VOLUME 12A: SUPPLEMENTAL GUIDANCE ON THE FEED THE FUTURE EVALUATION DESIGN TEMPLATE

NOVEMBER 2015

INTRODUCTION AND PURPOSE

USAID's Bureau for Food Security (BFS) recently developed a Feed the Future Evaluation Design Template to guide third party evaluators and USAID staff in documenting how performance and impact evaluations will be conducted of Feed the Future projects and activities. The Template encourages third party evaluators to consider and describe all aspects of the evaluation being undertaken, including its intended purpose, the type of research questions to be answered, methods of data collection and analysis, and personnel needs. Before an evaluation begins, relevant USAID staff review and approve the Evaluation Design.[1]

This supplemental guidance serves to assist third party evaluators in completing the Feed the Future Evaluation Design Template. It is also intended to guide USAID Operating Unit (OU) staff in their review to ensure critical components of the Evaluation Design are documented, practical, and respond to the OU's evaluation needs. This guidance specifically provides instructions to evaluators on how to complete section 2 of the Template with an adequate level of detail that meets USAID's evaluation criteria. The components covered under section 2 of the Template in which this guidance document provides additional instruction on includes: 1) Evaluation Purpose (section 2a of the Template), 2) Evaluation Research Questions (section 2b), 3) Methodology for Quantitative and Qualitative Data Collection (section 2c), 4) Methodology for Quantitative and Qualitative Data Analysis (section 2d), and 5) Outcome Measures (section 2e).

The methods and approaches described in each section of this guidance are not intended to be exhaustive or prescriptive, but rather provide information to evaluators on various options available to them in designing a Feed the Future performance or impact evaluation. Whether an evaluator chooses to use the methods described in this guidance or not, it's important the chosen methods and approaches are fully described in the Evaluation Design Template.

3a. EVALUATION PURPOSE

At the outset of developing and completing the Evaluation Design, it is important that evaluators, OUs, and other relevant stakeholders agree on the main purpose or purposes of the evaluation at hand, and identify any possible limitations in meeting those intended purpose(s).The purpose(s) of an evaluation should inform the timeline, resource allocation, stakeholder involvement, and context in which the evaluation is taking place.

Most often, the evaluation purpose is identified and specified by OUs in the original Statement of Work (SoW). However, where evaluators encounter stated purposes that are unclear, unrealistic, impractical, or do not address the specific evaluation needs of the OU, there will

[1] An Evaluation Design can often be referred to as an Evaluation Protocol or, more simply, an Evaluation Plan. Whichever term is used, these documents are drafted by the selected evaluator post-award to describe the various components that guide the implementation of an evaluation.

often need to be a choice about where resources are devoted. Discussions to define final evaluation questions and, as necessary, prioritize resources should take place between the evaluator and USAID prior to initiating further work.

Evaluation in USAID has two primary purposes: *accountability* to stakeholders and *learning* to improve effectiveness. While this also applies to Feed the Future evaluations, OUs often state more detailed evaluation purpose(s) in their SoW to guide the evaluator in understanding how information from the evaluation will be used to support OU needs. In general, Feed the Future evaluations will have stated purposes that fall under one of the following categories:

1. **To contribute to a broader evidence base** that informs future policy and practices by those inside and outside the Agency

2. **To improve project effectiveness** in achieving intended results

3. **To inform strategy development or project design,** such as Country Development Cooperation Strategies

4. **To demonstrate results and return on investment** (i.e. best value for money)

5. **To sustain legitimacy across stakeholders**, including the US Government, national governments, and US taxpayers

A Special Note on Impact Evaluations: On occasion, OUs may specify, when stating the Evaluation Purpose for an impact evaluation, the specific type of research design to be undertaken by the evaluator (i.e. experimental, quasi-experimental, or non-experimental). Where OUs do not specify this, the evaluator conducting the impact evaluation should describe the type of research design they plan to use in the Evaluation Purpose (section 2a of the Template) or under the Methodology for Quantitative and Qualitative Data Collection section (section2c of the Template). For more information on the types of impact evaluation and research design, please refer to Volume 4 of Feed the Future's M&E Guidance Series which discusses impact evaluations.

