USAID
FROM THE AMERICAN PEOPLE

TOOLKIT FOR MONITORING AND EVALUATING GENDER-BASED VIOLENCE INTERVENTIONS ALONG THE RELIEF TO DEVELOPMENT CONTINUUM

9 May 2014

This publication was produced for review by the United States Agency for International Development. It was prepared by Jessica Menon, Victoria Rames, and Patricia T. Morris, PhD, of Development and Training Services.

Section 3

This document is available online. Online documents can be found on USAID's website at www.usaid.gov/gbv and the USAID Learning Lab at http://usaidlearninglab.org/. Documents are also available through the Development Experience Clearinghouse (http://dec.usaid.gov) and on the dTS website at www.onlinedts.com.

Prepared for the United States Agency for International Development, USAID Contract Number AID-OAA-I-10-00014-TO-2-00051.

Implemented by:
Development & Training Services, Inc. (dTS)
4600 North Fairfax Drive, Suite 402
Arlington, VA 22203
Phone: +1 703-465-9388
Fax: +1 703-465-9344
www.onlinedts.com

Toolkit for Monitoring and Evaluating Gender-Based Violence Interventions along the Relief to Development Continuum

9 May 2014

Section 3

Acknowledgments

The development of the Monitoring and Evaluation (M&E) Toolkit was made possible through the generosity of the United States Agency for International Development (USAID). It was prepared through the Transparency, Accountability and Performance Indefinite Quantity Contract (TAP IQC), Gender-Based Violence Strategy Research Agenda Project. Development and Training Services, Inc. (dTS) fielded a three-person team: Jessica Menon (independent consultant), Patricia T. Morris, PhD (director, Gender Practice at dTS), and Victoria Rames (independent consultant) to develop the Toolkit. Peter Pawlak, an independent consultant working through Chemonics, participated in the fieldwork that formed the basis of the Toolkit. Staff at dTS—Alex Ginn, Ashley Mills, and Dawn Traut, and dTS interns Payal Chandiramani, Aicha Cooper, and Megan Sullivan—contributed time and effort in the development of the Toolkit.

The Toolkit is the result of a collaborative learning process that took place from December 2012 to November 2013. USAID staff based in Washington, DC, including Misrak Brhane, Tiare Cross, Niloufer De Silva, Carolyne Siganda, and Kelli Young, provided leadership and technical guidance during all stages of the Toolkit's development. The research team further thanks Niloufer De Silva and Tiare Cross for drafting select sections of the Toolkit. The USAID country mission staff, including Katherine Reniers and Nettie Jannini (Haiti), Passanna Gunasekera (Sri Lanka), and Betty Mugo and Monica McQueary Azimi (Kenya), shared their firsthand experience on the M&E of gender-based violence (GBV) interventions along the relief to development continuum during the research team's field missions in March–July 2013. They also organized focus groups and meetings with staff of USAID implementing partners, United Nations (UN) agencies, and other national and international nongovernmental and community-based organizations, as well as with beneficiaries of select GBV programs, in Haiti, Sri Lanka, and Kenya. Through these consultations, the research team gathered data and insight into the M&E for GBV interventions, which ultimately informed and shaped the Toolkit.

The research team also benefitted enormously from the opportunity to learn from GBV project/ program beneficiaries, women and men, in Haiti, Sri Lanka, and Kenya. They shared their views on what they perceived to be the most important changes that the GBV programming had made in their lives. The research team extends its appreciation for field test inputs received in July 2013 from representatives at United Nations Children's Fund, UN Women, Gender-Based Violence Recovery Centre, International Rescue Committee (IRC), Federation of Women Lawyers Kenya (FIDA), Coalition on Violence Against Women (COVAW), National Gender Equality Commission (Kenya), MSF-France, No Means No Worldwide, Population Council, Neighborhood Alliance Initiative, Liverpool Voluntary Counseling and Testing (LVCT), PSI/Pathfinder, Femnet, and Aphia-Pathfinder. The research team is also grateful to the IRC/Peace Initiative Kenya and its local implementing partners—COVAW, FIDA, and Rural Women Peace Link—for their valuable technical guidance and contribution of staff resources and time in the field-testing of the Toolkit.

The research team appreciates the feedback received on the first draft of the Toolkit from implementing partners and USAID staff in the field, including Passanna Gunasekera (USAID/Sri Lanka) and Kathy Kantengwa, MD, MPA (MSH/Haiti). Lastly, the research team thanks independent consultants and GBV experts Jeanne Ward and Julie Lafreniere, Christine Heckman of the Inter-Agency Standing Committee (IASC) GBV Area of Responsibility Rapid Response Team, and Samira Sami, Health Scientist at the US Centers for Disease Control and Prevention for their enormous support and technical guidance during the development of the Toolkit.

