impact evaluations.

Outcome monitoring at the international and national level usually relies on information collected through routine health information systems, nationally representative surveys, such as the Demographic and Health Survey (DHS), and the census or other vital statistics registries. Monitoring outcomes at subnational levels may require special surveys, community surveillance systems, and qualitative methods. Rapid quantitative methods are optimal for collecting outcome monitoring data. Two such methods, cluster sampling and Lot Quality Assurance Sampling, are widely used in global health, including by the World Health Organization's Expanded Programme on Immunization (EPI). Both methodologies are low cost, time efficient, and less complex than population-based surveys, and provide managers with information to make decisions about program resources. Due to the use of small sample sizes, however they have low statistical power, and should not be used to measure incremental changes in outcomes. Qualitative methods, which are used less often, allow for in-depth understanding of program outcomes and may suggest changes that could further improve outcomes. Qualitative methods are particularly useful to monitor changes in domains, such as motivations, social norms and program acceptability/quality, not easily or comprehensively measured with quantitative surveys. Qualitative methods can also capture unintended or indirect effects of interventions. Qualitative tools traditionally used to monitor outcomes include direct observations, in-depth interviews, and focus group discussions. Additionally, certain methodologies such as outcome mapping and the most significant change, are especially participatory and are used to measure changes in outcomes as perceived and experienced – and thus valued – by the people interventions are intended to reach and

affect. Both outcome mapping and the most significant change are multi-faceted methods that can be tailored to and adapted for the contexts within which they are used.

Selecting appropriate indicators is an important step in outcome monitoring. As with all indicators, they should produce the same result when used repeatedly; measure only the condition or event they are intended to measure; reflect changes in a state or condition over time; represent reasonable measurement cost; and provide the information needed to make programmatic decisions. Additionally, outcome indicators need to be specific to the population and program they are monitoring; practical to collect routinely; and responsive to program effects.

To support the process of outcome monitoring, program planners must decide on information needs; whether primary data collection is necessary (or if secondary data sources can be used), and the frequency with which outcomes will be measured. They must also understand the information environment in which the program will be operating, including knowing what types of data are already being collected in the target area, how often the data are collected, and the quality of the data. The program must assess the resources, such as money, time, and staff effort, they can allocate to outcome monitoring, including possible primary data collection. Finally, program planners must decide who to share the data with, the best formats for presenting data, and what types of decisions will be made based on the data.

In general, we find that outcome monitoring practice could benefit from the incorporation of qualitative data collection methods, especially in conjunction with quantitative methods. Program management should also consider how outcome monitoring data and information can help support the health information system in which the program is operating.

Section 2. Rationale for Outcome Monitoring

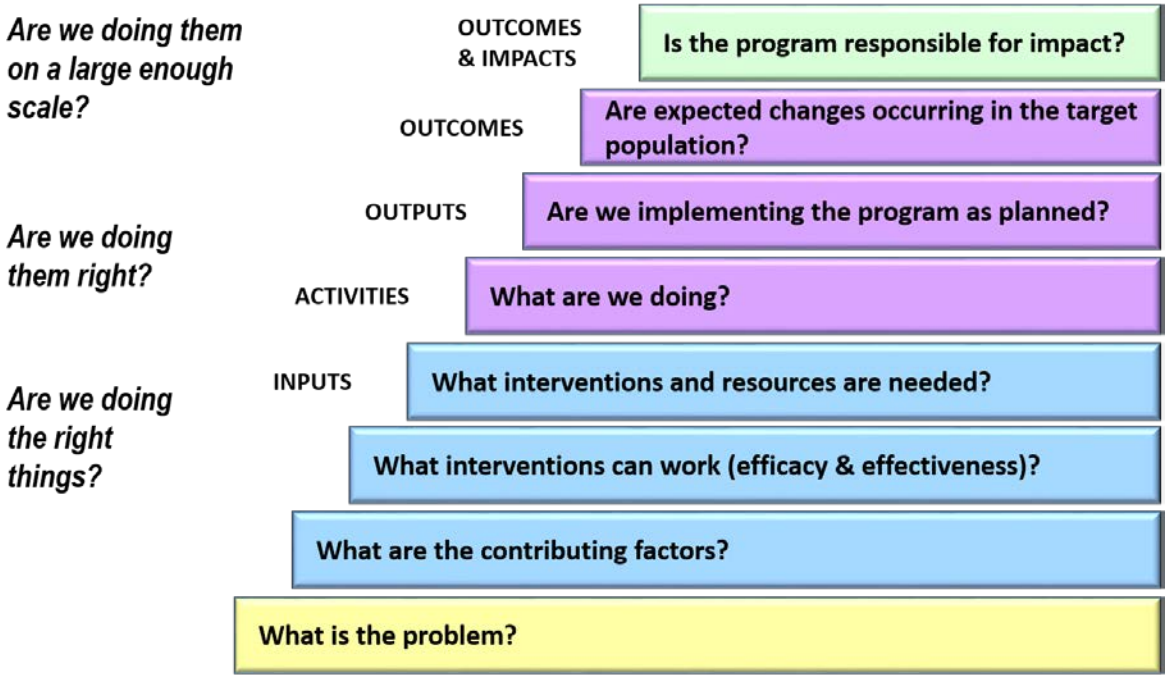
Outcome monitoring is the continuous measurement of the knowledge, behaviors and/or practices that a program or intervention intends to change. The purpose of outcome monitoring is to provide routine, timely information for program management on whether the program is making progress towards meeting its objectives. Outcome monitoring tracks changes in knowledge, behaviors, and/or practices of people over time and can provide an indication of whether the program is having (or will have) the expected impact in the target population. Failure to monitor outcomes on a regular basis can hide fundamental program inadequacies, delay action to improve effectiveness, and ultimately lead to wasted investment.