3b. EVALUATION RESEARCH QUESTIONS

In section 2b of the Evaluation Design Template, the evaluator should specify the questions the evaluation will seek to answer. Clearly defining the evaluation questions to be answered is a fundamental step in the evaluation design process as it lays the foundation for the type of quantitative or qualitative data collection and analysis methods to be used. Often, the SoW provided by the OU will include a list of evaluation research questions, but collaboration between the evaluator and OU may yield a more refined set of questions.

Evaluation questions should be clearly stated and easy to understand. The questions should be objective, i.e. not suggesting an answer which could potentially bias the reader or respondent. Moreover, evaluation questions cannot and should not address every aspect of an activity or project; instead, evaluation questions should address specific issues where more information is needed and can be addressed by the evaluation team in the agreed upon timeframe. A rule of thumb is to include between five to 10 questions per evaluation. Note that follow-on or sub questions would be counted as part of the original question. For example, “Has the intervention caused changes in participants’ economic behaviors? How? To what extent?” would be counted as one question as those questions would all be treated or answered through similar data.

Evaluation questions must be answerable with quantitative or qualitative data that are already available (i.e. secondary data) or can be collected at a reasonable cost (i.e. primary data). It is important to remember that a question for a sector assessment or a needs assessment is not an evaluation question. Evaluation questions must relate back to the performance of a specific project or activity and should fall within the manageable interests of a project or activity.

For OUs looking for helpful tips in writing good evaluation questions, refer to USAID's Learning Lab Checklist for Defining Evaluation Questions.

3c. METHODOLOGY FOR QUANTITATIVE AND QUALITATIVE DATA COLLECTION

The methods used to collect data for an evaluation will be defined by the questions the evaluation seeks to answer. Some questions can best be answered through quantitative data, such as whether the intervention has had an effect in reaching the desired outcomes, to what extent outcomes have changed over time, or which approach is most effective. Other evaluation questions can best be answered through qualitative data, such as how beneficiaries are perceiving Feed the Future interventions or how and why interventions are making changes occur. Most often, evaluations need both quantitative and qualitative data (often referred to as mixed methods) to comprehensively answer the questions the evaluation seeks to answer. For example, questions such as "how and to what extent is an intervention causing a desired outcome" can best be answered through both quantitative and qualitative data.

There is no preferred type of data (quantitative or qualitative) that USAID evaluations must use. The decision to focus on quantitative or qualitative data, or a mix of both, should be based on the intended purpose of the evaluation and the questions it seeks to answer. In the following sections, BFS provides instruction to evaluators on the various components that should be included under the Methodology for Quantitative and Qualitative Data Collection section in the Evaluation Design Template (section 2c).

Selecting Quantitative and Qualitative Methods of Collection

When completing section 2c of the Template, the evaluator must describe the method(s) of data collection for obtaining quantitative and/or qualitative information. A variety of data collection methods can be used, and evaluators should seek to use the methods that are the most effective and cost efficient.

Quantitative methods of data collection usually involve a survey questionnaire to a sample of respondents, although other forms of quantitative data, such as geo-spatial, agro-ecological, or climatic information, can be obtained from other sources. Qualitative methods of collection are numerous and involve a variety of approaches to elicit information from respondents across themes or domains. A list of qualitative data collection methods with brief descriptions is available in Annex 1.

In addition to selecting and describing the method(s) of data collection, section 2c of the Template must also explain how quantitative and qualitative data will be recorded while data collection is underway. For quantitative surveys, this section should explain, for instance, if the data will be collected through paper surveys or using a computer-assisted personal interviewing

approach (CAPI), such as tablets and PDAs. It should also describe how data will be transferred from paper or electronic sources into a final format for analysis.

For qualitative methods, it is recommended to record individual and group interviews and later transcribe and translate the recordings into the language in which they will be analyzed (usually English). If the context or situation in which data are collected makes recording the conversations unfeasible or problematic, the evaluator should describe in this section the method by which the information will be accurately recorded. This could mean, for example, having a separate note taker documenting the conversation and translating notes into the language of analysis.

Sampling Strategy

Any social science research, including evaluations, requires the design and selection of samples for study. Even with very small populations or case studies, evaluators must decide which people, contexts, or organizations need to be sampled to answer the evaluation questions and determine how the sampling of those populations will be conducted.