CONTENTS

ACRONYMS

GBV	Gender-based violence
GBVIMS	Gender-based violence information management system
IASC	Inter-Agency Standing Committee
IDP	Internally displaced persons
M&E	Monitoring and evaluation
MTE	Midterm evaluation
PIRS	Performance indicator reference sheet
RDC	Relief to development continuum
SMS	Standard Messaging System
SOPs	Standard operating procedures
USAID	United States Agency for International Development
USAID/OFDA	USAID's Office of U.S. Foreign Disaster Assistance
USG	United States Government

SECTION 3

3. IMPLEMENTING THE M&E PLAN

Ideally, the M&E plan will be implemented after careful planning, particularly during the pre-crisis and post-crisis phases. But in the crisis phase, there is often not enough time for such planning before starting GBV programming. Section 3 provides broad guidance on implementing the M&E plan along the relief to development continuum (RDC) in this exact situation. It focuses on the collection of monitoring data, the assessment of data and program quality, and the realization of RTEs, MTEs, and final evaluations.

3.1 COLLECT MONITORING DATA

KEY CONSIDERATIONS:
COLLECTING MONITORING DATA FOR THE M&E OF GBV INTERVENTIONS

You can conduct performance monitoring in accordance with the M&E plan, discussed in **Section 2.5.1**, and through the use of the PIRS (**Section 2.6**) or the project/program indicator tracking table. You can also use **Annex C** to support the selection of data collection tools (**Table 3-1**) if your organization did not already do this when it developed the M&E plan.

You may also use different tools to work around limitations on monitoring direct service provision to GBV survivors. As an alternative to direct observation/monitoring of providers who serve GBV survivors you can interview medical, mental health, and legal aid providers.

Table 3-1. Data Sources for Performance Monitoring

Data Sources for Performance Monitoring
Secondary Data Sources
• Existing national statistics, databases, and reports, including national census
• Existing national and local plans, strategies, policies, laws, and frameworks related to GBV and gender equality
• Existing institutional/academic demographic, socioeconomic, reproductive health, and GBV surveys
• Existing evaluations, baseline surveys, or other documents from existing projects in the area of influence, or assessments and reports from other clusters/sectors (child protection, etc.)
• Existing mapping (stakeholders/services)
• GBV Area of Responsibility 3/4/5W service mapping tool
• Media (newspapers, radio, television)
• Regular project/program reporting, reviews, and evaluation reports
Primary Data Sources
• Multi-cluster/sector initial rapid assessment
• Review and analyze case data or trends (including from GBVIMS)
• Police reports and court records review/analysis

Data Sources for Performance Monitoring
• GBV legal case files review/analysis
• Ministry of Health statistics data or GBVIMS reporting
• Tracking of referral documents
• On-site observation
• Surveys
• Key stakeholder analysis
• Key informant interviews/peer-to-peer interviews • (Qualitative)
• Mapping of GBV prevention and response services provision
• Community mapping
• Safety and security mapping
• Focus groups
• Case studies
• Protection monitoring
• Community consultations to discuss GBV issues, contributing factors, and specific problems requiring action
• Community-based monitoring
• Pre- and post-tests, or other methods to assess changes in knowledge as a result of awareness-raising activities
• Print media and social media (e.g., Facebook)
• SASA! (Start, Awareness, Support, Action) Outcome Tracking Tool, based on skills, behavior, attitude and knowledge in the SASA! Raising Voices.

Example from the field: Using community-based qualitative performance monitoring tools

The Neighborhood Initiative Alliance that operates in Kajiado, Kenya, uses community-level meetings in which community outreach workers can observe certain changes that they feel are important indicators of change (and project success). The Alliance's staff suggests that it may be possible to develop a simple 1- to 2-page form with the specific data that the organization wants to capture (important qualitative indicators), including the rating "openness of community to discussing GBV" on a scale of 1–5 (1 minimum, 5 maximum), and clearly defining the level of openness at each rating. This would be a welcomed monitoring tool that they could report on quarterly in a systematic way, which ultimately would also assist in evaluations.