Monitoring outcomes cannot, however, be used to determine if the changes in outcomes are a direct result of the program. In other words, the attribution of change, or impact, of the specific program or intervention cannot be determined through outcome monitoring. For example, a key component of a malaria prevention program may be the distribution of bednets to pregnant women and mothers of young children. Monitoring the number of bednets distributed by the program will be important to

ensure that targets are met, however, it will not provide information on whether the target population is actually sleeping under those nets (Foreit and Cummings, 2006). Outcome monitoring during program implementation can fill this information gap and inform the program on whether additional steps need to be taken to support the proper use of bednets once distributed.

Outcome monitoring is an important element of monitoring and evaluation (M&E). The M&E process, using a questions-based approach, has been depicted by a "staircase", as shown in figure 1 below (adapted from Rugg, Peersman and Carael, 2004). The bottom steps of the staircase are related to program planning and include activities to identify and understand the health problem, as well as to determine the appropriate interventions and resources needed to address the problem. The next steps in the M&E approach are related to program monitoring activities, which help ensure that program managers know exactly what the program is doing and whether it is being implemented according to plans. Outcome monitoring is one step above this, set between program monitoring activities and impact evaluation activities. This is because outcome monitoring provides information on whether there are observed changes in outcomes over time, but does not reach the step of answering the question of whether the program is responsible for the observed changes. To do this, one must move another step up the staircase to the impact evaluation activities that can address issues related to attribution and scale.

FIGURE 1. M&E Staircase



Adapted from: Rugg, Peersman, & Carael, 2004

Sometimes there is an overlap in terminology. Program monitoring is also referred to as "process evaluation" or "formative evaluation" because it focuses on the implementation process and addresses key questions such as: "How well is the program being implemented?" and "How much does implementation vary from site to site?" Outcome monitoring moves beyond the implementation process by specifically focusing on how much change is taking place in the program's target population. However in practice, data collection for process evaluation and outcome monitoring may overlap. This would be the case, for example, if indicators of outcomes are also used as indicators for program performance, such as when service provision interventions track client satisfaction.

There has been increased interest in outcome monitoring in the global health community. One reason is that outcome monitoring allows stakeholders to assess the progress of large, multi-national, or global initiatives, such as the Millennium Development Goals (MDGs), FP2020, and A Promise Renewed, among others, by tracking population-based trends in important health indicators. Outcome monitoring can also move attention beyond the process of program implementation, as indicated by the Institute of Medicine's 2013 evaluation of the President's Emergency Plan for AIDS Relief (PEPFAR), which found that PEPFAR indicators were focused primarily on inputs and outputs and not on outcomes and impact; therefore the initiative had limited ability to determine the effectiveness of its efforts (IOM, 2013). Finally, not all programs require or are suitable for impact evaluation, perhaps due to the size or scope of the program, budget and/or resource constraints, or even because the effectiveness of an intervention is already well-established. Such programs may, however, benefit from outcome monitoring to help steer program activities and ensure that there are positive changes in the population the program is trying to reach. Indeed, even when impact evaluation is planned for a program, outcome monitoring remains an important aspect of program monitoring, allowing for midcourse corrections during implementation and indicating if changes in the target population warrant an impact assessment (if there are no changes, it would not be helpful to continue with an impact evaluation).

It is worth noting that not all global health programs will be able to conduct outcome monitoring. Outcome monitoring may be ill-advised when the budget is not sufficient to collect the appropriate information, when the time and expertise needed for designing and implementing outcome monitoring is not available, and when programs are not implemented at a large enough scale to elicit change in a target population.