Section 2c of the Template should clearly describe the sampling strategy, or sampling design, to be used in the evaluation and why the sampling strategy is appropriate. The sampling strategy should describe the population (people, contexts, or organizations) to be sampled, the proposed sample size, and the factors that make it an indicative (i.e. representative) sample. In quantitative evaluations, probability sampling is typically considered the most ideal and rigorous sampling strategy. Probability sampling entails the random selection of subjects so that each subject has a known and non-zero probability of selection. This makes it possible to generalize findings from the sample to the total population.

For qualitative evaluations, probability sampling can be used, but is often not ideal nor the most efficient means of collecting qualitative information. In non-probability samples, units or respondents are deliberately selected to represent characteristics of a group within the sampled population. Samples in qualitative evaluations are often purposive, or criterion-based, meaning respondents are selected because they represent a particular characteristic the evaluator wants to examine and are most likely to generate useful data for the evaluation. Purposive sampling strategies may focus on socio-demographic characteristics or may link to shared experiences the evaluation seeks to explore. The characteristics, experiences, or other criterion used to select a purposive sample should always be determined by the purpose of the evaluation. Even when non-probability sampling is used, samples should be chosen in a systematic way and described in the Evaluation Plan to ensure that the data is a credible and indicative sample. See Annex 2 for more information on the types of probability and nonprobability sampling strategies.

Sample Sizes

Defining the sample size in the Evaluation Design is a critical part of the sampling strategy in both quantitative and qualitative evaluations. Whatever the purpose of the evaluation may be, evaluators can only draw precise and accurate findings with an appropriate sample size. An undersized sample (or study) can be a waste of resources by not having the capability to produce useful results, while an oversized one can use more resources than necessary.

In general, it is much better to increase the accuracy of data collection than to increase the sample size after a certain point. In quantitative evaluations, the sample size is determined by the number of observations needed to ensure statistical accuracy and precision of specific variables and to detect statistically significant differences between observations. As a general

rule, the desirable sample size in quantitative evaluations, specifically for calculating continuous variables (i.e. means), is determined by the expected variation in the data: the more varied the data are, the larger the sample size we will need to obtain the same level of accuracy.[2] Sample size for proportional indicators, such as prevalence of poverty, will also take into account the amount of change expected within the indicator, and other survey design factors, such as the design effect. A quantitative study should aim to quantify well-defined variables, for example, the proportion of under-five-year-olds who received Vitamin A supplementation. Sample size calculations are based on estimates of what these proportions are likely to be based on the baseline estimates and targets established by the activity, or an informed guess or results of previous surveys. In many Feed the Future evaluations and especially any contemplating primary quantitative data collection, it is recommended that evaluators employ a statistician to help in calculating adequate sample sizes for quantitative data collection.

Qualitative samples use different parameters for determining size. Usually, qualitative samples are much smaller in size than quantitative samples for several reasons: 1) there is a point when very little new evidence is obtained from each additional observation (unit or respondent), often called the "saturation point," 2) qualitative research does not aim to provide estimates of prevalence or incidence with statistical precision; and 3) qualitative data collected through an individual observation are typically extremely rich and complex, making it unmanageable or unnecessary to collect hundreds of observations to assess the intended research questions.

According to Ritchie et al. (2003),[3] there are a number of criteria to consider when determining qualitative sample sizes: the heterogeneity of the population; the number of selection criteria; the degree to which defining criteria are related or interrelated; groups of special interest that require intensive study; multiple samples within one study; type of data collection methods; and the budget and resources available. The determination of a qualitative sample size must be made based on judgment across those and potentially other criteria. Evaluators may use an approach where they estimate a maximum sample size and integrate a process in data collection to determine when the saturation point is reached, such as having interviewers periodically discuss the types of information their interviews are yielding and if they believe there are many new concepts being uncovered. To determine maximum qualitative sample sizes, Ritchie et al. (2003) provide a few general rules of thumb:

- Individual interview samples should be under 50 and can become difficult to manage in terms of the quality of data collection and analysis if over 50. Sample sizes as high as 70 to 80 should be seriously questioned and allowed only if there are clear reasons for having a larger sample.