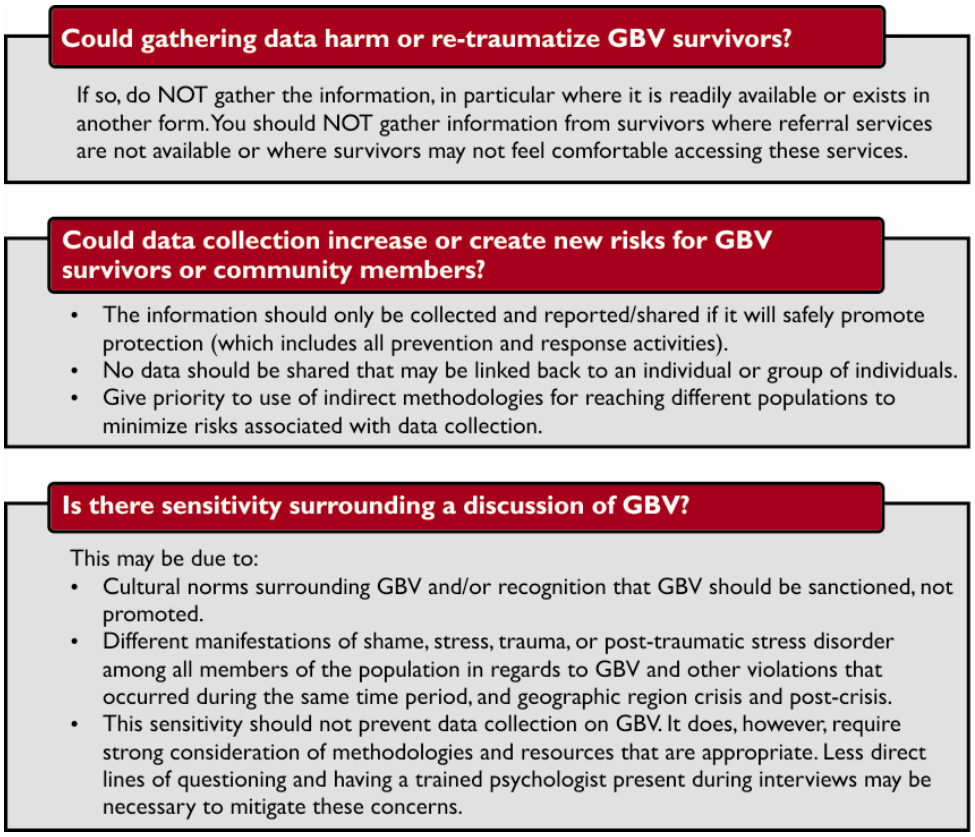
In Sri Lanka, for example, the organization WIN used regular (monthly) community reporting to gauge community perceptions of GBV. "Everyone knows everything in the community," so WIN was able to identify new reports of GBV in the past month and respond accordingly with community awareness or response activities.

Also in Sri Lanka, a rural women's development organization suggests that focus groups be organized by CBOs through temple contacts to facilitate the solidification of trust between GBV survivors. They suggest that it is important for a local CBO to partner with an international organization that spends time (weekly meetings for 1–3 months to build trust rather than a one-off focus group discussion) to get real and accurate information. This is important to build into evaluation time frame, costs, and required human resources.

3.1.1 Identify Whether Data Collection Is Feasible and Ethical

One of the key considerations for primary data collection is to determine whether it is feasible and advisable in the context in which your organization is undertaking GBV programming. During pre- and post-crisis phases, it is usually feasible to collect GBV prevalence data. Major security or political concerns, ethical considerations, or significant stigma all may affect feasibility; as does secrecy associated with discussing GBV (including a socially repressive environment for women and girls). **Figure 3-1** provides an overview of the ethics of primary data collection and, by extension, secondary data collection.

Figure 3-1. Ethical Considerations in Selecting and Adapting Data Collection Tools



Both the Inter-Agency Standing Committee GBV guidelines and the Sphere Standards are unequivocally clear about collecting primary data during a crisis phase. They state that, even in the absence of "proof," all humanitarian actors have a responsibility to assume that GBV—especially sexual violence—is happening and to plan and implement their interventions so as to mitigate GBV-related risks. It may be nearly impossible to gather such data due to limited safe and secure access to communities, limited time to respond, and ethical considerations. The latter include discussing GBV issues with a population during a crisis and the possibility of re-traumatizing survivors, their families, or communities; political sensitivities; or potential interview fatigue.

In the crisis phase, gather primary qualitative data on the nature and scope of GBV from reported GBV cases, to have some sense of GBV prevalence. Focus groups, key stakeholder interviews, community mapping, service mapping, and anecdotal information may also help you to identify the perceived

magnitude of GBV and the availability of GBV response services. You can use these data to design the M&E plan and initiate programming.

Once the crisis subsides, or reduces in intensity, you may gather GBV prevalence data if it is safe, appropriate, and clearly useful to do so. Many organizations on the ground, particularly in the crisis phase, find that using ongoing project/program interventions (such as community theater) can be invaluable tools for ongoing needs assessment. These can allow outreach workers to observe and document behaviors, attitudes, and knowledge that can then influence future programming.

Does your organization have the mandate and capacity to collect data?