While many global health programs are engaged in the practice of outcome monitoring, have specific training and data collection tools designed for outcome monitoring, and report findings to stakeholders and others, including on their program websites, there are few general resources on the practice of outcome monitoring itself. Exceptions to this include a handbook to guide the data collection process, including the identification of information needs and tool development and implementation (Parkinson and Wadia, 2010). The handbook also contains a useful list of resources and publications for other aspects of outcome monitoring and a "Top Tips" section, which includes advice such as, "Keep your tools short, simple, and clear" and "Make sure people collecting data are trained and supported" (Parkinson and Wadia, 2010, Appendix A). A guide on the analytical process needed for outcome information was developed by the Urban Institute and is available on their website (Urban Institute, 2004). The

MEASURE Evaluation guide on collecting PEPFAR Level 2 MER indicators (Chapman, Foreit and Parker, 2014) includes guiding principles and methodological considerations; survey approaches and tools; and procedures for recruitment, data collection, and data management specific to conducting outcome monitoring on orphans and vulnerable children (OVC) programs.

Section 3. Key Considerations for Outcome Monitoring

Outcome monitoring starts with the M&E framework. The framework serves a guide for all M&E activities of the program and will help identify the outcomes that the program will need to measure. Selecting outcome indicators is an important step. Sometimes donors provide the outcome indictors that they want monitored, other times programs need to determine the indicators themselves. In both cases, the selection and determination of the outcomes and specific indicators is important to ensure that the correct information is collected to monitor the progress of the program.

Indicators for Outcome Monitoring

As with all good indicators, indicators for outcome monitoring should produce the same results when used repeatedly to measure the same condition or event (i.e. are reliable); measure only the condition or event they are intended to measure, and are defined in clear and unambiguous terms (i.e. are valid); reflect changes in a state or condition over time; represent reasonable measurement costs; and provide the information needed to make programmatic decisions (i.e. are timely) (Frankel and Gage, 2007). The mnemonic acronym SMART is used to promote the use of indicators that meet these criteria, and that are Specific, Measureable, Achievable, Relevant, and Time-bound.

Indicators designed to monitor outcomes provide information on changes within the target population and whether changes are occurring to the desired degree and at the desired pace to indicate that the program is on track toward meeting its objectives. Outcome monitoring indicators need to be specific to program outcomes and cover all the relevant outcomes of the program. Additionally, outcome indicators need to be practical for routine data collection, responsive to program effects and able provide information on the potential effectiveness of the program. Given that outcome monitoring is not meant to attribute changes in the knowledge, behavior, or practices of the target population to the program, the interpretation of the outcome monitoring indicators must occur in the proper context. This means that consideration must be given to the potential non-programmatic effects on change as well.

Case Study: Development of Global Indicators and Tools for Assessing Child Well-Being in OVC Programs

Historically, one of the persistent challenges to monitoring and assessing the impact of programs for orphans and vulnerable children (OVC) was the lack of standardized measures and measurement tools for child and household well-being that were tailored to the OVC population. To address this problem, MEASURE Evaluation in collaboration with 49 stakeholders representing implementing partners, donors, national OVC teams, universities, projects and task forces, developed a group of core indicators and tools for monitoring and assessing the wellbeing of children and households. Indicators and data collection tools were designed to enable and standardize the production of population-level child and caregiver well-being data beyond what are available from routine surveys. The data produced are intended to inform programs and enable mid-course corrections, and to enable comparative assessments of child and caregiver well-being and household economic status across a diverse set of interventions and geographical regions.

The indicators and tools were designed to measure changes in child, caregiver and household well-being that can reasonably be attained with program interventions. The indicators selected are amenable to change from diverse program interventions, and are relevant across a wide range of program settings. Some of the outcome indicators may take a long time to manifest and may rely on referrals to other high-impact child survival, education, protection, and economic strengthening programs. In addition to program outcome indicators, the tools include a limited number of indicators that programs may not be able to change, such as household composition and age, but which may enhance or inhibit program success. The tools were also aligned to both the PEPFAR OVC Programming Guidance (2012) and the US Government Action Plan on Children in Adversity (2012).

Most of the core indicators are included in the new PEPFAR Monitoring, Evaluation, and Reporting (MER) indicator list (PEPFAR, 2013). This means that PEPFAR-funded programs will be supported to report on many of these outcome indicators regularly. The tool kit can be found at: http://www.cpc.unc.edu/measure/our-work/ovc/ovc-program-evaluation-tool-kit.