- For group discussion samples, maximum sample sizes should be around 90 to 100 (12 to 14 groups) above which the sample becomes difficult to manage. Samples of 140 to 150 (around 20 groups) should be seriously questioned and allowed only if there are clear reasons for a larger sample.

---

[2] Fresle, D. et al. 2004. *How to Investigate the Use of Medicines by Consumers*. World Health Organization and University of Amsterdam. Retrieved from: http://apps.who.int/medicinedocs/en/d/Js6169e/

[3] Ritchie, J., Lewis, J., and Elam, G. (2003) in Qualitative Research Practice: A Guide for Social Science Students and Researchers. Sage Publications: London.

For data collection methods other than interviews or focus groups, evaluators can refer to the criteria and rules of thumb above to determine an appropriate sample size.

Triangulation

In evaluation, as in all research, triangulation is used to validate results or findings by cross verifying collected data with other sources of information. It is important for the evaluator in section 2c of the Template to describe whether and how triangulation will be done to strengthen the credibility of evaluation findings. In social science research, including evaluation research, triangulation uses two (or more) methods and/or data sources to corroborate results. In section 2c (and section 2d, where relevant) of the Template, evaluators should explain how they will use different types of methods and data sources to support findings and/or identify inconsistencies in evaluation results. Triangulation is often the reason evaluations use a mixed-method approach, and the Evaluation Design should specify how and when multiple methods will be used to triangulate findings. For more information on triangulation, please refer to USAID's Technical Note on Conducting Mixed-Method Evaluations.

3d. METHODOLOGY FOR QUANTITATIVE AND QUALITATIVE DATA ANALYSIS

Analyzing data to summarize information, elicit patterns, and identify phenomena is an important part of every evaluation. While methods of data analysis are wide-ranging, they are generally grouped into two categories: quantitative and qualitative methods.

Strategies for data analysis and how the data will be synthesized to answer evaluation questions should be decided early on and described in section 2d of the Template. Whether an evaluation is using quantitative, qualitative, or a mix of both methods for data analysis, this section should identify and describe the methods and tests that will be used to answer the evaluation questions, including any tools or software that will be used to aid in the analysis. The full range of methods and tests that should be used to answer the evaluation questions may not be known at the time of writing the Evaluation Design, but this section should, at a minimum, lay out the primary methods of data analysis. In the paragraphs below, we discuss various quantitative and qualitative data analysis methods, but a more comprehensive list can be found in Table 1.

Quantitative data are most commonly analyzed statistically, using descriptive and/or inferential methods, such as measures of dispersion, central tendency, or multivariate analysis to examine the factors contributing to the direction and magnitude of change. In addition, statistical tests are often used to test the significance of differences between groups or of changes over time.

Quantitative data analysis methods can also be referred to as numeric analysis methods. Both descriptive and numeric analyses allow evaluators to identify recurring and systematic patterns in observations, which, if present, provide insights into the relationships between observed phenomena. Standardization across measures in quantitative data analysis makes it possible to aggregate measures and to make statistical comparisons among individuals, households, regions, and time periods.

Qualitative methods for data analysis are more diverse and focus on meaning, providing complex descriptions and examining detailed information about relationships between variables rather than quantifying variables or variable relationships. Data collected using qualitative

methods can sometimes be referred to as textual data and can be analyzed and interpreted descriptively or numerically. The objective of descriptive analysis is to understand the unique characteristics of observations - the particular context, household, organization or individual.

When conducting qualitative analysis, observations should be reviewed until the point of saturation. Saturation is the point at which there are few or no new items or themes that emerge from findings. Codebooks are often developed and used to code patterns or themes that appear. Computer Assisted/Aided Qualitative Data Analysis Software (CAQDAS), like *Atlas.ti* or *NVIVO*, can also be used at the outset to assist in identifying themes that exist in the entire set of observations or after the codebook has been developed to run queries on the codes and themes collected.