A critical aspect of collecting data is your organization's mandate, programming, strengths, and capacity to gather primary data. For example, consider the following examples of data collection needs:

- **GBV prevention efforts in schools**. Does your organization have the expertise necessary to interview children, which may include child survivors of GBV?
- **Provision of psychosocial support or medical services to GBV survivors**. Does your organization have psychologists trained to provide psycho-social support to traumatized populations, or can it secure the resources necessary to hire someone with that profile?
- **Working with survivors**. Does your organization have the training and capacity to adapt existing situational/needs assessment tools to ensure that they are survivor-centered (safety, confidentiality, respect, and nondiscrimination)?
- **Sensitive data collection**. Does your organization have the mandate and capacity to undertake potentially sensitive data collection—including storage facilities, staff training, and background—and the trust of the community?

If your organization does not have the capacity or human resources required to collect GBV-related secondary and primary data, it should not do so. One solution would be to partner with another organization that does have this capacity and organizational mandate.

3.1.2 Consensus on the Type of GBV to Monitor

Consensus and clarity on the type of GBV that your organization will address and therefore monitor and evaluate are essential. This understanding is crucial for the selection of appropriate monitoring tools and methods and staff to conduct monitoring. For example, before selecting focus groups to monitor changes in the perceptions of risk of sexual harassment, all staff must be absolutely clear that they will be monitoring only sexual harassment. Going beyond addressing the perceptions of the risk of sexual harassment to include sexual violence could be dangerous. It could endanger participants or even re-traumatize them (especially if a trained counselor is not present). It is critical not to conduct focus group discussions with groups of GBV survivors about their experiences of violence. Only as a last resort should you interview key stakeholders.

3.1.3 Decisions on Monitoring Priorities

Projects can be monitored for implementation progress, sectoral technical quality, adherence to best practices, and for fraud and corruption. Your organization should develop monitoring standard operating procedures (SOPs) specific to the project/program and have them vetted by the project management and M&E technical staff. The SOPs should cover frequency, method of monitoring, personnel, safety considerations, and reporting lines to ensure accountability of results and swift action to correct project deficiencies.

3.1.4 Selection and Training of the Project Monitoring Team

One of the key steps in conducting M&E is the selection and training of the project monitoring team, from project staff to community members. This should be part of the M&E plan (**Section 2.5.4**). General guidelines for selecting and training the project monitoring team are shown below. Members of the team should:

- Have appropriate training and experience on how to put into practice the GBV guiding principles: *safety*, *confidentiality*, *respect*, and *nondiscrimination*
- Have training and expertise on GBV and (if conducting interviews) on interviewing GBV survivors
- Understand how to obtain voluntary and informed consent when using focus groups and key stakeholder interviews (see **Annex T** for guidance)
- Be trained and held accountable for maintaining the confidentiality of all data collected (see **Annex T**)
- Understand and be able to implement measures to safely store and protect data
- Take into account language, ethnicity, religion, political orientation/affiliation, region of origin, sex, and related safety and protection concerns (e.g., Although selecting project monitoring staff who are of the same ethnicity as GBV survivors may seem appropriate, this may not always be the case)
- Have access to and know about available services. The team should know how to safely and appropriately provide referrals to GBV survivors who identify themselves so that they have the option to receive services and support if they so choose. If no services exist, we strongly recommend that you do not interview GBV survivors. By the same token, if interviewing survivors is absolutely necessary, your organization should make someone available to speak with survivors during and after the interview if they express an interest in speaking to a counselor.

RDC CONSIDERATIONS

Crisis Phase

- **Constraint 1:** Existing national organizations and development actors may be unable to continue using monitoring tools, which may result in inconsistencies or interruptions in data collection.
- **Solution:** Support national organizations and development actors to adapt monitoring the tools to the best of their ability. Ensure data analysis describes challenges and potential data inconsistencies.
- **Constraint 2:** Frequent turnover of personnel that typically occurs in a crisis may have an impact on the collection of data for performance monitoring.
- **Solution:** Engage and train several personnel within your institution and community stakeholders in performance monitoring to ensure consistency. Make sure M&E plans are of high quality and detailed so that new personnel may use them consistently. Most important, focus on strengthening existing local organization's efforts and their M&E plans rather than making new parallel ones. This will work toward more sustainable programming and longer-term M&E beyond the crisis phase.

Post-crisis Phase

- **Constraint 1:** M&E plans used for shorter term relief efforts may not have planned for ongoing monitoring of important outcomes beyond the crisis.
- **Solution:** Collaborate with national organizations and development actors to provide data and take-up continued monitoring to be folded into M&E plans developed post-crisis where there are synergies.