Logistics for Outcome Monitoring

A number of decisions need to be made to support the outcome monitoring process. For example, program planners will need to determine the frequency with which outcomes will be measured. Again, sometimes programs are given this information by donors or other stakeholders as part of the program requirements. Other times, program planners should consider how often the information will be needed to make the appropriate adjustments in the program. As mentioned earlier, this could be at the beginning, middle, and end of a program. Another rule of thumb is that outcomes are monitored on a yearly basis. Note that the OVC outcomes measured for the PEPFAR Monitoring, Evaluation, and Reporting (MER) are collected biennially. The decision on frequency of data collection will be based in part, on the number of outcomes to be measured and the time, resources, and budget needed to carry

out the outcome monitoring. In general, programs would not likely need to assess and report outcomes more than once a year. Some funders require that outcome monitoring is conducted by independent or external consultants who are not involved in program implementation. In such cases, it is likely that a larger budget and time commitment will be needed to conduct the outcome monitoring.

Program planners also need to understand the information environment in which the program will be operating. This includes knowing what types of data are collected in their target area, how often the data is collected, and the general quality of the data. This knowledge will help guide programs in determining whether primary data needs to be collected for outcome monitoring, and if so, how much and how often. Knowing the information environment can also help the program understand how to position itself within the local information system. As much as possible, a program should build off of pre-existing data. For example, because outcome monitoring requires information on the target population, access to census data or other vital registries is often needed. Such information is typically not cost-efficient for a program to collect. Also, as much as possible, a program should not deliberately create a parallel information system. This means that program data collection is, and can be, embedded in local systems as much as possible. This is an especially important consideration for low-resource settings, in which parallel health information systems can contribute to a continuation of weak and under-resourced systems.

Preparing for Outcome Monitoring

The following is a checklist of some of the important steps and questions related to the planning and implementation of outcome monitoring:

- Determine Rationale: Is outcome monitoring part of the M&E framework? Will information obtained from monitoring outcomes help program management to know whether the program is making progress towards meeting its objectives? What information is needed?

- Decide on Frequency: How often are outcome data needed for reporting to stakeholders? How often are data needed for decisions related to program planning and implementation?

- Select Indicators: Are the outcome indicators SMART? Are there standardized indicators for program outcomes?

- Assess Local Data Environment: What sources of data are available for the target population? How will outcome monitoring data for the program fit within the health information infrastructure already in place?

- Assess Resources: How much money, time and staff effort will it take to implement the outcome monitoring as planned? Will the budget cover the collection of primary data? Are staff available and capable to collect and analyze outcome data? Will consultants need to be hired?

- Select Data Collection Methodologies: Will quantitative, qualitative or both types of data be collected? Are there pre-existing tools that can be used or adapted to your program or will new data collection tools need to be developed? This issue will be explored in the next section.

Consider Use of Data: Who does the outcome monitoring data need to go to? What types of decisions will be made based on these data? What would be the best formats for presenting the data to help facilitate these decisions? Are there additional ways in which these data can be used, either internally by the program or externally by others? This issue will also be explored in the next section.

Section 4. Data Collection for Outcome Monitoring

There are several data sources and data collection tools that can be used for outcome monitoring. Usually, the type of data needed and what is already available will direct the decisions on sources and data collection activities. Outcome monitoring at the international and national level usually relies on information collected through routine health information systems; nationally representative surveys, such as the Demographic and Health Survey (DHS) or Multiple Indicator Cluster Survey (MICS); the census or other vital statistics registries; and any other source of population-based data. However, these sources of data can be increasingly limited in use when looking at outcomes at a subnational level (i.e. province, district/county, or lower level). Additionally, program catchments, populations and geographic areas of interest for outcome monitoring may not follow administrative boundaries, and may not always be linked directly to national surveys or information systems. In these cases, special surveys are often needed to monitor outcomes. For example, rider surveys can be included in larger survey activities to collect specific information from a population already under study; community surveillance systems can be set up to closely monitor trends within specific administrative zones; qualitative methods can be used to collect information on outcomes that are difficult to quantify, and include one-on-one interviews, focus groups, and outcome journals, among others; and rapid assessments can be designed to collect specific information on outcomes from target populations.

All the types of data methods mentioned above, and many others, are valid and used in global health outcome monitoring. The choice of what data source and collection methodology are used should be decided by the program context and specific information needs. Whenever possible, program managers should try to use available data, using special studies to fill in the gaps. The following sections describe several common data collection methodologies for tracking outcomes in global health.

Rapid Assessment Methods for Outcome Monitoring

Rapid assessments are optimal for collecting data for routine outcome monitoring; necessary population-based data can be obtained at a fraction of the cost, human resource commitment, and time required by larger, national level surveys, such as the DHS, MICS, or AIDS Indicator Survey (AIS). Because of this, rapid assessments can be conducted as often as needed throughout the life of a program, typically annually or at the baseline/midterm/endline of a program. Additionally, the information that can be obtained by rapid assessments is aggregated at an optimal level for program management, whether that be at the provincial, district, or sub-district level. In contrast, this level of information is often not attainable through large-scale national surveys.