Table 1. Options for analyzing quantitative/numeric and qualitative/textual data[4]

Quantitative or Numeric Analysis Methods Options
• **Summary statistics (e.g. means, medians, proportions, ranges):** providing a quick summary of data, which is particularly useful for comparing one project to another, before and afterwards. • **Frequency tables:** arranging collected data values in ascending order of magnitude, along with their corresponding frequencies, to ensure a clearer picture of a data set. • **Measures of central tendency:** a summary measure that attempts to describe a whole set of data with a single value that represents the middle or center of its distribution. • **Measures of dispersion:** a summary measure that describes how values are distributed around the center. • **Cross-tabulations:** obtaining an indication of the frequency of two variables (e.g., gender and frequency of school attendance) occurring at the same time. • **Correlation:** a statistical technique to determine how strongly two or more variables are related. • **Multivariate descriptive:** isolating the contribution of and relationships among two or more related variables. • **Parametric inferential:** carried out on data that follow certain parameters. The data will be normal (i.e., the distribution parallels the bell curve); numbers can be added, subtracted, multiplied and divided; variances are equal when comparing two or more groups; and the sample should be large and randomly selected. • **Non-parametric inferential:** relates to data that are flexible and do not follow a normal distribution; also known as “distribution-free;” data are generally ranked or grouped. Examples include: ranking, the chi-square test, binomial test and Spearman's rank correlation coefficient.
Qualitative or Textual Analysis Methods Options
• **Thematic coding:** recording or identifying passages of text or images linked by a common theme or idea, allowing the indexation of text into categories.

[4] Peersman, G. (2014).Overview: Data Collection and Analysis Methods in Impact Evaluation, Methodological Briefs: Impact Evaluation 10, UNICEF Office of Research, Florence.

- **Content analysis:** examining coded text to elicit themes and quantify the number of times a theme is mentioned in an observation. Reduces large amounts of unstructured textual content into manageable data relevant to the (evaluation) research questions.
- **Typological analysis:** analyzing data to create a classification system, taken from patterns, themes, or other kinds of groups of data.
- **Analytic Induction:** examining phenomena in textual data to develop a hypothetical statement, then systematically examining similar phenomena to see if they fit the hypothesis. Hypothesis is revised as more observations of phenomena are examined until all observations are assessed.
- **Logical Analysis/Matrix Analysis**: uses coded data to develop an outline of generalized causation or logical reasoning processes. Data are summarized in flow charts, diagrams, etc. to pictorially represent or in narrative descriptions.
- **Archival Analysis:** reviewing the documentary archive related to a project/activity, as well as the context in which it is operating, to gain a deep understanding of the chronology of events related to the project - from conception to implementation.
- **Discourse Analysis:** analyzing text in documents or interview transcripts, to see if patterns and themes emerge in the way that text authors or interview participants frame and talk about things that indicate how they conceive phenomenon and experience them.

3e. OUTCOME MEASURES

Section 2e of the Template should describe and identify the outcome measures that can be used to translate the evaluation questions into specific and measurable results. In many instances the SoW will include a set of outcome measures that are of interest to the OU. If it does not, or the evaluator identifies other outcomes measures, the Plan should describe these outcomes and explain how they relate to the evaluation questions.

Outcome measures are used by projects and activities to document changes that should occur as a result of targeted interventions. Where output measures document the tangible products or services an activity produces, outcome measures document the benefits, learning, or behavioral changes that happen as a result. Outcome measures capture the steps required to meet the intended goal of an activity or project. They are usually included as part of the project or activity performance monitoring plan.

Similarly, evaluations should draw on outcome measures to assess the extent to which a project or activity has or is meeting its high-level goals. Outcome measures should always be specific and measurable to ensure that data can be collected, analyzed, and interpreted to answer the evaluation questions. Depending on the evaluation question, evaluators may be need to look at quantitative, qualitative, or a combination of both quantitative and qualitative outcome measures.

If an evaluation is being conducted of a specific activity, the implementing partner will have a set of outcome measures laid out in its monitoring and evaluation plan. Where appropriate, these outcome measures, and associated data collected by the implementing partner, should be used to answer the evaluation questions. If an evaluation is being conducted at the project-level, the OU should also have identified outcome measures in its Project Appraisal Document, which should be considered by the evaluator to guide the analysis. In some instances, however, the evaluator may have to identify other outcome measures, either to replace or complement the existing ones, and better focus its analysis of the research questions.