3.2 MONITOR FOR PROGRAM QUALITY

KEY CONSIDERATIONS:
MONITORING PROGRAM QUALITY IN THE M&E OF GBV INTERVENTIONS

It is important to use and interpret performance and situational data to monitor program quality consistently. This involves monitoring progress toward achieving the targets detailed in the M&E plan in **Section 2.5** and the PIRS in **Section 2.6**. Regular monitoring reports (monthly, quarterly, annually) should indicate progress toward indicators as planned, with particular attention to program quality. Questions to answer may include:

- Are those benefitting from the project/program the ones who are specified in the M&E plan?
- Are there any intended beneficiaries or other segments of the population who are excluded from the project benefits?
- Are there biases in programming?
- If services are being provided, are they of the quality expected, as detailed in the M&E plan? Are they meeting the (international or national) standards that were detailed in the M&E plan?
- Are there occurrences of fraud or corruption that are related to the project/program activities?

Early warning systems can serve as a performance monitoring tool to gauge whether targeted awareness efforts and contingency planning are serving their intended purpose.

GBV case management (monitoring) can be a very effective performance monitoring tool in both relief and development contexts.[1] The Suriya Development Organization in Sri Lanka keeps detailed case management files, which offer rich information on the progress towards the achievements of medium- and longer-term outcome and impact indicators. Information on changes in the survivor's attitudes and confidence levels, ability to seek support from others, self-sufficiency, community and family responses to the survivor needs, police responses, and the like are all regularly documented. Additionally, the organization has weekly case management meetings to gauge effectiveness and quality of care.[2]

The collection of project data on a monthly or weekly basis (needed in crisis settings) and analyzing the reports in comparison to previous months are important to determine trends in usage of services or changes within the population, or to identify project misperceptions or issues that need immediate attention. These trends will often need to be investigated further in order to completely understand the reasons for the change. In general, drops or increases in service usage rates of +/-10% should be flagged for further investigation.

To mitigate bias or flaws in one type of monitoring method, it is important to include a variety of methods across the project cycle to capture information in different ways. Monitoring methods might include in-person visits by GBV or M&E technical staff, monthly output reports of activities, analysis of beneficiary list by type of vulnerability or other relevant criteria, and/or quality checks using checklists relevant to the given project (i.e., checklist of provision of GBV services in a primary healthcare setting).

[1] Personal communication, interview with Francesca Rivelli of International Red Cross/Haiti, 11 March 2013, Port-au-Prince, Haiti.

[2] Personal communication, interview with Sarala Emmanuel, Director of Suriya Development Organization, May 2013, Sri Lanka.

You must also determine what level of monitoring is appropriate for your context. In general, higher risk contexts require a greater level of monitoring, particularly when project management staff do not have regular access to project sites or where corruption is especially high. There is no consensus within the humanitarian community on how many aid recipients should be monitored directly or indirectly. However, there are minimum considerations for follow-up. You should consider, based on project design, minimum thresholds that will be affected by whether or not the project includes service delivery or awareness components. For example, an organization may set a threshold of 5–15% of GBV clients who will receive a post-service client survey (hard copy, in-person, text, or in a follow-up appointment), which will help to gauge client satisfaction with the service and determine whether quality standards were met. Alternatively, organizations may set as a goal that 100% of GBV awareness-raising sessions will include a way to allow communities to directly express complaints or concerns to project staff who will be trained in proper follow-up.

Monitoring traditional and social media can be an effective tool for monitoring changes in community attitudes towards GBV. Technology, such as Standard Messaging System (SMS), can be used to obtain information quickly in a crisis or post-crisis phase and can be used for baseline assessments, performance monitoring, and evaluation. In Sri Lanka, this technology has been used to send out multiple-choice SMS to gauge attitudes on GBV. The responses to the mini-survey were then used to tweak anti-GBV messages.

3.3 MONITOR FOR DATA QUALITY

Monitoring for data quality is a core function of the implementation of performance monitoring. Though there are always trade-offs between the cost and quality of data, USAID missions and implementing partners should balance these two factors to ensure that the data are of sufficiently high quality to support the appropriate level of management decisions by both entities. Performance data should be as complete and consistent as management needs and resources permit (USAID ADS 203).

KEY CONSIDERATIONS:
MONITORING DATA QUALITY IN THE M&E OF GBV INTERVENTIONS

Annex I of the Toolkit includes a **Data Quality Assessment Checklist** adapted from the USAID Learning Lab[3] to verify the internal quality and consistency of the data collected in the M&E plan. The checklist provides a series of key questions to ensure that the data meet USAID's data quality standards. It is organized according to five key categories:

1. Validity: Do data clearly and directly measure what we intend?

2. Integrity: Are mechanisms in place to reduce the possibility that data are manipulated for political or personal reasons, or incomplete due to management problems?