There are two sampling methodologies used for rapid assessments; one is a two-stage cluster sampling methodology, often referred to as 30 x "n". Two-stage cluster designs begin with a sample of 30 clusters within a defined program area. From each of these clusters, a small sample is taken: the first is randomly selected; often the remaining sample points are obtained from the same or next closest household. The other rapid assessment methodology is Lot Quality Assurance Sampling, or LQAS. LQAS requires that program areas be divided into sub-areas, or "lots", from which a small random sample is taken. Both cluster sampling and LQAS are widely used methodologies in global health. For example, cluster sampling, specifically the 30 x 7 design, has long been used by WHO's Expanded Programme on Immunization (EPI) (Henderson and Sundaresan, 1982). A 30 x 30 cluster design has also been used by food and nutrition programs during humanitarian crises (Binkin et al., 1992). Adaptations of the cluster design have also been investigated for use in emergency settings (for example, 33 x 6 and 67 x 3) (Deitchler, Deconinck and Bergeron, 2008) and by the SMART Initiative (2002). In contrast, LQAS has its roots in the manufacturing sector as a quality assurance tool. In its adaptation to the health setting, LQAS has become an increasingly popular methodology for rapid assessment. In a review of LQAS applications from 1984-2004, a total of 805 LQAS surveys were implemented, with the period of 2000-2004 experiencing more than a ten-fold increase in the number of LQAS surveys per year as compared to the period of 1984-1999 (Robertson and Valadez, 2006). LQAS has been used in a wide range of health fields, including immunization and vaccination, malaria, growth and nutrition, and neonatal tetanus mortality, among others. Recent adaptations of the LQAS design include multiple category-LQAS (MC-LQAS) for use to classify results into low, medium, and high categories (Olives et al., 2012); LQAS in combination with an area sampling methodology to conduct a school-based (vs. community-based) study (Myatt et al., 2005); and large country-LQAS (LC-LQAS), which integrates elements of cluster sampling with the LQAS methodology (Hedt et al., 2008).

Cluster sampling and LQAS are similar in that both methodologies provide frequencies for chosen indicators for a program as a whole and thus, both provide managers with information to make decisions about program resources. Some evidence suggests that a benefit of cluster sampling is that it can be less costly than LQAS and take less time for data collection, as only 30 sites need to be visited (Singh et al., 1996). Due to the long history of use with EPI, cluster sampling is also perhaps the more widely recognized methodology for rapid assessments. Conversely, a main benefit of LQAS is that it allows for the collection of additional information at a sub-program level and can therefore be used to identify sub-program areas that are not performing well.

A main disadvantage of both of these methodologies is that they have low statistical power. Both use small sample sizes and have thus traded off statistical power for savings in cost, complexity and speed. As a result, neither method can reach the level of precision that is obtained by DHS and other large surveys, and should not be used for measuring incremental change. Cluster sampling is a good choice when sub-program area information is not needed, when there are no defined sub-program areas, or when LQAS is too costly, as might be the case if large projects or programs are divided into a high number of sub-areas. LQAS on the other hand is useful when sub-program information is desired and when a project is divided into meaningful and well-defined sub-areas.

There are a number of implementation guides for rapid assessments. For example, a thorough and clear description of both methodologies and implementation issues can be found in the Rapid Household Survey Handbook (Davis et al., 2009). On-line training resources, such as those for LQAS provided by the CORE Group at http://www.coregroup.org/our-technical-work/working-groups/monitoring-and-evaluation and published research articles on the application and suitability of the methods, such as those cited above, are available for rapid assessments in a wide range of health fields.

Case Study: Implementation of LQAS for Maternal, Neonatal, and Child Health Outcome Monitoring in Kenya

MEASURE Evaluation implemented rapid assessments using Lot Quality Assurance Sampling (LQAS) for a number of maternal, neonatal, and child health (MNCH) behavior and knowledge indicators in Kenya. Working with the Ministry of Public Health & Sanitation and the Ministry of Medical Services, a series of outcome monitoring surveys were conducted to provide information both for the government and the USAID-funded Aids, Population and Health Integrated Assistance (APHIA II) Programs from 2009-2012. The surveys were designed to provide information for programming and monitoring of indicators related to breastfeeding, use of bed nets, nutrition, immunization, treatment of infectious diseases, prenatal care and delivery, and sanitation and hygiene. The surveys were conducted in 5 contiguous districts within each province, and identified both priority indicators (where overall coverage was low) and priority districts (where performance in the district for a particular indicator was low). In Kenya, baseline information on the targets at the district level was missing, so districts were assessed by the average coverage estimate for each province. An example of how the information was used for programmatic decision-making can be seen by tracking indicator performance in Western Province. In 2009, the indicator for the "percentage of mothers of children aged 0-23 months who boil or chlorinate their drinking water to make it safe" had a coverage of 45% for the sampled area. While performance for this indicator was uniformly low across most districts, one district, Teso South, was flagged as underperforming. As a result, during the following year, the APHIA II program increased training for community health workers so they could intensify public health talks to the community on water and sanitation issues. The focus on sanitation initiatives included: providing more tablets; initiating efforts to educate community members on positive health aspects of clean and safe drinking water; encouraging communities to have water tanks and assisting 20 schools to build water tanks; encouraging the communities to build Ventilated Improved Pit latrines; and educating the community on the importance of washing hands. In 2012, a second LQAS was conducted in Western Province. Results for the same indicator showed an average coverage of 56%, and though the confidence intervals of this estimate overlapped with the previous estimate, the districts were assessed at a higher standard. In this round, Teso South showed marked improvement, and was not flagged for low performance.