Annex 1. Brief Descriptions of Qualitative Methods of Data Collection

1. **In-Depth Key Informant Interviews** involve conducting intensive individual interviews with a small number of respondents to explore their perspectives on a particular idea, program, or situation. One of the advantages of an in-depth interview is that it provides more detailed information than what might be available through other data collection methods, such as surveys. In-depth interviews are usually conducted using a semi-structured questionnaire, with open and closed-ended questions, and can be conducted over the telephone or face-to-face.

2. **Focus Groups** are small groups of six to ten people who are led through an open discussion by a skilled moderator. It is a semi-structured qualitative data collection method in which a purposively selected set of participants gather to discuss issues and concerns based on a list of key themes developed by a moderator. Usually no more than 10 questions are asked in a focus group and the discussion is free-flowing. Focus groups are economical to conduct, and they can yield detailed qualitative information about a project or program from a relatively large number of respondents.

3. **Positive Deviance Inquiry (PDI)** is an assessment tool used to explore existing capacities and resources in the community. The approach does not primarily focus on identification of needs and the external inputs necessary to meet those needs, but instead seeks to identify and optimize existing resources and solutions within the community to solve prevalent problems such as malnutrition. In doing so, PDI involves an in depth study of households within the community whose behaviors and practices enable them to find better solutions to problems than their neighbors.

4. **(Social) Network Analysis** is a methodology for analyzing social or organizational networks. Network analysis is useful to understand how individuals, communities or organizations providing services interact with each other. Network Analysis involves mapping and measuring interactions and relationships between actors in the network.

5. **Participatory Mapping,** also referred to as indigenous mapping, counter mapping, and community mapping, involves the creation of maps by community members and stakeholders to provide a visual depiction of how the community members perceive the space in which they live and interact. Often these maps are used for decision-making at the community and regional levels, for advocacy or for targeting interventions, but can have other unintended positive impacts such as creating unity, raising awareness of social issues, and empowering members to engage in the betterment of their community.

6. **Seasonal Calendars** are a participatory tool used to explore seasonal variations in livelihoods in regards to agriculture and food security and to identify periods of particular stress and vulnerability. Either individually or in group, community members create a local context calendar that dissects the year in a way that is relevant to the community (months, seasons, etc.). The calendars capture various types of information such as relative amounts of rainfall, female and male labor efforts, food availability, water availability, income and expenditures, and periods of hunger. Livestock, fodder, markets and health/diseases information can also be captured in this method.

7. **Participatory Photography** is a qualitative research method that consists of taking photographs of photo points or fixed places in the community to help monitor changes in the landscape over time. It is useful to stimulate discussions on changes to land use, land cover, water bodies, and erosion over time. The participatory photography method should

be used in combination with other qualitative methods to capture social changes as well as physical changes.

8. **Participant Observation** is an anthropological data collection method that requires the researcher to become a participant in the culture or context being observed. Participant observation can provide highly useful insights in understanding how participants interact with one another, how they communicate, and how much time they spend on certain activities. This method may be conducted, for instance, to obtain an understanding of the daily routine of farm life or observe farming practices first hand. It is one of the most intensive qualitative methods as it requires the researcher to become accepted as a member of the culture in order to ensure observations are of a natural phenomenon.

9. **Direct Observation** is a method of collecting valuable information on activities, behaviors, and physical features of a project or intervention without verbal responses from community members. Direct observation is a particularly useful evaluation method when the indicators of interest are behaviors, interactions, and attitudes. This method typically involves extensive note taking so that analysis of the text can be performed to increase objectivity.

10. **Most significant change** is a participatory monitoring method where stakeholders tell and document stories of significant change in their lives, communities, etc. From the collection of stories, the most significant ones are used to understand the intervention impact. Sometimes the stakeholders are asked to identify issues that are important for them and then tell stories to articulate the most significant change related to those issues. This may be followed by a discussion to derive consensus on the most significant change. This method is particularly useful when program outcomes are difficult to quantify and may vary across participants.