3. Precision: What margin of error is acceptable given the likely management decisions to be affected?

4. Reliability: Using the same measurement procedures, can the same results be replicated?

5. Timeliness: Are data sufficiently current and available frequently enough to inform management decision-making at the appropriate levels?

[3] USAID. n.d. Data Quality Assessment Checklist and Recommended Procedures. http://usaidlearninglab.org/sites/default/files/resource/files/Data%20Quality%20Assessment%20Checklist.pdf

The checklist may also be used even in the planning for M&E to anticipate and address key data quality issues even before beginning the implementation of GBV programming.

Data quality assessment procedures can be compressed during a crisis phase. Data quality assessments are not required to be submitted to USAID/OFDA. However, humanitarian partners should discuss known data limitations, methods to triangulate data in the field, and methods to ensure the objectivity of the data reported.

It is also important that some specific factors or changing circumstances that could affect GBV data quality be taken into account. Though it may be possible to anticipate some factors, you will have to address others as they become apparent, through the regular monitoring of data quality. These may include any or all of the following:

- Changes in access to affected areas and affected populations due to roads or entire communities becoming inaccessible or unsafe
- Changes in the ability to communicate with key partners in the project/program area due to a breakdown in telecommunications or lack of translation
- Little time to review and ascertain the quality of initial baseline data gathered in the pre-crisis period, resulting in data quality issues during the crisis period
- High/rapid turnover of both trained international and national staff, including M&E officers, which may impact the application of consistent data collection methods
- Changes in the willingness or availability of key informants working with community groups, resulting from a number of factors such as increased political tensions, threats, and increased stigma surrounding GBV and GBV survivors
- Unpreventable loss or destruction of GBV data (in particular during a conflict, where data may be intentionally destroyed by parties to the conflict)
- Changes in the government policy regarding the collection of GBV data, in particular the collection of survey data.

RDC CONSIDERATIONS

- **Constraint:** During a crisis, the quality of GBV data may be affected by numerous logistical, safety, and political constraints, as well as increased sensitivities surrounding GBV.
- **Solution:** As soon as data quality issues are discovered, brainstorm and implement modified or alternative methods to gather quality data, while continuing to respect the guiding principles on ethical GBV data collection outlined in **Section 1**.

Crisis Phase

- **Opportunity:** During pre-crisis contingency planning, anticipate potential GBV data quality issues, and take measures to address them, including increasing the number of skilled data collection staff, training local partners and communities to collect data (methods, ethics, technology, and data storage).

Post-crisis Phase

3.4 CONDUCT REAL-TIME, MIDTERM, AND FINAL EVALUATIONS

When evaluating programs and projects, it is useful to consider the principles below, which are outlined by OECD/DAC for the Evaluation of Development Assistance. Humanitarian and development actors and donors, including USAID, use them. These principles are often included in evaluation scopes of work, and evaluations reports must address all of them in their findings. Further, M&E data collection activities must be organized around the lines of inquiry related to each principle. It is of great importance for any evaluation—particularly for the evaluation of GBV interventions—to achieve the overarching goals of the USG Strategy on GBV to improve learning and understanding of effective GBV interventions via high-quality M&E.

1. **Relevance:** The extent to which the activity is suited to the priorities/policies of the target group, recipient, and donor.

   Evaluation Questions:

   - To what extent are the objectives of the program still valid?
   - Are the activities and outputs of the program consistent with the overall goal and the attainment of its objectives?
   - Are the activities and outputs of the program consistent with the intended impacts and effects?

2. **Effectiveness:** Measures the extent to which an activity attains its objectives.

   Evaluation Questions:

   - To what extent were the objectives achieved/are likely to be achieved?
   - What were the major factors influencing the achievement or nonachievement of the objectives?

3. **Efficiency:** This measures the qualitative and quantitative outputs in relation to inputs. This generally requires comparing alternative approaches to achieving the same output, to see whether the most efficient process has been adopted.

   Evaluation Questions:

   - Were activities cost efficient?
   - Were objectives achieved on time?
   - Was the program or project implemented in the most efficient way compared to alternatives?

4. **Impact:** The positive and negative changes produced by a development intervention—directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental, and other development indicators. The examination should be concerned with both intended and unintended results. It must include the positive and negative impacts of external factors, such as changes in terms of trade and financial conditions.

   Evaluation Questions:

   - What has happened as a result of the project or program?

- What real difference has the activity made to the beneficiaries?
- How many people have been affected?

5. **Sustainability:** Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.

   Evaluation Questions:

   - To what extent did the benefits of a program or project continue after donor funding ceased?
   - What are the major factors that influenced the achievement or nonachievement of sustainability of the program or project?