Materials related to the implementation of LQAS in Kenya are available on request. Implementation of LQAS in Kenya led to the development of a training package on outcome monitoring and evaluation using LQAS by the APHIA II Evaluation Project with funding from USAID/Kenya. The package can be found at: http://www.cpc.unc.edu/measure/training/materials/outcome-monitoring-and-evaluation-using-lqas. Similar MNCH outcome monitoring work was conducted in Liberia, 2011-2014, and is available at: http://www.cpc.unc.edu/measure/publications/.

Qualitative Methods for Outcome Monitoring

Qualitative methods allow for an in-depth understanding of program outcomes. Qualitative methods are particularly useful in monitoring changes in attitudes, perceptions, motivations and social norms – domains that are not easily or comprehensively measured with quantitative methods. Qualitative methods can also be used to capture any unintended or indirect effects of interventions and can assess the quality and character of the intervention implementation, which may help explain why intended outcomes were or were not achieved. Additionally, qualitative data collection often costs a fraction of the cost of collecting quantitative data, especially from representative samples. As such, it is ideal to use for midcourse correction of programs, particularly when budget constraints do not allow for extensive quantitative data collection. Qualitative tools traditionally used in global health to monitor outcomes include direct observations, in-depth interviews, and focus group discussions. Each of these is best used to elicit a specific type of information. For example, direct observations may be used in the waiting room of a clinic, to assess the process by which clients are registered for their visit and how long they must wait. In-depth interviews, which have been referred to as one-on-one conversations (Rubin and Rubin, 1995) or "social encounters" (Holstein and Gubrium, 1999), are best used to understand people's knowledge and experience with, for example, certain health services or a type of contraception. In contrast, focus group discussions (FGDs) use group interaction to provide data and insights that would be less accessible without the dynamics found in a group (Ulin et al., 2002). FGDs are best used to understand social norms, expectations, values, and beliefs. FGDs are an ideal tool to understand gender norms and roles and how they relate to phenomenon such as early marriage and pregnancy of girls.

While qualitative methods, by their nature, encourage interaction between researchers and research participants, some outcome-focused qualitative methods are especially participatory. These methods, such as outcome mapping and most significant change (MSC), value documenting changes as perceived and experienced by the people programs are designed to reach and affect. Both outcome mapping and MSC may be tailored to and adapted for the context within which they are used, including the type of program and the specific country or region.

Outcome mapping is a multi-faceted method that focuses on outcomes as changes in behaviors, relationships, activities, or actions and recognizes the complex and parallel processes of multiple actors (individuals, groups, or organizations) that lead to such change. While outcome mapping can be used for all stages of a program (planning, monitoring, and evaluating), there are specific steps and sub-steps involved in using the tool for outcome monitoring. These steps include using outcome journals to monitor changes in key actors (referred to as "boundary partners", or the individuals, groups, or organizations with whom the program works directly and who have influence in the development of the program; this group may include beneficiaries, gatekeepers, and the research community, for example). An outcome journal is established for each boundary partner that has been identified as a priority. The outcome journal includes progress markers; a description of level of change (low, medium, or high) for each progress marker; which boundary partners exhibited change; a description of and reasons for change; people and circumstances that contributed to change; evidence of change; unanticipated

changes; and lessons for the program. The program must determine values for high, medium, and low ratings so that outcomes are measured consistently over the course of the program (both across time and across boundary partners). The actual data that populates each outcome journal may be collected through observations, interviews, focus groups, or other methods. To monitor progress over time, outcome journals can be quantified using a predetermined scoring scheme. For example, any value (i.e. high, low, medium) of "expect to see" markers can be set at one point each; any value of "like to see" markers set at two points each; and any value of "love to see" markers set at three points each. Such values are tallied for each outcome journal, and can be used to monitor the progress of each boundary partner over time. While outcome journals may also be used to compare one boundary partner to another, the program must take into consideration factors such as boundary partners "starting" from different places, or having different objectives, where some objectives may be inherently more challenging than others (Earl, Carden, and Smutylo, 2001). Resources on outcome mapping can be found on the Better Evaluation website at: http://betterevaluation.org/plan/approach/outcome_mapping and through the Outcome Mapping Learning Community, which can be accessed via http://www.outcomemapping.ca/.