11. **Transect Walks/Land-use Transects** help to explore environmental, economic, and social resources within a community along a given transect. Local participants are divided up into various groups (by sex, occupation, etc.) and asked to identify a line through their community that they believe transects the main variations in topography, land-use, and resources, such as residential areas, schools, water points, natural resources, land use, etc. Facilitators probe for details about the area while taking notes and drawing sketches. The output of the land-use transect is usually a map or diagram that can be used to stimulate and inform discussions around land-use patterns, resource allocation and distribution, conflicts, problems and planning.

12. **24-hour Daily Activity Profile** is a research tool to appraise the schedule of each family member within a 24 hour period. It is useful for identifying activities and rest periods of family members to better plan and target development efforts. It is best used when wanting to compare the work done by each family member during the day. The activity profile is developed by drawing a matrix on the ground or on paper together with men, women, youth, etc. The interviewer plots time (in hour intervals) against the activities of each family category. This information can help to identify problems, constraints or possibilities for interventions as well as to provide data on changes and trends over time. Time allocation data can be used to stimulate discussion on the time use dimension of women's empowerment, potential trade-offs if economic activity and child care ability and quality etc.

Annex 2. Types of Sampling Methods

Type of Sampling	Method
Probability Sampling: sample has known probability of being selected	
Simple random sampling (SRS)	Evaluator randomly selects targeted number of population elements. Each element in the population has a known and non-zero probability of selection.
Systematic sampling	Evaluator randomly selects the first element of the sample, then subsequent elements are selected using a fixed or systematic interval until the desired sample size is reached.
Stratified sampling	Evaluator first separates target population into mutually exclusive, homogeneous segments (strata), and then random sample of elements are selected from each segment (stratum).
Cluster sampling	Evaluator randomly selects elements of the population from naturally occurring or administratively created groups (i.e. clusters).
Non-Probability Sampling: sample does not have a known probability of being selected	
Purposive sampling	Evaluator deliberately selects (i.e. non-random) information-rich elements for in-depth study.
Intensity sampling	Evaluator selects information-rich elements that manifest the phenomenon intensely but are not extreme cases.
Deviant case sampling	Evaluator selects elements with highly unusual manifestations of the phenomenon in question.
Stratified purposeful sampling	Evaluator selects elements within particular subgroups that share characteristics of interest; used to facilitate comparisons between different groups.
Snowball or chain sampling	Evaluator selects elements whereby existing elements identify future elements from among their

	acquaintances.
Maximum variation sampling	Evaluator selects elements with a wide variation on dimensions of interest; can help to identify common patterns that cut across variations.
Convenience sampling	Evaluator selects elements based on those easiest to reach.

Annex 3. Characteristics of Quantitative and Qualitative Approaches

Evaluation Activity	Quantitative Approach	Qualitative Approach
Selection of subjects or units of analysis	Random sampling to ensure findings can be generalized, and to permit statistical testing Subject selection methods are fully documented	Choice of selection procedure varies based on purpose of study Purposive sampling often used to ensure representation of all important groups Important to clearly document selection criteria
Research protocol	Data usually recorded in structured questionnaires Extensive use of pre-coded, closed-ended questions Standard protocol must be followed consistently throughout study	Interview protocols are the most common instrument, often semi-structured. Data collection instrument may be modified during the course of the study as understanding grows
Data Collection and Reporting Methods	Mainly numerical values or closed-end variables which can be subjected to statistical analysis Some open ended questions may be included Observational checklists with clearly defined categories may be used	Textual data, recorded verbatim or notes Focus groups (usually less than 10 people) and meetings with large community groups Key informant interviews Participant and nonparticipant observation Photography
Triangulation	Data source triangulation using secondary data (e.g., census data, national household surveys, etc.) Consistency checks built into questionnaires to provide independent estimates of key	Investigator triangulation: A monitor can observe a focus group or group meeting to identify any potential bias resulting from how the session was conducted and to provide an independent perspective

	variables	
	Triangulation in mixed methods: qualitative information is used to validate quantitative data and vice versa	
Data Analysis	Descriptive statistics Multivariate analysis to examine factors contributing to the magnitude and direction of change Significance tests for differences between groups	Each subject treated separately to examine the unique characteristics of each person or group Analysis emphasizes context of study and how it affects understanding of findings Extract themes, patterns