Further, when GBV activities are incorporated into larger projects/programs, which may focus on different sectors, GBV considerations should be integrated into the project's sector-specific evaluation questions. Although outside of the scope of the Toolkit, users may consider GBV-specific evaluation questions for GBV activities that may be incorporated into larger project/program evaluations.

KEY CONSIDERATIONS:
FINALIZING THE EVALUATION PLAN AND CONDUCTING MIDTERM AND FINAL EVALUATIONS OF GBV INTERVENTIONS

The finalization and implementation of the evaluation plan should follow from the M&E plan for evaluations, as outlined in **Section 2.5.2**. Three key steps are discussed below.

1. Prioritize the evaluation questions

The evaluation questions were developed in **Section 2.5.2**. When you prioritize the evaluation questions, it is essential that they be closely linked to the input, process, and output indicators for a performance evaluation, and the outcomes and impacts for an impact evaluation. Prioritize evaluation questions according to the following criteria to ensure that they:[4]

- Are important to program staff and stakeholders
- Address important program needs
- Reflect the program goals, strategies, and objectives of your organization's project/program
- Can be answered with available resources, including funds and personnel expertise
- Can be answered within the available time frame
- Provide information to make program improvements
- Will be supported by the partners of the program

[4] Adapted from US Centers for Disease Control and Prevention, Department of Health and Human Services, 2009. Evaluation Briefs, No. 4.

- Link to the program. Once the questions are determined, they can and should be checked/verified against the program strategic plan, M&E framework, and work plan to make sure they remain relevant.

2. Select evaluation methodology and tools

Once the evaluation questions have been prioritized, determine who will collect the data and what tolls will be used to provide information for answering the evaluation questions (outlined in **Section 2.5.2** as part of the M&E plan). Possible data sources may include secondary sources (project/program monitoring data and reviews of case management files and police records, and primary sources (stakeholder interviews, focus groups, and on-site inspections). **Annex C** can support the identification of any additional needed tools.

Surveys using randomized samples are a useful evaluation tool and cited numerous times in the PIRS as a key data source. The box below provides an overview of different sample methods. Surveys using randomized samples should be planned well in advance of implementation. They will typically first be implemented to inform a baseline assessment (**Section 2.7**) and should be carried out by highly trained and qualified M&E experts experienced in sampling and survey design. This is why the M&E plan (**Section 2.5**) puts greater emphasis on identifying a local or international academic partner to assist the development or humanitarian actors with evaluations and programming. It is beyond the scope of the Toolkit, however, to go into great detail about the art of survey sampling.

Determine who will be responsible for collecting and analyzing the data to answer the evaluation questions. Independent third-party evaluation specialists often conduct evaluations. **Figure 2-11** (**Section 2.5.4**) provides an overview of considerations in selecting staff persons for conducting evaluations.

A note on sampling for surveys

Both probability and informal sampling procedures may be used for surveys. Their descriptions are brief and elementary, however. The focus here is on probability sampling, given the emphasis on randomized survey sampling in the PIRS in the Toolkit. In most cases, particularly for randomized survey sampling that is used at baseline and for evaluations, a third-party evaluator with expertise in survey design should be contracted to identify the most appropriate sampling strategy and sample size, which depends completely on the specific project context. The underlying concept is that large groups of people, organizations, households, or other units can be accurately examined by scrutinizing a small number of the group. A formula is used to draw inferences from the sample for the whole population.

Probability sampling vs. informal sampling

In probability sampling, each unit in the population has an equal chance of being selected for the sample. The selection of units for the sample is carried out by chance procedures, and with known probabilities for selection. Informal selection uses convenience or common sense rather than mathematical reasoning.

Illustrative probability sampling methods

- **Simple random sampling:** Each unit of population (individuals, households, organizations, etc.) has an equal chance of being selected, and may be drawn by lottery or numbering all units and entering in a program to select random numbers.
- **Stratified random sampling:** Uses zoning to divide the sample into three or four layers. Once the strata are identified, sampling is completed using select primary sampling units in each stratum equal to the total proportion of the total population in the stratum.
- **Cluster sampling:** Clusters of a population (such as farms, neighborhoods) are identified and random samples are chosen from each cluster. A 30 X 30 cluster sampling is popular, particularly in the humanitarian context where 30 households are sampled from 30 clusters (yielding a sample size of 900 households, and with an average family of 6 persons, or 5,400 individuals).

Informal sampling methods

- **Convenience sampling:** Only those easily reached by interviewers are included in the sample.
- **Judgment sampling:** Uses the judgment or advice of experts or the survey designer to construct samples.
- **Snowball sampling:** Begins with few population units but increases until it ends up with the required sample size.
- **Quota sampling:** population is divided into various strata, and a predetermined number of people, or quota, is selected for each.