Similar to the process of outcome mapping and outcome journals, the "most significant change" (MSC) approach involves a number of steps that can provide insight into programs that are complex and produce diverse and emergent outcomes. The process involves identifying the stakeholders who are committed to participate in MSC and, along with the stakeholders, establishing domains of change to be monitored. These domains are not precisely defined like outcome indicators, but deliberately left broad so they can be defined by actual users. The program and stakeholders also decide how often to monitor changes in these domains. Next, stories of significant change are collected from those most directly involved, such as participants and field staff. Data collectors ask simple questions such as: "In your opinion, what is the most significant change that took place for participants in the program during the last three months?" Respondents are encouraged to discuss why they consider a particular change to be significant. These stories can be collected through in-depth interviews, group discussions, or beneficiaries who document their own stories. Initially, respondents allocate their stories to one of the chosen outcome domain categories.

To help reduce the potentially large volume of locally important stories to a small number of more widely valued stories, the most significant of all change stories at one level of a program or organization are submitted to the next level up. As most organizations have hierarchical structures, it makes practical sense to use this existing structure to organize the selection process. The use of multiple levels of selection, called 'summary by selection' enables the whittling down of stories to the outcomes and domains that are most pertinent without burdening any individual or group with too much work. The final set of selected stories provide outcome monitoring data that are then used to understand if the program is on the right track, and what midcourse corrections may need to be made.

After the most significant of change stories are selected and submitted, the results of the selection process are fed back to those who provided the stories. This feedback helps explain which stories were selected as most significant and why, and provides information how the selection process was

organized. Providing feedback aids participants' searches for stories in the next reporting periods, because knowing which particular type of change is valued can lead to further searches for similar changes in that domain. In addition to these essential steps in the MSC method, in some settings it may be useful to verify stories, quantify stories (within individual stories, or across sets of stories), and conduct content analysis (i.e. analyze themes across a set of change stories). A detailed guide to the MSC technique can be found at http://www.mande.co.uk/docs/MSCGuide.pdf (Davies and Dart, 2005). Furthermore, many program websites contain stories of most significant change related to their programs. For example, stories from the introduction and scale-up of the Standard Days Method into the national family planning programs of the Democratic Republic of Congo, India, Mali and Rwanda can be found on the website for the Institute for Reproductive Health at Georgetown University at: http://irh.org/resource-library/5140-2/.

Section 5. Summary and Recommendations for Outcome Monitoring

The global public health community has increasingly embraced outcome monitoring as a necessary part of program M&E. As the field continues to progress, there are a number of considerations to keep in mind. Programs should consider incorporating qualitative methodology to monitor outcomes. The heavy reliance on quantitative methods is perhaps due to the relatively "straightforward" nature of assessing observed changes in indicators such as "Percent of children whose primary caregiver knows the child's HIV status". One strategy is to incorporate qualitative methods into the midline of an outcome monitoring plan. For M&E plans that include yearly monitoring of outcomes, programs may consider alternating between quantitative and qualitative methods, thus collecting both types of information in a cost-efficient manner. Though qualitative data collection and analysis may be perceived to be less rigorous than quantitative methods, it is imperative to have data collectors who are trained in how to effectively conduct one-on-one interviews, FGDs, and observations. The lack of such training precludes obtaining in-depth information, a marker of qualitative methods that can help the program assess if it is "on track".

As a result of outcome monitoring work conducted with the governments in Kenya and Liberia, it was clear that rapid assessments for outcome monitoring can be part of a robust health information system. These country examples are not isolated. Any program should consider the benefits that could be accrued by thoughtful attention to program data needs and the information system in which the program is placed. However, in order to incorporate routine outcome monitoring into health information systems, some basic elements are required. These elements include good planning so that information is available when it's needed for strategic planning and decision-making; an understanding of the type of information that outcome monitoring can provide so that stakeholders know what information they will get and how it will be useful to them; an intention to use the information that is generated in order to strengthen programs and services; and finally, the human and financial resources

necessary to adequately carry out the rapid assessments. With such elements in place, outcome monitoring can be a routine and sustainable M&E practice for both national and sub-national scale programs.

References

Binkin N, K Sullivan, N Staehling, P Nieburg. 1992. Rapid nutrition surveys: How many clusters are enough? Disasters. 16(2):97-103.

Chapman J, K Foreit, L Parker. 2014. Collecting PEPFAR Level 2 monitoring, evaluation, and reporting (MER) indicators: A supplement to the Orphans and Vulnerable Children Survey Tool Kit. MEASURE Evaluation: Chapel Hill, NC.

Davies R, J Dart. 2005. The 'Most Significant Change' (MSC) Technique: A Guide to Its Use. RJ Davies and J Dart: Cambridge, UK and Hastings, Australia.

Davis R, J Luna, A Rodriguez-Lainz, E Sarriot. 2009. The rapid household survey: How to obtain reliable data on health at the local level. ICF Macro: Calverton, MD; and Public Health Institute: Oakland, CA.

Deitchler M, H Deconinck, G Bergeron. 2008. Precision, time and cost: a comparison of three sampling designs in an emergency setting. Emerging Themes in Epidemiology. 5:6.

Earl S, F Carden, T Smutylo. 2001. Outcome Mapping. International Development Research Center: Ottawa, ON, Canada.

Foreit K, S Cummings. Annual monitoring of health outcome indicators: Description and assessment of alternative data collection methodologies. 2006. MEASURE Evaluation: Chapel Hill, NC.

Frankel N, A Gage. 2007. M&E Fundamentals: A self-guided minicourse. MEASURE Evaluation: Chapel Hill, NC.

Hedt BL, C Olives, M Pagano, JJ Valadez. 2008. Large Country-Lot Quality Assurance Sampling: A new method for rapid monitoring and evaluation of health, nutrition and population programs at sub-national levels. The International Bank for Reconstruction and Development/The World Bank: Washington, DC.

Henderson RH, T Sundaresan. 1982. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bull World Health Organ. 60(2):253-60.

Holstein JA, JF Gubrium. 1999. Active interviewing. In Bryman A, RG Burgess. Qualitative research. Vol II. Methods of qualitative research, pp.105-121. Sage: London.

Hoshaw-Woodard S. 2001. Description and comparison of the methods of cluster sampling and lot quality assurance sampling to assess immunization coverage. World Health Organization: Geneva.

Institute of Medicine (IOM). 2013. Evaluation of PEPFAR. Washington, DC: The National Academies Press.

Myatt M, NP Mai, NQ Quynh, NH Nga, HH Tai, et al. 2005. Using lot quality-assurance sampling and area sampling to identify priority areas for trachoma control: Viet Nam. Bull Worl Health Organ. 83(10):756-63.

Olives C, J Valadez, S Brooker, M Pagano. 2012. Multiple Category-Lot Quality Assurance Sampling: A new classification system with application to Schistosomiasis control. PLoS Negl Trop Dis. 6(9):e1806.

Parkinson D, A Wadia. 2010. Assessing Change: Developing and using outcomes monitoring tools. Charities Evaluation Services, at http://www.ces-vol.org.uk/Resources/CharitiesEvaluationServices/Documents/assessingchange-740-748.pdf [Accessed 30 June 2014].

Robertson S, JJ Valadez. 2006. Global review of health care surveys using lot quality assurance sampling (LQAS), 1984-2004. Social Science & Medicine. 63(6):1648-60.

Rubin HJ, IS Rubin. 1995. Qualitative interviewing: the art of hearing data. Sage: Thousand Oaks, CA.

Rugg D, G Peersman, M Carael (eds). 2004. Global advances in HIV/AIDS monitoring and evaluation. New Directions for Evaluation, 103.

Singh J, DC Jain, RS Sharma, T Verghese. 1996. Evaluation of immunization coverage by lot quality assurance sampling compared with 30-cluster sampling in a primary health centre in India. Bull World Health Organ. 74(3):269-74.

SMART (Standardized Monitoring and Assessment of Relief and Transitions). 2002. http://smartmethodology.org/ [Accessed 30 June 2014].

Ulin R, ET Robinson, EE Tolley, ET McNeil. 2002. Qualitative Methods: A field guide for applied research in sexual and reproductive health. Family Health International: Research Triangle Park, NC.

The Urban Institute. 2004. Analyzing Outcome Information: Getting the most from Data. In: Series on outcome management for nonprofit organizations. At: http://www.urban.org/UploadedPDF/310973_OutcomeInformation.pdf [Accessed 10 August 2014].

United States Government. 2012. Action plan on children in adversity: A framework for international assistance 2012-2017. Washington, DC.

United States President's Emergency Plan for AIDS Relief (PEPFAR). 2012. Guidance for orphans and vulnerable children programming. Washington, DC.

United States President's Emergency Plan for AIDS Relief (PEPFAR). 2013. Next generation indicators reference guide. Version 1.2. Washington, DC.

MEASURE Evaluation

Carolina Population Center
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517

http://www.cpc.unc.edu/measure/