Generally, in determining the sample size, it is important to have confidence that your survey results are representative. For a 95% confidence level (which means that there is a 5% chance of your sample results differing from the true population average), a good estimate of the margin of error (or confidence interval) is given by $1/\sqrt{N}$, where N is the number of participants or sample size. A reasonable rule of thumb in a larger population is selecting 3–5% of the population (ACF-International 2010).

- There are also instances where purposive sampling may be used. For example, you may decide to target both your survey and services to a specific population of an IDP camp known to be at risk for sex trafficking prior to coming to the camp. In this case, targeted surveys may be suitable, particularly if they require skilled and trained psychosocial staff to address the specialized needs of the population surveyed.

See the Resources in **Annex Y** for further guidance on sampling methodologies, including ACF-International (ibid.), which includes more in-depth guidance on sampling methods and choosing sample sizes.

3. Prepare the Evaluation Report

Once the evaluation is complete, use **Annex V** as an illustrative outline to prepare the evaluation report. Follow the data analysis plan and methodology detailed in the evaluation plan and the evaluation inception report. The evaluation report template can serve as a guide for preparing meaningful, useful, and credible evaluation reports that meet quality standards. It does not prescribe a definitive section-by-section format that all evaluation reports should follow. Rather, it suggests the content that should be included in a quality evaluation report.

Criteria to ensure quality of the evaluation report (USAID Evaluation Policy 2011)

- The evaluation report should represent a thoughtful, well-researched, and well-organized effort to objectively evaluate what worked in the project, what did not, and why.
- Evaluation reports shall address all evaluation questions included in the SoW.
- The evaluation report should include the SoW as an annex. All modifications to the SoW, whether in technical requirements, evaluation questions, evaluation team composition, methodology, or timeline, need to be agreed upon in writing by the technical officer.
- The evaluation methodology will be explained in detail and all tools used in conducting the evaluation such as questionnaires, checklists, and discussion guides will be included in an annex in the final report.
- Evaluation findings will assess outcomes and impact on males and females.
- Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (selection bias, recall bias, unobservable differences between comparator groups, etc.).
- Evaluation findings should be presented as analyzed facts, evidence, and data and not based on anecdotes, hearsay, or the compilation of people's opinions. Findings should be specific, concise, and supported by strong quantitative or qualitative evidence.
- Sources of information need to be properly identified and listed in an annex.
- Recommendations need to be supported by a specific set of findings.
- Recommendations should be action-oriented, practical and specific, with defined responsibility for the action.

Recommendations in the report should be formulated in a way that will facilitate the development of a project/program management response. Recommendations must be realistic and reflect an understanding of the commissioning organization and potential constraints to follow-up. Each recommendation should clearly identify its target group and stipulate the recommended action and rationale.

The lessons learned from an evaluation comprise the new knowledge gained from the particular circumstance (initiative, context outcomes) that is applicable to and useful in other similar contexts. Frequently, lessons highlight strengths or weaknesses in preparation, design, and implementation that affect performance, outcome, and impact.

RDC CONSIDERATIONS

Crisis Phase

- **Constraint 1:** Evaluators and those that are participating in/subject to an evaluation may be exposed to violence. Ensuring safe "evaluation space" in areas where either conflict or disaster has taken place may be of a concern.
- **Solution:** Postpone the evaluation until safety is secured for all parties involved.
- **Constraint 2:** It may not be possible to conduct a transparent and public evaluation process that brings together actors and beneficiaries that may be in conflict to hear one another's viewpoints. Meeting separately with actors in conflict may also create tensions/distrust.
- **Solution:** Consult with local experts and determine the best method for engagement, possibly meeting separately with actors in conflict or meeting publicly only with leaders of groups.

Post-crisis Phase

- **Constraint 1:** Baseline and monitoring data, including information on implementation, are often lacking for periods during crisis.
- **Solution:** Identify data available and conduct a performance evaluation to the extent possible. Use qualitative data collection tools that may provide contextual information on the "baseline" before the project/program was implemented. At the end of a crisis-phase, in the design of an "exit strategy," these data can feed into the evolution of evaluating outcomes in the post-crisis phase, such as medium-term outcomes, including changes in behavior, shifts in power balances, more networks, facilities and services to support survivors, strengthened interagency coordination, or more open and responsive agencies.
- **Constraint 2:** Following a crisis, previous plans for impact evaluations may face challenges due to inconsistency in data collection methods related to security, safety, or ethical issues that arose during the crisis.
- **Solution:** Do your best to continue plans for the evaluation. Make sure challenges in completing the evaluation per the original plan are clearly identified. Work with local partners and humanitarian and development actors to identify strategies to fill data gaps. Focus on longer-term outcomes such as changes in social norms and practices to reduce the prevalence of GBV.